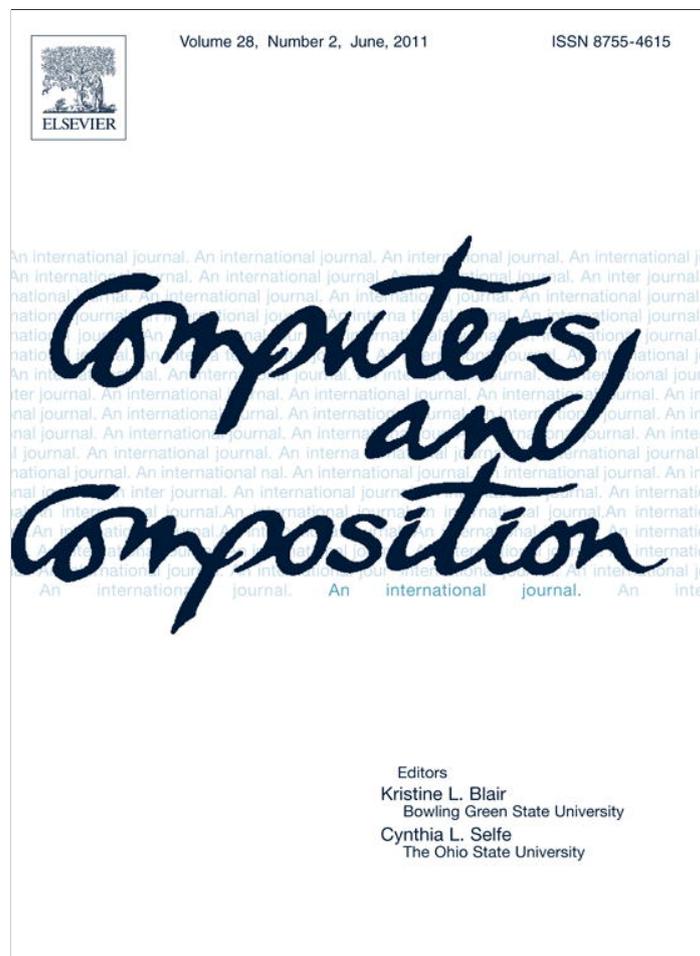


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Technology-Mediated Writing Assessments: Principles and Processes

Bill Cope*, Mary Kalantzis, Sarah McCarthey, Colleen Vojak, Sonia Kline

University of Illinois, Urbana-Champaign, IL, USA

Abstract

This paper explores developments in technology-mediated writing environments that may support new forms of formative assessment and the closer relation of formative to summative assessment. Not only might these provide more learner-responsive and effective assessment of writing, but they may also support the assessment of disciplinary knowledge embedded in written and multimodal texts. After an overview of current debates on contemporary assessment practice, the paper goes on to develop six principles for effective assessment of writing. On this basis, the paper identifies potentially promising aspects of emerging processes of technology-mediated writing assessment.

© 2011 Elsevier Inc. All rights reserved.

Keywords: Writing; Assessment; Technology; Internet; Multimodality

1. Writing Assessment Today

Today's assessment practices have been widely criticized on a number of grounds, two of the most important of which are their emphasis on summative assessment at the expense of formative assessment and their focus on the kinds of empirical and conceptual verities that can be measured in multiple choice texts (Ravich, 2010; Ryan & Shepard, 2008). This paper explores possibilities presented by emerging writing assessment technologies to address these fundamental criticisms.

The critique of summative assessment practices has been thoroughly articulated in general terms in recent decades, during which time high-stakes testing has become central to systems of educational accountability. The focal point of much of this criticism has been the kinds of summative testing that have been deployed in schools under the aegis of the No Child Left Behind Act (Ravich, 2010). In this context, the quality of students' learning experiences has been questioned when teachers focus their efforts on test success, particularly for students close to critical score thresholds (Government Accountability Office, 2009; Hollingworth, 2009; Madaus, Russell, & Higgins, 2009; McCarthey, 2008). The impacts have been especially problematic for low performing schools, English Language Learners, and students with disabilities (Harlen & Crick, 2002).

The critics of today's testing regime have noted that most assessment does not feed directly back into learning and that ex post facto judgmental summative assessment is prioritized at the expense of formative assessment of direct relevance to the learner in the learning process (Baker, 2007; Bass & Glaser, 2004; Black & Wiliam, 1998;

* Corresponding author.

E-mail address: billcope@illinois.edu (B. Cope).

Centre for Educational Research & Innovation, 2005; Pellegrino, Chudowsky, & Glaser, 2001). Research shows that “situated assessment” in the form of regular and multiple forms of feedback produces enhanced learning outcomes (Greeno, 1998; Mislevy, 2006b; Palinscar & Brown, 1984; Scardamalia & Bereiter, 1996; Shepard, 2000; Windschitl, 2002).

Critics have also questioned a system that will produce readily computable scores by breaking knowledge up into discrete items, each with an unequivocally correct answer. Indeed, teaching for success in high-stakes, multiple choice, summative tests might mean that at times, we are not teaching and assessing the right things. For example, the International OECD PISA assessment shows that U.S. students fall far behind many comparable countries in tasks that require complex problem-solving skills (*The Forum for Education & Democracy*, 2010). The testing we have today, the critics argue, does not necessarily nurture ways of thinking required for a future of work and community life that will increasingly value creativity, innovation, problem-solving, collaboration, and risk taking (Gee, 2004; Wagner, 2008). Critics have also criticized the focus on in-your-head, individualized knowledge when more and more knowledge is situated, contextual, and socially embedded; they have questioned the value of memorized knowledge when working knowledge can readily be accessed on the fly through today’s information and calculation technologies. Finally, critics have critiqued the accuracy and effectiveness of today’s standardized testing as the basis for educational accountability systems (Ryan & Shepard, 2008). In sum, the case has now been frequently made that much testing has been to the detriment of student learning and to the profession of education.

The 2010 National Educational Technology Plan highlights the need to develop “assessments that give students, educators, and other stakeholders timely and actionable feedback about student learning, “improved” assessment materials and processes for both formative and summative uses,” and new technology-mediated “assessments to engage and motivate learners and to assess complex skills and performances” (*Office of Educational Technology*, 2010, pp. xii-xiii). The Plan’s recommendations are backed by a growing body of research that underlines the importance of formative assessment that feeds directly into the learning process (Baker, 2007; Bass & Glaser, 2004; Black & Wiliam, 1998; Centre for Educational Research & Innovation, 2005; Pellegrino, Chudowsky, & Glaser, 2001).

This paper focuses on the possibility that writing might become a site of assessment not just of writing as a subject area in its own right but also writing across a number of other discipline areas. For instance, whereas multiple-choice summative tests break scientific knowledge into atomized fragments with definitive yes-or-no answers, a science report can show deeper understandings of the nature of scientific evidence, reasoning, and argumentation. Assessing science writing brings us closer to the practices of science and scientists, central among which is the task of reading and writing science. Whereas inferences about knowledge and understanding multiple choice science tests require highly mediated inferences about underlying understanding of science, assessing science writing allows us to assess scientific practice and understanding in a far less mediated way. It is not possible to get closer to science than scientific writing.

Compared to multiple-choice tests, writing in this sense is both an integrative process of “complex performance” (such as presenting evidence, reasoning, and argumentation) and a uniquely powerful window into the substance of disciplinary practice (for instance, by representing the narratives of scientific working hypotheses, literature consulted, theories tested, experiments undertaken, and effects observed). If we can test with writing, we can get closer to the practices and ways of thinking that define a discipline. Complex performance means thinking and acting like a scientist or a historian or a journalist in a context of “situated learning” (Gee, 2004; Latour & Woolgar, 1986). It means working in a “community of practice” (Lave & Wenger, 1991; Wenger, McDermott, & Snyder, 2002). It involves higher order thinking (Etkina, Mestre, & O’Donnell, 2005; Mestre, Ross, Brookes, Smith, & Nokes, 2009). It means presenting a “scientific argument” (Abi-EI-Mona & Abd-EI-Khalick, 2006; Driver, Newton, & Osborne, 2000; Jimenez-Aleixandre & Eurduran, 2008; Osborne, 2005; Toulmin, 2003), a historical interpretation, or a balanced journalistic report. It involves “evidential reasoning” (Cetina, 1999). It requires students to act as disciplinary practitioners, for instance as they “talk science—observing, describing, theorizing, questioning, challenging, arguing, designing experiments, following procedures, judging, evaluating, deciding, concluding, generalizing, reporting (Lemke, 1990 p. ix). It includes learning that addresses “the epistemological bases of the discipline” (Grotzer, 2009 p. 61).

Today’s social writing environments on the World Wide Web are ideal sites for obtaining multiple forms of feedback, including formative and summative assessment information that reflects progress in the development of disciplinary knowledge over time. Teachers and students are increasingly using Web-based writing portfolio spaces such as Wikispaces, PBWiki, WordPress, and Google Apps. These spaces become living, Web-accessible, and assessable records of learning tasks that students have undertaken. However, none of these are designed specifically for educational use; none draws on the insights of the learning sciences for their foundational architecture; and none includes formative

or summative assessment infrastructure to scaffold recognizably reliable and valid assessment. In short, the specific educational potentials of web-based writing technologies have barely been explored.

There is an urgent need to create dedicated educational applications for these Web-based writing environments that integrally incorporate formative and summative assessment. Indeed, there is enormous potential to develop new modes of assessment in which student activity is continuously assessed without disrupting time spent for learning in the classroom. We want to imagine a technology-mediated writing environment of the near future where it will be possible to provide continuous and specific feedback to learners; to use a multimodal format that can accommodate different expressive needs and learning styles; to collect and analyze the work students do both separately and in comparison with each other; to track progress of individuals over time; to compare individual performance to groups; and to track cohorts of students. Eventually, we believe these data may be so thorough and so accurate that they could displace expensive and often problematic summative assessments. In an ideal—and we hope not too distant—future, assessment could be a continuous, seamlessly integrated, and painless part of learning. All assessment would be formative, and summative assessment would simply be a retrospective reading through a progress “dashboard.” These objectives might be read as a counterpoint or corrective to several decades in which summative assessment has been a major priority and teacher and school accountability has been focused through large-scale, standardized summative assessments.

Behind the often heated and at times ideologically gridlocked debate is a genuine challenge to address gaps in achievement between different demographically identifiable groups of students. There is an urgent need to lift whole communities and cohorts of students out of cycles of underachievement. For better or for worse, testing and public reporting of achievement seems to be one of the few tools capable of clearly informing public policy makers, educators, and communities alike about how their resources are being used to expand the life opportunities for learners.

Our stance in this paper is as follows: let's not argue about objectives with even the most vehement of the supporters of mass-institutionalized testing. Let's agree that we need measurement, and a lot of it. Let's agree that we need transparency and that we need accountability when it comes to measuring and reporting learning outcomes. To the extent that we take issue with existing processes and systems of measurement and analysis, it is because they are inadequate to our needs—to the point that at some times and in some respects, these processes and systems become counterproductive. The current system's critics tell us the tests purport to do one thing and do another, and when the tests do that “another” thing, they do it poorly. Inasmuch as this claim is true, our task is to transform fundamentally our systems of educational measurement. That may mean we do more measurement and do it more accurately, perhaps by measuring different things and measuring them differently. We want to make the case that evolving technologies, applied to the discursive and social relations of new writing spaces on the Internet, provide us for the first time with a paradigm-changing opportunity. We have the tools at hand to transform assessment. —and to the extent that we can transform assessment, we may also be able to transform learning.¹

2. Principles: Towards a Renewed Agenda for Writing Assessment and an Expanded Role for Writing in Assessment and Learning

We propose six transformations that will be required if we are to develop a new agenda for learning and assessment in which writing becomes pivotal across a wide range of disciplinary fields and subject matters. These transformations relate to measurable knowledge attributes that are frequently neglected in today's tests. The literature on learning and assessment suggests that we should be measuring these things if our assessments are to be relevant to learners' needs and the nature of knowledge in the contemporary world (Bransford, Brown, & Cocking, 2000; Pellegrino, Chudowsky, & Glaser, 2001).

We phrase each of these transformations as a “should” statement, a principle that should be turned into practice if we are to make our assessments more effective. Meaningful assessment should

- be situated in a knowledge-making practice
- draw explicitly on social cognition
- measure metacognition

¹ This research is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090394 to the University of Illinois at Urbana-Champaign: the “Assess-As-You-Go Writing Assistant”. <<http://assess-as-you-go.com>>.

- address multimodal texts
- be “for learning”, not just “of learning”
- be ubiquitous.

3. Assessment Situated in a Knowledge-making Practice

Deep, disciplinary knowledge is most effectively acquired in contexts that focus on whole, socially authentic, and meaningful tasks in the context of a community of practice (Lave & Wenger, 1991; Wenger, McDermott, & Snyder, 2002). In the real world of knowledge formation or the simulated world of the classroom, this is, for instance, the practice of being a scientist, a news reporter, a social scientist, or a local historian (Gee, 2004). Knowledge is created in social settings of knowledge production. It is about creating kinds of people, types of epistemic stance, and open-ended knowledge-making capabilities.

By comparison, narrower instances of discrete item testing measure reified facts and concepts that can be broken into atomized, component parts with correct/incorrect answers. They generalize about student knowledge from a patchwork of specific empirical and conceptual objects, mere cognitive fragments predetermined by the test-maker to be indicative of a domain of knowledge and deemed by them to be unequivocally correct. Having pronounced upon match/mismatch of knowledge fragments, by fiat of inference, the assessor extrapolates to “understanding” of a knowledge system. Our question to such approaches is, rather than work with extrapolations, why not assess knowledge systems in situ and in action?

Situated assessment has the potential to take the form of regular and multiple forms of feedback, analyzing the ways in which learners capitalize upon the distributed character of knowledge (various sources of knowledge, some of which are at hand and others at a distance), and the formation of peer and broader communities for the construction of knowledge (Greeno, 1998; Mislevy, 2006b; Palinscar & Brown, 1984; Scardamalia & Bereiter, 1996; Shepard, 2000; Windschitl, 2002).

The reason for situated assessment is not just that it should be more reliable; it also has the potential to make assessment more valid for contemporary conditions of knowledge formation and acquisition. For instance, much of the knowledge and many of the human capabilities that are most relevant to today’s workplace, community, and educational settings are instances of what we would call “complex performance”: problem solving, coming to modulated conclusions which are beyond yes/no dichotomies, knowledge that only makes sense when contextualized by broader knowledge and social systems, knowledge created in social collaborations, and innovation or initiative involving calculated risk-taking (Baker, 2004; Mislevy, 2006a; Pellegrino et al., 2001; Wagner, 2008). Complex knowledge performance requires demonstration of a capacity to deploy a range of interconnecting and complementary forms of higher order thinking such as empirical, theoretical, analytical, and applied reasoning (Kalantzis, 2006; Kalantzis & Cope, 2008). A key challenge for educators in this moment of dramatic social change is to develop ways to assess these complex performances (Darling-Hammond, Aness, & Falk, 1995; Darling-Hammond & Wood, 2008) and to assess knowledge in a way that is authentic to the conditions of knowledge creation and learning (Darling-Hammond, Aness, & Falk, 1995; Janesick, 2006).

Writing is an ideal site for the application of the principle of situated assessment of complex performance (with considerable extension of what we mean by “writing” discussed in the multimodality section below). Writing is itself an instance of complex performance. It is also an ideal medium for the representation of many varieties of complex performance—such as reports or essays that describe factual investigations, which persuade through juxtaposing the factual with the interpretative, or which contain the results of empirical scientific inquiry (Lemke, 1990; Martin & Halliday, 1993). In other words, writing can be assessed both for its representational forms and its represented contents. Writing uniquely provides an opportunity for students to express their understanding of bodies of knowledge in a holistic and nuanced way. It is also particularly well suited as a site of formative assessment because detailed, specific feedback can be provided that contributes directly to learning.

What passes for most literacy assessment today is heavily biased to reading because it is more readily assessed through discrete item comprehension tests. By comparison, writing is expensive to assess because it requires slow, human reading and is not necessarily very reliable in the application of an overall grade. Comprehension tests, however, are founded on an unsustainably simplistic, static theory of meaning and representation. They assume that there are intrinsic and indisputable authorial meanings. It is as if the text contains transparent truth, and there is no interpretative role for the reader (Barthes, 1976; Eco, 1981). For this reason, comprehension tests tend to ask about empirical banalities

in a text that can often be easily passed over when reading for meaning (the narrow specifics of who, what, when, where) rather than more significant semantic issues which will in their nature be open to interpretation (such as motivations, purposes, alternatives, interests, lessons). Moreover, by making comprehension a proxy for “literacy,” such tests value a receptive meaning capacity (reading) over a productive meaning capacity (writing). This is an anachronism given the conditions of work, citizenship, and community life in the twenty-first century. Today, we value constructive team contributions over taking orders, creativity and problem solving over compliant operationalization of systems, risk taking and entrepreneurship over line management and bureaucracy.

As we will argue later in this paper, the productive complexities of writing need not be neglected as a site of assessment. In fact, writing is particularly well suited to new, technology-mediated modes of assessment in the era of the social Web.

4. Assessment of Social Cognition

Traditionally, assessment has focused on things in your individual head, or even more narrowly, on things that can be remembered. In every part of the world outside of tests, however, knowledge is always at hand and accessible—in communities of practice, in readily available knowledge sources, and through search and calculation algorithms. Our society also depends on a division of intellectual labor, meaning that although we may become experts in some things, we can never be experts in everything and so habitually rely on others' expertise. This is a deep, structural characteristic of modern societies, a fundamental characteristic of our conditions of epistemic existence that conventional assessments choose to ignore. The social web and ubiquitous connectivity make knowledge even more immediately sociable. Our contemporary communications environment makes the cognitive-epistemic assumptions of traditional tests even stranger. Memory tests have forever been the butt of the old-school folklore about learning by rote and cramming for exams. Today's social web makes this kind of testing anachronistic.

The most basic premises of conventional testing as we have come to know it are peculiar to testing itself: the idea that there is an exteriorized body of reified knowledge, the facts and logics of which can be transferred to memory. In reality, learning involves social communication (learning in a community of practice); it consists of a series of prompts and reference points into social knowledge. This means that what needs to be learned is not the putative content of external knowledge, but how to access and deploy knowledge resources. Today's cognitive science tells us that learning is inextricably social. Knowledge is intrinsically distributed. It is necessarily the result of social interaction (Meltzoff, Kuhl, Movellan, & Sejnowski, 2009). So, assessment should now also attempt to measure this irrepressibly sociable thing—knowledge—and measure it on its own social terms.

Learning is also integrally related to learner identity—you don't learn unless you feel you belong in a learning context where you know you can act as a knowledge maker and that your action will work and be seen to work by your fellows. Unless learning engages your identity, unless its objects are focal to your motivated attention, the learning outcomes will be diminished (Gee, 2004).

Here are some possible social knowledge assessment scenarios: The world is an open book. There is nothing to be memorized, only strategies for access. What we need to assess, then, is how we can locate and use available knowledge resources. Or several students may be under taking a collaborative task, and the assessable outcome may not even be a single script— but we can nevertheless assess differential contributions to a shared, digital document. In these assessment frames, the heroic individual with memory capacities may be good for scoring well at Trivial Pursuit—but this can no longer be our measure of epistemic excellence. Rather than attempting to calculate individual cognitive competence, we will assess collaborative knowledge competence. The focal interest may still be an individual, but the measure of their learning will be how they construct their understandings within socially embedded knowledge ecologies. To assess this, we would allow the test-taker access to as much information as they can get, give them as many algorithmic tools of search and calculation as they can use, and afford them as much opportunity as they need to ask others for “answers” or respond to their requests for assistance. Finally, rather than having every isolated individual on the same page of the same test at the same time, we might have individuals collaborate in knowledge ecologies that work because they are based on the logic of differential and complementary rather than identically templated and replicated expertise.

5. Assessment of Metacognition

Research demonstrates that higher levels of academic performance result from the development of metacognition, the capacity to think about one's thinking processes (Bransford, Brown, & Cocking, 2000; Pellegrino et al., 2001). Alongside situated social cognition, assessment should also have a metacognitive element. Experts in a subject domain typically organize knowledge into schemas and make sense of new information through processes of pattern recognition. Such knowledge representations are useful tools for understanding, knowledge making, and knowledge communication. Furthermore, thinking is more efficient and effective when accompanied by the process of monitoring and reflecting upon one's own thinking (Bransford et al., 2000; Grotzer, 2009; Pellegrino et al., 2001).

Traditional testing measures one's schema-absorbing capacities more than one's schema-forming capabilities. It also measures cognition, bit by bit, at the expense of metacognition or the ability to think holistically about one's thinking. This has always been a deficiency in conventional assessment—its prosaic epistemic narrowness. However, such tests have become even more anachronistic in an era when our new information order provides peculiar tools for navigating the enormous body of available text. In our textual journeys through the social Web we encounter multiple ersatz identifications in the form of icons, links, menus, file names and thumbnails. We work over the algorithmic manifestations of databases, mashups, structured texts, tags, taxonomies, and folksonomies in which no one ever sees the same data represented the same way. The person browsing the social Web is a machine-assisted fabricator and analyst of meanings.

The social Web, in other words, is not just a pile of discoverable information. Users can only navigate their way through its thickets by understanding its architectural principles and by working across layers of meaning and levels of specificity and generality. This is a new cognitive order, some elements of which arise in earlier modern times but which in their intensity and extensiveness require a peculiarly abstracting sensibility. They also demand a new kind of critical literacy (not “comprehension”) in which fact is moderated by dialogues about the status of knowledge and critical discussions of authorial interests (for instance, in the “edit history” pages in wikis or the comments in blogs). Meanings and knowledge are more manifestly modal, contingent, conditional—not that serious knowledge has ever been anything but this, despite the implicit epistemes of tests. It is just less avoidably the case in the era of the social web. We need metacognition to get around in today's textual and knowledge environments, and we need ways to measure this metacognition.

6. Assessment of Multimodal Texts

The nature of what we conventionally understand to be “writing” or “literacy” has been increasingly questioned in recent times. The communications environments in which we now live have transformed the ways in which students communicate: via email, Facebook, Twitter, instant text and image messaging, blogs, videos and the like (Burbules, 2009; Hawisher & Selfe, 2000). These spaces are multimodal (Kress, 2009; Kress, Jewitt, Ogborn, & Tsatsarelis, 2001); in other words, they are places in which text, sound, and still and moving image are integrally related. What we do in schools under the rubric of literacy, and particularly what we measure in our literacy assessments, has not caught up with these profound changes. In many of the subject or knowledge areas that use writing for representation, to use words alone is simply not enough. It is inconceivable, for instance, that a written report of a scientific experiment or a social studies field report could be adequate without reference to a range of tables, diagrams, graphs, images, or audio or video representations. Inexpensive ubiquitous recording technologies and new social media sites for assembling and sharing multimodal texts make this cheap and easy to do.

For these reasons, effective assessment today needs to be in spaces in which learners can represent their knowledge multimodally using a mix of written text, image, audio, and videos of gestural and spatial relations. These fundamental aspects of multimodality are addressed in the multiliteracies theory of our contemporary communications milieu (Cope & Kalantzis, 2000; Cope & Kalantzis, 2009a). The World Wide Web is the first medium of representation and communication to provide such an accessible space for multimodal expression. It is ideally suited for knowledge representation on the part of learners and for assessment of those representations. The result is that curricula and assessments in their traditional formats and media need updating in order to make optimal use of the affordances of these digital spaces, as well as to create learning and assessment environments that are manifestly contemporary in the communicative options they allow.

7. Assessment “For Learning,” Not Just “Of Learning”

The utility of assessment is enhanced if it also provides students and their teachers with specific feedback, not just a holistic score. Attempting to create assessments that play a more constructive role in learning raises the following questions: How can assessments be constructed which make learning goals clearer, assessments that interpret student performance in a way that is more meaningful to learners, teachers, and parents, and that suggest what an individual learner or a group of learners still needs to learn? Also, how can assessment be made more interactive and dynamic (as opposed to static and product-based), providing assistance as part of the assessment process and designed to show learner potential by influencing and helping to change their performance (Black, McCormick, James, & Pedder, 2006; Campione, 1989; Shepard, 2000)? How can assessment fulfill the promise of school reform by connecting more closely to daily curriculum activity with mandated standards (Baker, 2004; Resnick, 2006)?

Researchers have attempted to address these questions in the area of formative assessment (Bass & Glaser, 2004; Black, McCormick, James, & Pedder, 2006; Centre for Educational Research & Innovation, 2005; Frederiksen, 1994; Shepard, 2000). Considerable empirical support exists for the educational effectiveness of assessment closely linked to instruction (Black et al., 2006; Centre for Educational Research & Innovation, 2005; Pellegrino et al., 2001). Black and Wiliam (1998) provide a meta-analysis of over 250 studies of classroom-based formative assessment, concluding that students learn better when they receive systematic feedback on the particular qualities of their work with a view to improving that work. Rapid feedback provides more powerful learning outcomes (Cumming, Wyatt-Smith, Elkins, & Neville, 2006; Lincoln & Neville, 2006; Wyatt-Smith & Cumming, 2003).

Specifically in the case of the assessment of writing, one of the demonstrated benefits of the use of technology in formative assessments is the possibility of immediate or rapid feedback (Baker, 2005). Educators can also use technology-based assessments to make sense of large amounts of data arising from complex performance (Mislevy, 2006a; Mislevy, Steinberg, Breyer, Almond, & Johnson, 2002), bringing to bear psychometric analyses not practically applicable by teachers in a normal classroom situation (Pellegrino et al., 2001). Technology can also create new opportunities for seamlessly integrated assessment in support of self-paced instruction (Chudowsky & Pellegrino, 2003), offering unprecedented possibilities for customizing learning to meet the varied needs of diverse learners (Abell, 2006; Ketterlin-Geller, 2006).

Researchers have also expressed concern about the nature and form of teacher feedback, with or without technology assistance. Much of the research on teachers' responses to writing suggests that written comments are often too global, vague, and inconsistent to help improve writing (Smith, 1997; Straub, 1997). However, written feedback can be helpful in fostering some types of revisions such as adding details, improving coherence, and editing (Conrad & Goldstein, 1999). While content-based feedback can result in higher quality drafts than error correction (Truscott, 1996), too many comments can overwhelm students (Beach & Friedrich, 2006). Students prefer comments about how to improve writing, including global and specific suggestions within supportive contexts (Lee, 2000). Varying comments to address each student's needs and difficulties is an important consideration when writing feedback, and this can be accomplished through teacher conferences and peer response groups (Ferris, 2003).

Because teachers do not have time to provide extensive feedback, peer conferences are a way to engage students in meaningful formative assessment dialogues with each other. However, while peers can provide helpful feedback, they need training in strategies and group processes. Without training, students tend to give negative or highly judgmental comments; trained students, however, can provide constructive feedback resulting in substantive revisions (Simmons, 2003; Straub, 1997). The challenge, then, is to develop new assessment and learning ecologies that harness the lateral energies of peers when the hierarchical knowledge authority of the teacher is spread too thinly to provide adequate, regular, specific, and detailed feedback to learners. The social Web provides an opportunity for this quite profound shift in pedagogical roles, relationships, and responsibilities.

8. Ubiquitous Assessment

We have seen the emergence over the past decade of a phenomenon that is often called “ubiquitous computing”. Laptop computers, mobile phones, game consoles, mobile music players, e-book readers, cameras, and the control boxes that drive television screens all have powerful computing functionalities. These functionalities significantly overlap and integrate from one device to the next. The practical challenge for educators is to create environments in which learning interactions are not confined to the classroom and the school timetable but instead can spill out to

work at home, in the library, and in the community—places that are in fact as ubiquitous as the new devices. This environment has been called “ubiquitous learning” (Cope & Kalantzis, 2009b). Such environments are well suited to support task-focused learning in which formative assessment is integral, making it available anywhere and anytime. This means that students might write in a learning and assessment environment designed to render in a Web browser (including browsers on phones and other mobile devices). Soon we will have interoperable platforms that can render directly to *all* major devices.

We also want to argue that the possibilities for thorough formative assessment in the context of the ubiquitous social Web are such that, some day in the not-too-distant future, we may be able to abolish summative assessments—or at least reduce the distinction between summative and formative assessments to a mere heuristic. Working over the noisy data of machine-mediated feedback, the social Web environments of our near future will be able to build more accurate views of individual student progress over time and capabilities in relation to cohorts, however defined (an individual in relation to a class, a demographic profile, etc.). Our goal should be to make assessment integral to all learning insofar as continuous feedback would be provided and progress tracked. Assessment would then become so pervasive that it all- but disappears. Web-based writing is ideal for the development of such a seamlessly integrated learning and assessment environment.

Indeed, we envision the possibility of abolishing the operational distinction between formative and summative assessment. The things you do to provide formative assessment also create data that can be aggregated and analyzed to provide summative measures. Such assessments would be grounded in more valid assessment data than possible in standalone summative assessments. Conventional summative assessment separates out the test as a peculiar end-of-learning game. In fact, it becomes a game unto itself, quite different in its logic to the processes of everyday knowledge and learning. As a consequence, the curriculum has come to include lessons in how to game the test. This is often called “backward design” (Wiggins & McTighe, 2005) without a hint of irony. All too easily, and despite the best intentions of the creators of this idea, backward design can become a process of playing the system by reverse engineering its winning game-logics. Tests, at least in part, end up testing the student’s test-taking capacities. As we have argued here, when contrasted to the wider world of knowledge making and application, today’s tests are a peculiar game indeed. The institutionalized disjunction between curriculum and its assessment becomes a core difficulty in test validity.

What if we move all assessment data collection into the space of learning? Then all we would ever measure is the substance of learning. There would be no need for ex post facto inferences, however valid or tendentious, because what we are doing is measuring learning itself at its source. And the more data we collect, and the more we view this data through the greatest variety of lenses, the more reliable our assessments will become.

9. Processes: Possibilities for Technology-Mediated Writing Assessment and Learning Represented through Writing

In this final section of our paper, we will analyze six technology-mediated processes for the assessment of writing in which it is possible to bring together formative and summative assessment (See [Appendix A](#)). Such assessments could be of writing itself (as “composition” in language arts or language learning) or of complex disciplinary performance as represented in writing. In other words, our focal point is the assessment both of the medium of writing (including the construction of multimodal texts) and of the disciplinary contents embodied in student production of written and multimodal texts.

We see potential in the following socio-technical domains, most of which are relatively fragmented into separate silos of disciplinary practice and software codebase:

1. Natural Language Analytics
2. Corpus Comparison
3. In-text Network-Mediated Feedback
4. Rubric-Based Review and Rating
5. Semantic Web Processing
6. Survey Psychometrics

First, we need to bring these coding silos closer together. Then we need to find ways to aggregate data drawn from these disparate technologies. There is potential to develop further and extend each of these technologies. Today, they offer us no more than a fraction of what may be possible.

But as a preliminary, how might we understand the potential role of technology as a factor in processes of assessment for learning? The technology to which we are referring has a direct manifestation in the form of server infrastructures and Web software environments along the lines of the recent “cloud computing” metaphor (Reese, 2009) and Web 2.0 architectures (O’Reilly, 2005). Technologies, however, are necessarily and simultaneously social. They are conduits of essentially social knowledge and communicative patterns, such as those that are characteristic of learning and its assessment. Technologies, in other words, are as much conduits of social knowledge as they are hardware and software arrangements in a narrower sense. They are repositories for the storage of human meanings. They embody human logics. They facilitate human interactions. Even in their moments of relative automation, the technologies we discuss here do nothing that is not a reflection and refraction of the social. So, we are just as concerned about change in the form of the social conditions of communication (in this case learning and the assessment of learning) as we are about analyzing the impact of changes in their media, which are its conduits.

10. Natural Language Analytics

We encounter natural language analytics each time we use a spelling or grammar checker in a word processing program. Since the first processing of natural language by computers, attempts have been made to create “automated essay scoring” technologies based on word lists and syntactic rules stored in the computer. These rules are applied to “text-in-isolation.” In other words, they simply apply the rules they have written into them with the text they are requested to process. From this they calculate a score based on the number of errors. They may also attempt to make inferences about generalities of “style.” These rules are reported to have generated statistically significant levels of agreement with human scorers (Valenti, Neri, & Cucchiarelli, 2003). Even in their most sophisticated forms, however, these assessment technologies have come under considerable criticism for their flaws—which we analyze in detail in a companion article in this issue of *Computers and Composition* (Vojak, Kline, Cope, McCarthy, & Kalantzis, 2011).

The key problem with most manifestations of natural language analytics in writing assessment is that they are grounded in a fixed, rule-bound view of language. Their key weakness is a bipolar view of right and wrong linguistic forms. They assume that rules drawn from the standard form of a language are universally applicable. However, language forms and uses vary enormously according to context, including the context-sensitive specificities of register, dialect, field, and tenor, or what Gee calls “social language” (Gee, 1996). We know the limitations of these technologies from our everyday experience of grammar and spell checkers in word processing.

We propose an alternative potential for natural language analytics based on an understanding of the writing process as a fluid, iterative struggle to make meaning. As we write, we are constantly reworking our text. Intrinsically, this has always been the case. In handwriting, we habitually cross things out, add things above the line or in the margins, and pen rewritten drafts another time. We do these constant revisions even more when we use a word processor because changes are so much easier to make and are never at the cost of making the text look messy. We are liberated from this resistant friction, which bedeviled the textual development process in the era of handwriting and typewriters. In this understanding of the writing process, the first iteration of a written sentence is not wrong. It is a first approximation at meaning that we look over and reframe however many times in order to capture our meaning more effectively. Furthermore, when we make a change at one point in a sentence (from a singular to a plural noun, for instance), it may prompt a change that needs to be made elsewhere (a verb agreement, for instance). Before it is changed, it is not an error but rather something we may choose to change as soon as we get a chance. And if we fail to change it, the reason may not be because we don’t understand verb agreement. It may be because we didn’t notice the change that needed to be made and that more crafting is needed. This process of reframing is potentially never-ending; for example, we could keep substituting a word or phraseology that makes our point even better. For the sake of practicality, of course, we do stop at a certain point when we think the text can be considered good enough for the present purposes.

In this conception, there are no errors in language, just things that might always be changed as we struggle to craft our meaning. For this reason, we propose as a basis for computer-mediated natural language analytics a taxonomy of change types. In a formative assessment, these take the form of a question along the lines, “you wrote _____, but could we suggest _____ as a possible change, which is an example of _____ type of change?”. The answer to this question might be “no” because you believe that what is already there works better to express your meaning. As the configurations of

language are infinite and not reducible to rules, the intelligence of this technology-mediated writing environment can become greater as the system is “trained” to know which kinds of change suggestions writers accept more than others. We envision a new generation of these technologies that rely upon an extensible and potentially massive repository of social knowledge of language and the meanings expressed through language.

11. Corpus Comparison

Corpus-based text analysis takes a large set of previously graded essays, all responding to an identical prompt, and compares a new text responding to the same prompt that has not yet been graded. Statistical natural language processing methods (Roth, 2004; Roth & Small, 2006) are used to compare the new text with an existing text. On this basis, it is possible not only to generate a score for a piece of writing but also to provide specific feedback for an initial or final draft. This method utilizes text clustering techniques to compare already-graded texts with the new text (Roussinov & Chen, 1999). The underlying principle of this writing assessment paradigm is that if a new text is similar to a previously human-scored text, it should receive a similar score.

One of the key drawbacks of any corpus comparison algorithm is the high cost of creating a new prompt since a large number of essays must be human-graded before the algorithm can be used. However, in an age when much of what we write is incidentally recorded, there is no reason why Web environments should not anonymously keep all student writing for the purposes of this kind of comparison. One direct parallel is the way in which GMail <<http://www.gmail.com>> stores and mechanically reads your emails in order to serve advertisements to which you are more likely to respond. Another example is the Google machine translation engine. Because Google now has so much translated material stored in two or more languages, it has a vast corpus to support machine translation.

The same could be achieved with student writing. If all writing were stored along with human grades, learning level, and discipline area, the corpus of stored writing could with time provide an increasingly nuanced basis for formative and summative assessment. By these means, social intelligence would be mediated through “machine intelligence.” We believe that corpus comparison techniques will become more powerful both as machine learning algorithms are refined and as larger and larger corpora of texts are collected to which these algorithms can be applied. In this context, a cloud computing environment in which every learner’s work is anonymously stored will become an ever-more powerful resource over time.

12. In-text Network-mediated Feedback

More powerful and lasting performance outcomes are achieved when learning is social, when responsive feedback systems are in place, and where cognition is, in effect, distributed. Using Web-based collaborative writing environments, it should be possible to support joint or team-created work while also tracking and assessing differential inputs. It should also be possible to offer both automated and machine-mediated human feedback. Research shows that responsive feedback has highly positive impacts on learning. It also shows that learning to give useful feedback is a skill that needs to be acquired, and that responsive feedback is as useful to the person learning how to give good feedback as it is to the recipient of the feedback (Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Simmons, 2003; Straub, 1997).

Computer-assisted human feedback systems are primarily designed to facilitate the processes of offering in-text feedback and making change suggestions. We have become familiar with the “track changes” function in word processors, including the “accept change” or “reject change” checking functions. We imagine a highly dialogical parallel conversation to the text, somewhat like the edit pages in Wikipedia. Annotation processes like these might be used as a feedback tool. However, such a tool might also be designed to support quantifiable assessment of the texts being analyzed (Shortis & Burrows, 2009).

Here is one possibility: to create a Web-based annotation tool where a human reader of a text (a teacher, a peer, a parent, an expert, or a critical friend) suggests possible changes and provides an explanation of each suggestion with a taxonomy of change types. The author subsequently accepts or rejects this suggestion. Author-validated change suggestions will then be stored as triggers for prompts pointing to possible changes in subsequent texts. A later writer may be served a prompt to the effect of “someone who wrote a sentence like yours agreed after review to change their sentence in this way: new sentence—old sentence—explanation.” In this way, we could build an infinitely extensible catalog of changes appropriate for something as endlessly complex as language. The more change data we collect, the closer the match to subsequent stretches of text and the more valuable the repository becomes as a reference point.

Such annotated corrections could also be used to complement collections of texts analyzed by corpus comparison. This illustrates the inextricable role of the social component in technology-mediated assessments. Such an annotation environment also supports metacognition. Students are not just analyzing the texts of their peers, annotating them with change suggestions; they may also be expected to provide a reason for the changes suggested. This reason could take the form of unstructured narrative commentary, or selection of an option from a taxonomy of change types.

There is also potential here for role-based weighting of annotations: self (e.g. “think aloud” protocol), peers, teacher, parents or invited expert readers, as well as rate-the-rater weightings offered by authors (how useful or accurate were this reader’s annotations?). In this way, a self-reflexive environment would be created in which students learn to give useful feedback and become more incisive readers of text. The more reflexive the processes of offering in-text feedback, the better its participants become in offering and taking on board such feedback and the more useful the assessment data that is generated.

13. Rubric-based Review and Rating

Computer-assisted review and rating against a rubric assesses the global qualities of a text against criteria, providing qualitative commentary and quantitative scores (Stevens & Levi, 2005; Wilson, 2006). The focus here is on the assessable global qualities of a whole text differentiated by review criteria and level of performance in relation to each criterion.

Web-enabled, rubric-related assessment is commonly used for peer review; however, the same mechanism may apply to self-assessment, teacher assessment, and parent or invited expert assessment. Such applications typically facilitate information sharing, collaboration, and portfolio management as well. In these systems, it is also possible for the reviewers themselves to be rated in order to weight the value of their comments and in time to improve the quality of their reviews (Carlson & Berry, 2003). In this respect, rubric-based writing and assessment environments would be characterized by mediated interaction and distributed social learning. Such an environment has the potential to create “disruptive innovation” (Christensen, Horn, & Johnson, 2008) that will progressively transform the role of the teacher from being a source of definitive knowledge to being a facilitator of a knowledge-producing community.

14. Semantic Web Processing

There has been considerable discussion in recent years about the next stage of the development of what is often called a “semantic web” (Berners-Lee, Hendler, & Ora, 2001). In its more technical manifestations, the Semantic Web speaks an overly formalized language, (RDF/OWL). In its more pervasive reality, a looser form of “semantic publishing” is gradually gaining a wider practical sway in the Web 2.0 world. Users encounter semantic publishing in the form of tagging, folksonomies, and taxonomies as well as “structural and semantic markup” (Cope, Kalantzis, & Magee, 2011).

There are several significant implications of semantic Web processing for student writing and the assessment of that writing. Structural and semantic markup requires that students think about markup, in effect exercising metacognition. Semantic Web processing requires, for example, tagged “paragraph” markers and not just putting in visual indents or line spaces; it also requires the specification of heading levels when it comes to designing an information architecture. It also suggests markup for content, which could be as simple as loosely structured “folksonomy” tags or taking a mind-map connecting these concepts that been created by a task-setting teacher and tagging key concepts and logical connectives as they appear in a text.

The potential of semantic Web processing for assessment has barely been explored. For example, take structural markup on the Web: paragraphs, sectioning, heading weights and so forth. If students use Web markup effectively, they are in reality applying a grammar to the architecture of their texts based on the structural features of those texts. If marked up in this way, textual architectures can be readily assessed. And, to take semantic markup a step further, if a teacher builds a “concept map” (Abd-El-Khalick & BouJaoude, 1997; Novak, 2010) of a content domain (e.g. a map of key concepts explaining how an ecosystem works) or applies a formal schema to student work in that domain (e.g. core chemistry concepts from Chemical Markup Language), the student’s text can be assessed with the assistance of semantic data mining processes. This will be all the more effective if the student has tagged the content of their text.

Concept maps are widely used in a variety of areas of teaching, such as science teaching (Abd-El-Khalick & BouJaoude, 1997; Novak, 2010) and have also been used for assessment (McClure, Sonak, & Suen, 1999; Rice, Ryan, & Samson, 1998). However, embedding concept maps within a technology-intensive and responsive learning environment has great potential to help students develop their scientific argumentation skills. Using probabilistic topic modeling technologies (Girju, Beamer, Rozovskaya, Fister, & Bhat, 2009a; Girju, Nakov, Nastase, Szpakowicz, Turney, & Yuret, 2009b; Paul & Girju, 2010), the potential in this space is to support pre-task diagnostics in which a student creates a concept map relative to the task specification, which is compared against a topic analysis of a canonical corpus such as Wikipedia or open source science journals. It could also support post-task analysis in which the topic map of a student text is compared to a canonical source, both by a human rater (self, peers, expert, teacher) and by automated comparison.

These semantic Web writing practices provide a foundation for more effective machine-supported assessment. For instance, a marked-up text by an author can be compared more reliably against domain or topic-based tag maps. Such maps can be made explicit as the teachers plan learning designs; students undertake their work; and peer reviewers, teachers, and expert examiners review completed student writing. These technologies have the potential to support both formative and summative assessment as well as to encourage powerful modes of metacognition during the learning process.

15. Survey Psychometrics

Although we are arguing for a fundamental shift in orientation toward assessment capable of measuring complex performance, we do envision an ongoing role for item-based surveys, particularly using the increasingly sophisticated methods of computer-adaptive testing (Chang, 2004; Chang & van der Linden, 2003; Chang, Ratnov, & Roth, 2007; Zhang, 2007). However, we suggest that there is a need to return these processes and technologies to their foundations as a variant of survey- psychometrics. Survey-psychometrics can be used to provide another perspective on the knowledge, understandings, perspectives, stances, or attitudes of learners before, during, after, or even parallel to a writing task. Such surveys might provide an additional perspective on background or acquired knowledge related to the complex performance as represented in the written text. They might provide a moderating counterpoint to the data emerging from the other modes of assessment surrounding the construction of the text and help to validate the results of the essay assessment.

Survey psychometric methodologies are used extensively for summative assessment but rarely for diagnostic assessment (Roussos, DiBello, William Stout, Hartz, Henson, & Templin, 2007). We envision Web-based writing environments in which teachers might create pre-writing task knowledge surveys to diagnose levels of student understanding, specifically identifying areas of relative strength and weakness in individual students (there by facilitating individualized instruction) and in cohorts of students. Teachers might also create post-writing task knowledge surveys identifying degrees of progress in acquiring disciplinary and content understandings. Such an environment could support a social network for the development and validation of test items and collections of items addressing a particular topic area. These items would be validated both by human users (teachers and students) and by psychometric statistical analysis specifically aimed to diagnose known cognitive hurdles to understanding (Chatterji, Koh, Choi, & Iyengar, 2009, pp. 6-7). With time, this could become a repository of collective professional intelligence as more and more topic-based collections of items were validated and the range of science topics covered became more finely differentiated.

Such Web environments might also use the most efficient of computer adaptive testing technologies. Unlike fixed-item tests in which the questions and their order are predefined for the examinee, computer adaptive testing selects the items based upon the examinee's previous responses (Chang, 2004; Chang & van der Linden, 2003). As the examinee correctly answers a question, the subsequent question becomes progressively more difficult based upon the difficulty value of items from the test pool. If the examinee answers a question incorrectly, the next item becomes easier. Research has shown that by utilizing computer adaptive testing, the reliability and validity of test scores are equal to or better than non-adaptive tests with twice the number of questions utilizing the same knowledge areas (Weiss & Kingsbury, 1984). Furthermore, in the context of the social Web, it is conceivable that a distributed open bank of surveys and survey items might be created and that the socio-technical system would constantly validate both individual survey items and thematically-linked groups of items.

16. Conclusion

We live in an extraordinary moment of socio-technical change. In some areas of social life, the changes are so profound that they are considered disruptive. The repertoire of socio-technical possibilities described in this paper has the potential to change the way our students learn and the way that we assess their learning. These changes constitute a shift from personal computing (my hard disk, my files) in which otherwise isolated works are tenuously and messily connected via email to a paradigm of socially integral interpersonal computing, which is barely foreshadowed in Facebook, Twitter, and Google Docs. This paradigm elevates writing to a place where it becomes a crucial site of learning and assessment.

The shifts we are experiencing today implicate our most fundamental relationships of representation and communication—writing, broadly conceived. They also offer enormous opportunities for the reform of assessment. We imagine a time in the not-too-distant future when there is no practical distinction between formative and summative assessment. These two forms of assessment will only ever be used to measure the same thing, for learning (as you go) and progress in learning (retrospectively) will converge. If and when this day arrives, there will be no need for the problematic end-of-program, high-stakes, or accountability-defining tests that we have today. Formative assessment will be pervasive—a repertoire of modes of machine-assisted social intelligence that provides continuous feedback to learners in a rapidly responsive manner and with sufficient substantive feedback.

Such a writing environment would generate more than enough data to measure individual learner progress and the performance of cohorts. The process of learning outcome assessment would remain rigorous—potentially more rigorous, in fact, than what one can achieve even with the broadest battery of tests currently available to the education community. We believe this could be a revolutionary step in education. Given our current state-of-the-art computer science and pedagogical understandings, this revolution is for the first time practically achievable.

Appendix A.

The New Assessment: A Matrix of Principles and Practices for Writing Assessment

Principles and Practices	Situated in Knowledge-Making Practice	Social Cognition	Metacognition	Multimodal Texts	For/Of Learning	Ubiquitous
Natural Language Analytics	Responds to a text's specific features with just enough information and just in time.	Environment becomes more socially and contextually intelligent as annotations are collected.	Reviewer annotations and queries to writers prompt metacognitive thinking about the writing and its contents.	Reading tags, captions, labels, descriptions of image, video and audio.	Assists the writer and provides data on their learning progress.	From a writer's point of view, substantially automated, instant responses regarding textual specifics.
Corpus Comparison	On-the-fly comparison of same-discipline, same-level texts.	Environment becomes more accurate as more texts are collected and aligned according to discipline, subject contents and learning level.	Writer provided generalizations from corpus comparison, and the opportunity to rewrite addressing these generalizations.	Reading text ancillary to multimedia objects.	Provides the writer with peer comparison of writing quality in relation to cohorts, including an opportunity to rewrite and reapply comparison.	Automated response regarding overall text in relation to standards and equivalent levels and content areas.

Appendix A (Continued)

Principles and Practices	Situated in Knowledge-Making Practice	Social Cognition	Metacognition	Multimodal Texts	For/Of Learning	Ubiquitous
In-text Network-mediated feedback	Immediate feedback on written work in a knowledge producing community of practice.	Synchronous or asynchronous person-to-person conversations around textual specifics.	Parallel conversation speaks metacognitively about the text form and contents.	Dialogical feedback on non-textual multimedia contents.	Specific feedback and quantification of plus/minus evaluations.	Participants need not be proximate for the around-text dialogue to happen.
Rubric-Based Review and Rating	Establishing an overall frame of reference for the knowledge work.	Review and rating by self, peers, and invited critical friends, thereby creating a social culture of constructive evaluation.	An explicitly defined frame of abstract outcomes criteria in relation to a performance scale.	Review and rating of purely multimedia objects, as well as written and multimodal texts.	Pre- during- and post-task access to rubric, along with the option to rework to address reviewer comments in relation to rubric.	Asynchronous, web-accessible review and rating.
Semantic Web Processing	A subject-specific schema for mapping a knowledge domain.	Collaborative construction of concept maps.	Conscious markup of structure and semantics.	Tagging of images, videos etc. using concept schemas.	Machine and person feedback on the application of concept map to a task.	Asynchronous, web-accessible semantic mapping and markup.
Survey Psychometrics	Task-embedded quizzes, surveys, item-based tests.	Surveys can measure stance, attitudes, and perspective.	Surveys can address underlying knowledge and understandings.	Addresses knowledge acquired from multimodal work.	Before, during, and after surveys to track what students already know and what knowledge they acquire.	Can be taken anywhere, anytime, for example when a task is completed.

Bill Cope is a Research Professor in the Department of Educational Policy Studies at the University of Illinois. He is also Director of Common Ground Publishing, developing internet publishing software for schools and scholarly publications, located in the Research Park at the University of Illinois. His most recent books are *The Future of the Academic Journal*, (with Angus Phillips, eds) Chandos, Oxford, 2009 and *Towards a Semantic Web: Connecting Knowledge in Academic Research*, (with Kalantzis and Magee), Woodhead, Cambridge, 2010. <http://wwcope.com>

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 or editor of: *Multiliteracies: Literacy Learning and the Design of Social Futures*, Routledge, 2000; *New Learning: Elements of a Science of Education*, Cambridge University Press, 2008; *Ubiquitous Learning*, University of Illinois Press, 2009; and *Literacies*, Cambridge University Press, forthcoming 2011. <<http://marykalantzis.com>>

Sarah McCarthy is Professor of Language and Literacy and Associate Head of Graduate Programs in the Department of Curriculum and Instruction at the University of Illinois at Urbana-Champaign. Current research focuses on the role of professional development in writing instruction. She is P. I. on the project “u-learn.net: An anywhere/anytime formative assessment and learning feedback environment” and writing consultant on the “Assess-As-you-Go Project.” She is co-editor (with Paul Prior and Mark Dressman) of the journal, *Research in the Teaching of English* and Co-Director of the University of Illinois Writing Project. mccarthe@illinois.edu

Colleen Vojak is an Adjunct Professor in Education Policy, Organization & Leadership at the University of Illinois, Urbana-Champaign. She is also the project coordinator for the Assess-As-You Go Writing Assistant: A Student Work Environment that Brings Together Formative and Summative Assessment, a U.S. Department of Education grant that involves the development and testing of a new online writing environment. Dr. Vojak’s past research interests have included religion and education; autonomy facilitating curriculum; the influence of market ideology and media on academic integrity; and the stigmatization of social service language. vojak@illinois.edu

Sonia Kline is a doctoral student in the Department of Curriculum and Instruction, Language and Literacy program at the University of Illinois at Urbana-Champaign, specializing in Writing Studies. She is also a graduate research assistant working along side a multidisciplinary team to develop the Assess-As-You-Go Writing Assistant, a web-based working environment for students. Prior to her work at the University of Illinois, Sonia taught

K-8 children, and worked as a technology curriculum manager in schools in Canterbury, Budapest, and New York. Her research interests evolve from points where issues of literacy, learning, and technology converge. kline4@illinois.edu

References

- Abd-El-Khalick, Fouad, & BouJaoude, Saouma. (1997). An exploratory study of the knowledge base for science teaching. *Journal of Research in Science Teaching*, 34, 673–699.
- Abell, Michael. (2006). Individualizing learning using intelligent technology and universally designed curriculum. *The Journal of Technology, Learning, and Assessment*, 5(3), 4–20.
- Abi-El-Mona, Issam, & Abd-El-Khalick, Fouad. (2006). Argumentative discourse in a high school chemistry classroom. *School Science and Mathematics*, 106, 349–361.
- Baker, Eva L. (2004). Aligning curriculum, standards, and assessments: Fulfilling the promise of school reform. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles. Retrieved from <http://www.cse.ucla.edu/products/reports/r645.pdf>
- Baker, Eva L. (2005). Improving accountability models by using technology-enabled knowledge systems (TEKS). National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles. Retrieved from <http://www.cse.ucla.edu/products/reports/r656.pdf>
- Baker, Eva L. (2007). Moving to the next generation system design: Integrating cognition, assessment, and learning. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles. Retrieved from http://www.cse.ucla.edu/products/download_report.asp?r=706
- Barthes, Roland. (1976). *The pleasure of the text*. (Richard Howard, Trans.) London: Johnathon Cape.
- Bass, Kristin M., & Glaser, Robert. (2004). Developing assessments to inform teaching and learning. National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA. Retrieved from http://www.cse.ucla.edu/products/download_report.asp?r=628
- Beach, Richard, & Friedrich, Tom. (2006). Response to writing. In Charles MacArthur, Steve Graham, & Jill Fitzgerald (Eds.), *Handbook of writing research*. New York: Guildford.
- Berners-Lee, Tim, Hendler, James, & Ora, Lassila. (2001 May). The semantic Web. *Scientific American*.
- Black, Paul, McCormick, Robert, James, Mary, & Pedder, David. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education*, 21, 119–132.
- Black, Paul, & Wiliam, Dylan. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Bransford, John D., Brown, Ann L., and Cocking, Rodney R., (2000). How people learn: Brain, mind, experience and school. NRC. Commission on Behavioral and Social Sciences and Education (Ed.) Washington, DC: National Academy Press.
- Burbules, Nicholas C. (2009). Meanings of ubiquitous learning. In BillCope, & Mary Kalantzis (Eds.), *Ubiquitous learning*. Champaign, IL: University of Illinois Press.
- Campione, Joseph C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities*, 22, 151–165.
- Carlson, Patricia, & Berry, Frederick. (2003). Calibrated peer review and assessing learning outcomes. In *33rd ASEE/IEEE Frontiers in Education Conference* Boulder, CO., F3E1–F3E6.
- Centre for Educational Research and Innovation., (2005). Formative assessment: Improving learning in secondary classrooms. Organisation for Economic Co-operation and Development. Paris.
- Cetina, Karin Knorr. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Chang, Hua-Hua. (2004). Understanding computerized adaptive testing: From Robbins-Monro to Lord and beyond. In David Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 117–133). London: Sage Publications.
- Chang, Hua-Hua, & van der Linden, Wim J. (2003). Optimal stratification of item pools in A-stratified adaptive testing. *Applied Psychological Measurement*, 27, 262–274.
- Chang, Ming-Wei, Ratnov, Lev, and Roth, Dan. (2007). Guiding semi-supervision with constraint-driven learning. *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, pp. 280–287. Retrieved from <http://acl.ldc.upenn.edu/P/P07/P07-1036.pdf>
- Chatterji, Madhabi, Koh, Nancy, Choi, Linda, & Iyengar, Radhika. (2009). Proximal assessment for learner diagnosis (PALD): A formative study of teacher practices and early student outcomes. Assessment and Evaluation Research Initiative, Teachers College, Columbia University, New York. Retrieved from http://www.tc.columbia.edu/i/a/document/11486_PALDReportR2web10-09.pdf
- Christensen, Clayton M., Horn, Michael B., & Johnson, Curtis W. (2008). *Disrupting class: How disruptive innovation will change the way the world learns*. New York, NY: McGraw Hill.
- Chudowsky S Naomi, & Pellegrino S James W. (2003). Large-scale assessments that support learning: What will it take? *Theory into Practice*, 42, 75–83.
- Conrad, Susan M., & Goldstein, Lynn M. (1999). ESL student revision after teacher-written comments: Text, context, and individuals. *Journal of Second Language Writing*, 8, 257–276.
- Cope, Bill, & Kalantzis, Mary. (2000). *Multiliteracies: Literacy learning and the design of social futures*. London: Routledge.
- Cope, Bill, & Kalantzis, Mary. (2009a). Multiliteracies: New literacies, new learning. *Pedagogies: An International Journal*, 4, 164–195.
- Cope, Bill, & Kalantzis, Mary. (2009b). Ubiquitous learning: An agenda for educational transformation. In Bill Cope, & Mary Kalantzis (Eds.), *Ubiquitous learning*. Champaign, IL: University of Illinois Press.
- Cope, Bill, Kalantzis, Mary, & Magee, Liam. (2011). *Towards a semantic Web: Connecting knowledge in academic research*. Cambridge, UK: Woodhead Publishing.

- Cumming, Joy, Wyatt-Smith, Claire, Elkins, John, & Neville, Mary. (2006). Teacher judgment: Building an evidentiary base for quality literacy and numeracy education. Centre for Applied Language, Literacy, and Communication Studies, Griffith University, Brisbane. Retrieved from http://www.qsa.qld.edu.au/downloads/publications/research_qsa_teacher_judgment.pdf
- Darling-Hammond, Linda, Ancess, Jacqueline, & Falk, Beverly. (1995). *Authentic assessment in action: Studies in schools and students at work*. New York: Teachers College Press.
- Darling-Hammond, Linda, & Wood, George H. (2008). Assessment for the 21st century: Using performance assessments to measure student learning more effectively. *The Forum for Education and Democracy*.
- Driver, Rosalind, Newton, Paul, & Osborne, Jonathon. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287–312.
- Eco, Umberto. (1981). *The role of the reader: Explorations in the semiotics of texts*. London: Hutchinson.
- Etkina, Eugenia, Mestre, Jose P., & O'Donnell, Angela. (2005). The impact of the cognitive revolution on science learning and teaching. In James Royer (Ed.), *The impact of the cognitive revolution on educational psychology*. Greenwich, CT: Information Age Publishing.
- Ferris, Dana R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Erlbaum.
- Frederiksen, Norman. (1994). The integration of testing with teaching: Applications of cognitive psychology in instruction. *American Journal of Education*, 102, 527–564.
- Gee, James Paul. (1996). *Social linguistics and literacies: Ideology in discourses*. London: Taylor and Francis.
- Gee, James Paul. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.
- Girju, Roxana, Beamer, Brandon, Rozovskaya, Alla, Fister, Andrew, & Bhat, Suma. (2009). A knowledge-rich approach to identifying semantic relations between nominals. *Information Processing and Management Journal*, 46(5), 589–610.
- Girju, Roxana, Nakov, Presley, Nastase, Vivi, Szpakowicz, Span, Turney, Peter, & Yuret, Denis. (2009). Classification of semantic relations between nominals. *Journal of Language Resources and Evaluation*, 42, 105–121.
- Government Accountability Office. (2009). Student achievement: Schools use multiple strategies to help students meet academic standards, especially schools with higher proportions of low-income and minority students. United States Government Accountability Office. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED507150>
- Greeno, James G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53, 5–26.
- Grotzer, Tina. (2009). Learning to reason about evidence and explanations: Promising directions in education. In Eamonn C. Allan, Tina Grotzer, Richard E. Jerome Kagan, David N. Nisbett, Perkins, David N., S. Lee, & Shulman (Eds.), *Education and a Civil Society: Teaching Evidence-Based Decision Making* (pp. 51–74). Washington, DC: American Academy of Arts and Sciences.
- Harlen, Wynne, & Crick, Ruth Deakin. (2002). *A systematic review of the impact of summative assessment on students motivation for learning*. London, UK: Evidence for Policy and Practice Information and Co-ordinating Centre.
- Hawisher, Gail, & Selfe, Cynthia. (2000). *Global literacies and the World Wide Web*. London: Routledge.
- Hollingworth, Liz. (2009). Unintended educational and social consequences of the No Child Left Behind Act. *Journal of Gender, Race and Justice*, 12(2), 311–328.
- Janesick, Valerie J. (2006). *Authentic assessment*. New York: Peter Lang.
- Jimenez-Aleixandre, Maria Pilar, & Eurduran, Sibel. (2008). Argumentation in science education: An overview. In Sibel Eurduran, & Maria Pilar Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research*. Amsterdam: Springer.
- Kalantzis, Mary. (2006). Elements of a science of education. *Australian Educational Researcher*, 33, 15–42.
- Kalantzis, Mary, & Cope, Bill. (2008). *New learning: Elements of a science of education*. Cambridge, UK: Cambridge University Press.
- Ketterlin-Geller, Leanne R. (2006). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning, and Assessment*, 4 (2). Retrieved from <http://www.jtla.org>
- Kress, Gunther. (2009). *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.
- Kress, Gunther, Jewitt, Carey, Ogborn, Jon, & Tsatsarelis, Charalampos. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London: Continuum.
- Latour, Bruno, & Woolgar, Steve. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Lave, Jean, & Wenger, Etienne. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lee, Amy. (2000). *Composing critical pedagogies: Teaching writing as revision*. Urbana, IL: National Council of Teachers of English.
- Lemke, Jay L. (1990). *Talking science: Language, learning and values*. Westport, CT: Ablex.
- Lincoln, Mary, & Neville, Mary. (2006). Middle phase of learning cluster project. Final Report 2006. Curriculum Branch, Education Queensland, Brisbane.
- Madaus, George, Russell, Michael, & Higgins, Jennifer. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, principals, schools and society*. Charlotte, NC: Information Age Publishing.
- Martin, J. R., & Halliday, M. A. K. (1993). *Writing science*. London: Falmer Press.
- McCarthy, Sarah J. (2008). The impact of No Child Left Behind on teachers writing instruction. *Written Communication*, 25, 462–505.
- McClure, John R., Sonak, Brian, & Suen, Hoi K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logical practicality. *Journal of Research in Science Teaching*, 36, 475–492.
- Meltzoff, Andrew N., Kuhl, Patricia K., Movellan, Javie, & Sejnowski, Terrence J. (2009). Foundations for a new science of learning. *Science*, 32(5), 284–288.
- Mestre, Jose P., Ross, Brian H., Brookes, David T., Smith, Adam D., & Nokes, Timothy J. (2009). How cognitive science can promote conceptual understanding in physics classrooms. In Issa M. Saleh, & Myint Swe Khine (Eds.), *Fostering scientific habits of mind: Pedagogical knowledge and best practices in science education* (pp. 145–171). Rotterdam, The Netherlands: Sense Publishers.
- Mislevy, Robert J. (2006). Cognitive psychology and educational assessment. In Robert L. Brennan (Ed.), *Educational Measurement* (pp. 257–305). New York: Praeger.

- Mislevy, Robert J. (2006b). Issues of structure and issues of scale in assessment from a situative/sociocultural perspective. Center for the Study of Evaluation National Center for Research on Evaluation, University of California, Los Angeles. Retrieved from http://www.cse.ucla.edu/products/download_report.asp?r=778
- Mislevy, Robert J., Steinberg, Linda S., Breyer, F. Jay, Almond, Russell, & Johnson, Lynn G. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, 14, 363–389.
- Novak, Joseph D. (2010). *Learning, creating and using knowledge: Concept maps as facilitative tools in schools and corporations*. New York: Routledge.
- O'Reilly, Tim. (2005). What is Web 2. 0? Design patterns and business models for the next generation of software. Retrieved from <http://oreilly.com/web2/archive/what-is-web-2.0.html>
- Office of Educational Technology. (2010). Transforming American education: Learning powered by technology. US Department of Education, Washington, DC. Retrieved from <http://www.ed.gov/technology/netp-2010>
- Osborne, Jonathon. (2005). The role of argumentation in science education. In Kerst Boersma (Ed.), *Research and the quality of science education* (pp. 367–380). Amsterdam: Springer.
- Palinscar, Aannemarie Sullivan, & Brown, Ann L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175.
- Paul, Michael, & Girju, Roxana. (2010). A two-dimensional topic-aspect model for discovering multi-faceted topics. In *Proceedings of the Annual Conference on Artificial Intelligence (AAAI)*.
- Pellegrino, James W., Chudowsky, Naomi, & Glaser, Robert. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.
- Ravich, Diane. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.
- Reese, George. (2009). *Cloud application architectures: Building applications and infrastructure in the cloud*. Sebastopol, CA: O'Reilly.
- Resnick, Lauren B. (2006). Making accountability really count. *Educational Measurement, Issues, & Practice*, 25, 33–37.
- Rice, Diana, Ryan, Joseph, & Samson, Sara. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching*, 35, 1103–1127.
- Roth, Dan. (2004). Learning based programming. In Lakhmi Jain and Dawn Holmes. In *Innovations in Machine Learning: Theory and Applications*. Berlin: Springer-Verlag., pp. 73–98.
- Roth, Dan, & Small, Kevin. (2006). Margin-based active learning for structured output spaces. In *Proceedings of the European Conference of Machine Learning (ECML)* (pp. 413–424).
- Roussinov, Dmitri G., & Chen, Hsinchun. (1999). Document clustering for electronic meetings: An experimental comparison of two techniques. *Decision Support Systems*, 27, 67–79.
- Roussos, Louis A., DiBello, Louis V., Stout S William S., Hartz, Sarah M., Henson, Robert A., & Templin S Jonathan H. (2007). The fusion model skills diagnosis system. In Jacqueline Leighton, & Mike Gierl (Eds.), *Cognitive diagnostic assessment in education*. New York: Cambridge University Press.
- Ryan, Katherine E., & Shepard, Lorrie A. (2008). *The future of test-based accountability*. New York: Routledge.
- Scardamalia, Marlene, & Bereiter, Carl. (1996). Student communities for the advancement of knowledge. *Communications of the ACM*, 39, 36–37.
- Shepard, Lorrie A. (2000). The role of classroom assessment in teaching and learning. National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles. Retrieved from http://www.cse.ucla.edu/products/download_report.asp?r=517
- Shortis, Mark, & Burrows, Steven. (2009). A review of the status of online, semi-automated marking and feedback systems. In *ATN Assessment Conference 2009*. Melbourne: RMIT University.
- Simmons, Jay. (2003). Responders are taught, not born. *Journal of Adolescent and Adult Literacy*, 46, 693–784.
- Smith, Summer. (1997). The genre of the end comment: Conventions in teacher response to student writing. *College Composition and Communication*, 33, 148–156.
- Stevens, Dannelle D., & Levi S Antonia J. (2005). *Introduction to rubrics*. Sterling, VA: Stylus.
- Straub, Richard. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English*, 31, 91–119.
- The Forum for Education and Democracy. (2010). *Creating a national culture of learning: Recommendations for the reauthorization of ESEA*. Washington: The Forum for Education and Democracy.
- Toulmin, Stephen. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Truscott, John. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369.
- Valenti, Semire, Neri, Francesca, & Cucchiarelli, Allesandro. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education*, 2, 319–330.
- Vojak, Colleen, Kline, Sonia, Cope, Bill, McCarthey, Sarah, & Kalantzis, Mary. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition*, 28(2), this issue.
- Wagner, Tony. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need, and what we can do about it*. New York: Basic Books.
- Weiss, David J., & Kingsbury, G. Gage. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361–375.
- Wenger, Etienne, McDermott, Richard, & Snyder, William M. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Cambridge, MA: Harvard Business School Press.
- Wiggins, Grant, & McTighe, Jay. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wilson, Maja. (2006). *Rethinking rubrics in writing assessment*. Portsmouth, NH: Heinemann.

- Windschitl, Mark. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72, 131–175.
- Wyatt-Smith, Claire Maree, & Cumming, Jacqueline Joy. (2003). Curriculum literacies: Expanding domains of assessment. *Assessment in Education*, 10, 47–59.
- Zhang, Jinming. (2007). Conditional covariance theory and detection for polytomous items. *Psychometrika*, 72, 69–91.

Further reading

- Page, Ellis B. (1967). *Statistical and linguistic strategies in the computer grading of essays*. Storrs, CT: University of Connecticut Press.