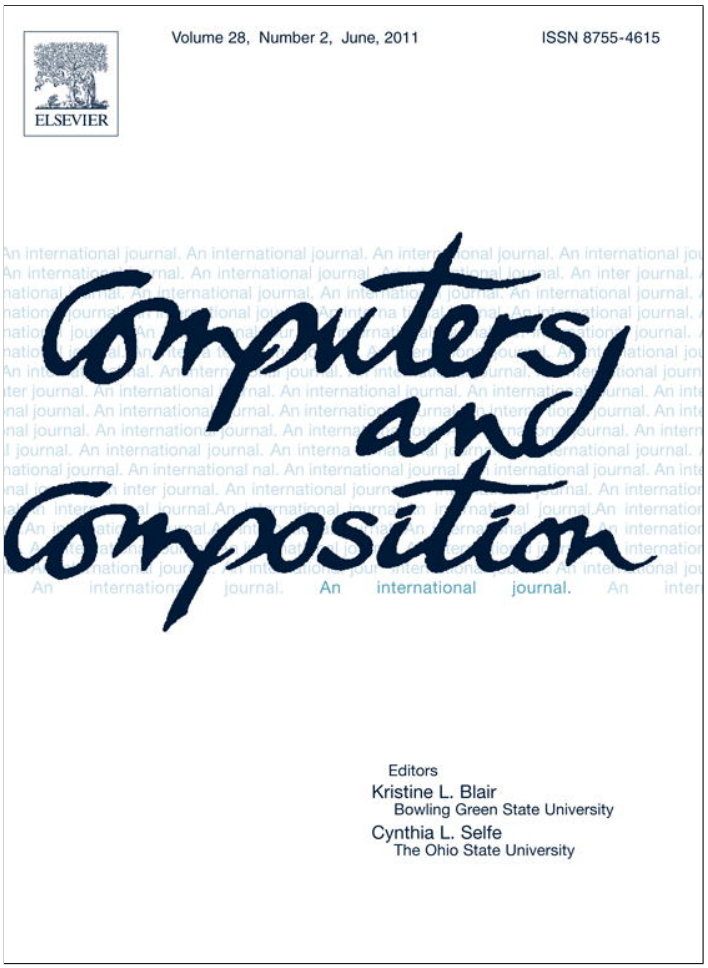

Available online at www.sciencedirect.com
ScienceDirect
Computers and Composition 28 (2011) 97–111
Computers and Composition
www.elsevier.com/locate/compcom

New Spaces and Old Places: An Analysis of Writing Assessment Software

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

University of Illinois, Urbana-Champaign, IL, USA

Abstract

This article examines the strengths and weaknesses of emerging writing assessment technologies. Instead of providing a comprehensive review of each program, we take a deliberately selective approach using three key understandings about writing as a framework for analysis: writing is a socially situated activity; writing is functionally and formally diverse; and writing is a meaning-making activity that can be conveyed in multiple modalities. We conclude that the programs available today largely neglect the potential of emerging technologies to promote a broader vision of writing. Instead, they tend to align with the narrow view of writing dominant in a more recent era of testing and accountability, a view that is increasingly thrown into question. New technologies, we conclude, are for the most part being used to reinforce old practices. At a time when computer technology is increasingly looked to as a way to improve assessment, these findings have important implications.

Keywords: Writing; Assessment; Technology; Learner feedback; Pedagogy

1. Introduction

Assessment has a powerful influence on what is valued, what is taught, and ultimately what is learned in our formal sites of institutional learning (Hillocks, 2002). Those who develop assessments, consequently, exert enormous influence and—for better and sometimes for worse—assume great responsibility. One would hope that assessments used in our schools promote the kinds of teaching and learning most beneficial to students. Clearly, however, other considerations also influence the development and use of assessments (Lemann, 1999).

In the United States, technologies have always played a major part in shaping what is assessed and how assessment is carried out (Huot & Neal, 2006; Newkirk, 2009). The invention of machine scoring of multiple choice test papers profoundly influenced testing practices to the extent that they have contributed to the dominance of this mode of assessment. Consequently, atomized facts are frequently tested to determine disciplinary subject knowledge. In the domain of literacy, reading tends to be assessed as a proxy because comprehension can be readily tested using multiple-choice questions whereas writing involves more complex and labor-intensive human assessment procedures. Applying computer technologies to increase the efficiency of writing assessment is therefore an attractive proposition that might facilitate new ways to assess disciplinary subject matter embodied in writing, such as a science report or a history essay, which may in turn help redress the imbalance in literary assessment.

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

8755-4615/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.compcom.2011.04.004

Starting in the 1960s with the now defunct Project Essay Grader (PEG) and continuing into the 21st century with technologies such as the Intelligent Essay Assessor (Landauer, Foltz, & Laham, 1998) and Intellimetric (Elliot, 2001), a number of computer programs have sought to emulate human grading of student essays. These programs have been shown to demonstrate a high degree of reliability (Shermis & Burstein, 2003); the scores that they generate closely align with scores assigned by human graders and are also praised as cost and time efficient (Dikli, 2006). More recently, in keeping with research on effective formative or diagnostic assessment (Black & Wiliam, 1998), automated essay graders not only provide a summative holistic score but also attempt to offer feedback for learning on various components of writing, such as focus and meaning; content and development; organization; language usage; voice and style; and mechanics and conventions. In addition to providing automated assessment, many computer programs offer technology-mediated assessment features that facilitate human feedback.

Despite these developments, many in the field of writing instruction are still cautious—with good reason—about the use of technology in the assessment of writing. In the view of the Conference on College Composition and Communication (CCCC), 2004, outlined in their position statement *Teaching, Learning, and Assessing Writing in Digital Environments* (2004), the advantage of the speed of machine-scoring is far outweighed by disadvantages. Writing to a machine, they argue, "violates the social nature of writing" and is thus detrimental to students' understanding of writing as a human form of communication. In addition, the CCCC questions the validity of machine scoring and the criteria on which machine scoring is based; the CCCC raises concerns that machine scoring of college papers will lead to high schools preparing students to write for machines rather than for human audiences. In the conclusion of the CCCC position statement, its view is explicit: "We oppose the use of machine-scored writing in the assessment of writing" (2004, para. 13).

Other educators such as Patricia Ericsson and Richard Haswell (2006) who edited the book *Machine Scoring of Student Essays,* stress the importance of educators remaining part of the dialogue. In their volume, which is the most comprehensive response to date from educators, key issues are considered. The validity of automated essay assessment is questioned; claims that machines can actually understand the meaning of text are refuted; and the dangers of students writing without legitimate human audience or purpose are discussed. These issues are highly pertinent as the government prepares to overhaul state assessments with their Race To The Top Assessment (RTTTA) Program, a shared vision for the two remaining RTTTA consortia who are competing for $350 million of federal funds is to "aggressively pursue technology-based solutions for more efficient delivery and scoring of state assessments" (National Governors Association & Council of Chief State School Officers, 2010, p. 3).

Over a period of a year, our research team conducted an extensive survey and evaluation of existing computer-based programs used to assess writing.[1] Our definition of assessment is broad, including all feedback ranging from summative scoring to specific qualitative comments that might support formative assessment. We not only examined computer programs that automatically assess writing but also technology-mediated writing assessment programs that facilitate teacher and peer response. In fact, we found that many programs offered both automated and technology-mediated assessment features. This finding is reflected in our discussion, which is much broader than previous research that tends to focus only on the automated component of computerized writing assessment. Our perspective consequently provides a fuller picture of the advantages and pitfalls of involving technology in the assessment of writing.

Our research also employes a broad definition of writing. In seeking to better understand the nature of writing, we draw on multiple theoretical frames. Our perspective is informed by socio-cultural theorists (Bakhtin, 1986; Vygotsky, 1978) and other theories and research that define literacy (literacies) in expansive terms: Multiliteracies (Cope & Kalantzis, 2000, 2009; The New London Group, 1996), New Literacy Studies (Heath, 1983; Street, 1984), and Social Semiotic theory (Halliday, 1978; Hodge & Kress, 1988; Kress, 2009). Based on these multiple perspectives, we identify three key understandings that provide the foundation for our analysis:

1) Writing is a socially situated activity.
2) Writing is functionally and formally diverse.
3) Writing is a meaning-making activity that can be conveyed in multiple modalities.

[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>.

We acknowledge that the creators of the writing assessment applications may not share our socially situated, diverse, and multimodal theories of writing. However, we maintain that these key understandings have been in the literature and informed practices for the past 30 years (Bazerman, 2008) and that they are essential for moving assessment practices into the 21st century.

This article examines the strengths and weaknesses of emerging writing assessment technologies. Our objective is not to provide a comprehensive review of each writing assessment program nor to debate whether technology should be used in the assessment of student writing. Rather, we take a deliberately selective approach, employing our three stated understandings about writing as a framework for review. First we delineate these key understandings and discuss the potential they hold for the use of technology in writing. We then outline our methods used to investigate existing computer-based programs. Next, in the main body of this article, we articulate our findings. We conclude that these programs largely neglect the potential of emerging technologies to promote a broader vision of writing. Instead, they tend to align with the narrow view of writing that was dominant in the more recent era of testing and accountability, a view that is increasingly thrown into question. New technologies, we conclude, are for the most part being used to reinforce old practices. In a companion paper (Cope et. al, this issue), we discuss possible ways forward given the technological and pedagogical developments in computer-mediated writing assessment.

2. Key Understandings

2.1. Writing is a socially situated activity

The stereotypical image of a writer is of the individual composing in isolation, "the solitary act of an atomistic writer who aims to produce a text" (LeFevre, 1987, p. 15). Karen LeFevre demonstrates that this view is limited and that its predominance has restrictive implications for the teaching and learning of writing—most notably, the common failure of educational institutions to consider fully the important social aspects of writing. LeFevre proposes a more comprehensive understanding of rhetorical invention as a social act. She believes that "to continue to neglect the ways inventors collaborate and the ways collectives help or hinder invention would be to settle for a limited view—an unfinished sketch, a look through a jammed kaleidoscope—of what happens when writers invent" (p. 94).

Similarly, Deborah Brandt (1990) puts social involvement at the heart of literate development. She argues that literacy is highly metacommunicative; in essence, it is about knowing what to do in the here and now to keep meaning from breaking down. From this perspective, literacy learning hinges on crucial knowledge about how people *do the work of* making sense of text. Brandt explains that this is why "literacy learning requires not merely ample experience with print but ample access to other people who read and write and will show you why and how to do it" (p. 6). Brandt posits that literacy is actually a hyper-social act—even when a writer composes alone, the writer must understand how an audience will make meaning from the text.

The notion of "community of practice" (Lave & Wenger, 1991; Wenger, 1998) emphasizes the socially situated nature of learning and acknowledges informal group networks (Barton & Tusting, 2005). Traditional writing classrooms tend to reduce audiences to one—the assessing teacher—and assessment to textual formalism. New technologies, however, provide great potential for connecting writing to situated learning experiences both in and out of the classroom. They comfortably support collaborative writing, commenting and editing by peers, and writing to portfolios, that can be read and evaluated by peers, parents, and critical friends as well as teachers. James Gee's (2005) vision of "affinity spaces" aligns even more closely with the complex and socially situated nature of writing made possible by new sites and modes of writing such as blogs, wikis, and social networking sites. "Affinity spaces" feature a collective communication endeavor, many different forms of and routes to participation, and status swapping where leadership is porous. Affinity spaces also support the intensive and extensive development of distributed, collective knowledge. Gee points out that young people are already highly familiar with affinity spaces out of school and argues that "the notion of affinity spaces can lead us to ask some new questions about classroom learning or ask some old ones in new ways" (p. 232). New technologies provide great potential to think in new ways about writing and writing assessment as a more social and collaborative learning process than the traditional notions and practices of the isolated, individual writer.

2.2. Writing is functionally and formally diverse

Language and life are intrinsically linked and connected to "spheres of human activity" (Bakhtin, 1986). Human activity shapes the functions and forms of texts but is to a large extent invisible in finished work. By the time this article has been finished, it has been shaped by various communicative events (Hymes, 1972): emails, meetings, cut paragraphs, texts scrawled in the middle of the night. During this process, the complex genre of the journal article for instance, in Bakhtin's words, has "absorbed," "digested," and altered primary genres such as a note or a conversation. Genres of writing, then, are not merely specific forms. As Charles Bazerman (2004) states, "genres are part of the way that humans give shape to social activity" (p. 317). Paul Prior (2009) similarly alludes that the "rhizomatic threads of genre spread just about everywhere we might look into human societies" (p. 18). Although experienced writers may recognize some of their influences and acknowledge those experiences and people who make their writing successful, inexperienced writers are less likely to do so. Traditional approaches to writing that equate genre with a collection of text features make it appear that these features are what is most significant in the text rather than whether the text accomplishes its socially intended function (Bazerman, 2004). In the view we present here, genre is a social process incorporating many textual moves and social interactions and producing varied communicative outcomes depending on the interests and identity-position of writers and readers (Cope & Kalantzis, 1993).

2.3. Writing is a meaning-making activity, and meaning can be conveyed in multiple modalities

Meaning-making practices have always been inherently multimodal; however, computer technologies have added a new dimension to this fact. New media have transformed the way in which students communicate (Cope & Kalantzis, 2009). These spaces are multimodal and include integrally related text, sound, and still and moving images (Kress, 2009; Kress, Jewitt, Ogborn & Tsatarelis, 2001). Jay Lemke (2002) distinguishes between multimodality and hypermodality, showing that hypermodality is much more than the juxtaposition of text, sound, and image. Hypermodality, according to Lemke, is "the conflation of multimodality and hypertextality" where we not only have "linkages among text units of various scales" but also "linkages among text units, visual units and sound units. And these go beyond the default conventions of traditional multimodal genres" (p. 301). Glynda Hull and Mark Nelson (2005), drawing on a case study of a community digital story project, argue that through multimodal composition in a computer environment, "the meaning that a viewer or listener experiences is qualitatively different, transcending what is possible via each mode separately" (p. 251).

When every child is required to work through a similar process that leads to similar textual products, setting standards for teaching and assessing is relatively easy. Teaching and assessing become far more complex, however, when students are expected to make a broad range of design decisions intended to provide the best way to convey meaning in a particular situation (Shipka, 2009). For instance, an excellent science report may require a mix of text, images, diagrams, tables, and video. How is this to be assessed? The drive for writing standards in schools, unfortunately, has often been interpreted as the need for standardization; consequently, students' meaning-making opportunities are often narrowly restricted to the formal elements of written text.

2.4. Investigation Methods

With these writing principles in mind, we examined computer applications that are currently used to assess student writing. The applications we analyzed were varied in terms of their modes of assessment and underlying technologies (see Appendix A). We examined programs such as *MY Access!*, *Criterion*, and *WriteToLearn* that relied heavily on automated assessment founded in the technologies of natural language processing. Other programs we examined, such as *Calibrated Peer Review*, *Choices,* and DiGIMS Online, provide technology-mediated writing assessment primarily by relying on human feedback with computer-assisted support.

We were not able to gain full access to every program; sometimes we were only able to try out a demonstration. Whenever possible, however, we tried and evaluated them from both the student and teacher perspective. The programs were also in various stages of development: some were still in active trials, others had been in use for several years, and still others appeared to be important in the history of automated essay scoring but were no longer functional (*Project Essay Grade*). The programs also ran the gamut from being totally free (*Calibrated Peer Review* and *Bayesian Essay Test Scoring System*) to being marketed intensively for profit. In an effort to survey the programs systematically,

we developed a chart of features and attempted to describe and assess the programs according to these features (see Appendix B). Whenever possible we scoured program websites, watched demonstrations, analyzed feature sets, and inputted sample texts for evaluation. We looked at the claims made by the programs and then evaluated how well they did what they purported to do. We also looked at available research reporting upon the programs in use, although for the vast majority of products there was none available—or of that which was available, we found that much was conducted in house by the companies who had developed and owned the technologies (Wang & Brown, 2007). There are, however, a few instances where independent research was conducted, which provided additional information on the effectiveness of these programs.

As we reviewed the programs, the following questions derived from our key understandings about writing guided our examination:

1. To what extent do these programs align with our understanding that writing is a socially situated activity? To answer this question, we examined the nature of *interaction and feedback* within these programs.
2. To what extent do these programs recognize that writing is functionally and formally diverse? In responding to this question, we considered how these programs represent *function, form, and genre*.
3. To what extent do these programs recognize that writing in the new media age increasingly and productively supports multimodality as a meaning-making activity and that meaning can be conveyed in multiple modes of representation and communication? To address these questions, we reviewed these programs to identify their *multimodal meaning-making* possibilities.

Our findings are broadly organized under these three sections of *interaction and feedback; function, form, and genre; and multimodal meaning-making* although in many cases these topics, and consequently the sections, overlap.

3. Findings

3.1. Interaction and feedback

The programs investigated provide feedback in response to student writing in various forms (quantitative and qualitative) and degrees of specificity. Feedback is only of direct value to students, however, if it enables them to work to improve their writing. It should promote reciprocal action or be interactive in nature. Feedback should also be situated within the social and generic context in terms of form and function and should connect writer with audience. Some of the computer-based writing programs we explored make an effort to connect students to real readers, such as peers, and deliver human-generated feedback; other programs primarily offer computer-generated feedback that can be monitored and supplemented by the teacher.

One example of a tool designed specifically to generate peer interaction is *Calibrated Peer Review* (CPR). This free program was created primarily for use in large lecture halls at the post-secondary level, although it is also used in high schools and could possibly be adapted for use in lower grades. The program's purpose is to encourage students to develop writing and assessment skills in a venue in which scoring essays is usually not practical due to the large volume of student-generated writing. The CPR assignment begins with a specific course-content-related prompt and links to additional online resources that students may use in conjunction with assigned readings. The CPR library currently contains several hundred assignments with prompts and accompanying calibration materials. Interestingly, most of these assignments are on science- and math-related topics. Students are constrained by not being able to choose the topic of their writing assignment; however, they are given the opportunity to develop writing skills in disciplines that are not usually noted for presenting topic choice opportunities.

Students using CPR anonymously score and comment on peer essays and then, later, score and comment on their own essay using a common rubric that contains items pertaining to writing form, style, and content. Before conducting the peer and self-review, the student is trained (or calibrated) by scoring three sample essays. Students who prove to be well calibrated are given more weight in the peer and self-review process than those who are less well calibrated. Student scores are comprised of four elements: the composite peer evaluations of the student's text, the calibration process, the student's evaluation of three peer texts, and the self-review. The more accurately the student assesses his or her own essay (judged by weighted peer reviews), the higher the self-assessment score; students are therefore rewarded for honesty (Nickle, 2006, p. 331). By the time students have completed an assignment, they have reviewed

three classmates' texts and have also been reviewed by three classmates, providing several opportunities for sharing evaluative comments with real audiences.

Choices is another online program that integrates human interaction with writing instruction. This program provides options for peer and teacher feedback during every step of the writing process, from initial brainstorming of ideas with classmates through a blog to peer and teacher reviews up through the final draft. Each student has a private workspace and a public page on which students may choose to share their writing. *Choices* presents students with six large writing assignments that tie into a composition textbook. Although these assignments prescribe the genre and provide general parameters, students are free to choose their own topics.

Other programs embed the basic tools for peer review, teacher review, or both within a larger program. For example, *MyCompLab* provides peer review as part of a broader palette of options, including diagnostic tests, remedial exercises, computerized assessment tools, and resources. *DiGIMS* online, on the other hand, relies entirely on instructor review, providing a way for instructors to give direct feedback to students and manage grading and record keeping while also aligning their writing curriculum with state standards-based rubrics. Although *DiGIMS* online has the advantages of direct teacher feedback and connection to state standards, it places considerable emphasis on the more technical aspects of writing and constrains students' work by adherence to highly structured rubrics used in the evaluation process. *DiGIMS* also does not require students to submit papers electronically, and therefore the feedback is not physically tied to the student text (as it would be with a text editor or a tracking system). Teacher comments appear on a printout that the teacher hands to the student. For teachers, *DiGIMS* offers the advantage of producing a detailed record of student errors; for students, however, it is difficult to see how this feedback delivery improves on the old-fashioned method of simply writing comments on the student's paper.

The Semi-Automatic Grader is similar in some respects to *DiGIMS*, but students submit their papers to *The Semi-Automatic Grader* electronically and the teacher feedback is inserted directly into the text. This Microsoft plug-in identifies possible errors (e.g., syntax, grammar, punctuation, word choice) and provides the writing instructor with the ability to track and quickly insert stock or customized evaluative comments into a text. The expanded, movable toolbar contains a series of icons with stock replies for each kind of error. It also has eight customizable scoring rubrics that provide the teacher and student with a structure for scoring the essay elements. The tool allows teachers to easily insert extended corrective comments with examples and relates these corrective comments to a rubric, which allows students to see specifically how their grades were computed; however, the tool primarily focuses on identifying errors, which is only a small piece of the total writing process.

The past two decades have also seen the development of a number of instructional writing programs based on automated essay scoring (AES) and feedback using natural language processing techniques. The more well-known examples of these include *Criterion, WriteToLearn, MY Access!* and the *Glencoe Online Essay Grader*. These online writing programs seek to support the writing process in a number of ways that seem to be congruent with the socially situated principles of writing earlier delineated. They offer immediate, targeted, personalized feedback with multiple opportunities to rewrite and thus engage students in a recursive writing process. However, although these programs provide prompt and targeted feedback as promised, our team's interactions with them have raised concerns about the quality and specificity of the feedback—and hence their usefulness in the revision process. Feedback provided by these applications includes holistic scores and accompanying narratives for the overall essay and for several traits, such as writing conventions (grammar, syntax, spelling, and punctuation), organization, word choice, and style. Feedback on writing conventions consists of identification of possibly errant text, with a clue about the nature of the problem, for example, subject/verb agreement or incorrect use of past tense. Feedback on essay organization typically consists of the identification of blocks of text that might represent parts of the essay, such as introduction, main points, supporting material, and conclusion, along with commentary about the presence or absence of such elements. Regarding word choice, the computer flags words that appear to be either over-used or words that could be replaced with ones that are more descriptive. In each instance, the student must decide if the suggestion is valid and whether or not to revise the text. Although the identification of possible errors or suggestions for improvement can be helpful, the over-identification of these can cause confusion and anxiety in the student who does not possess the necessary skills to distinguish between false positives and true errors.

AES systems have shown a high degree of accuracy—that is to say, the holistic scores assigned by computers have a high rate of exact or adjacent agreement (scoring within one point on either side of the exact score) with expert human raters. For example, *e-rater* (the scoring engine for *Criterion*) scores within a single point of expert human scorers 97% of the time, which is similar to the discrepancy rate between two human scorers (Shermis & Burstien, 2003, p.

114). *IntelliMetric* (the scoring engine for *MY Access!*) scores within a single point of expert human raters 99% of the time (Dikli, 2006, p. 27), and the *Intelligent Essay Assessor* (the scoring engine for *WriteToLearn*) scores a similar 98.1% (Pearson, 2007). However, "high reliability or agreement between automated and human scoring is a necessary, but insufficient condition for validity" (Chung & Baker, 2003, p. 29). Computers can approximate human scores, but what does this mean in terms of the computer's ability to deliver substantive feedback that may be used in the revision process? The high statistical correlation for exact and adjacent agreement is less remarkable when one understands that the typical score range in AES programs is 6 points. A score of 5 with a 1-point range in either direction would include scores of 4 and 6, or half of the possible scores. The Analytic Writing part of the GRE, scored by *e-rater*, uses a 6-point range with 98.89% of the total scores falling in the 3-6 point range and 88.74% falling in the 3-5 point range. Without looking at any exam, a scorer could achieve 88.74% adjacent agreement by assigning a score of 4 to every exam.[2] These facts alone do not necessarily negate the effectiveness of the scoring engines, but unfortunately, there have been few opportunities to assess the validity of automated essay scorers outside of industry-sponsored studies.

One concern frequently voiced about AES systems is their tendency to focus on writing products rather than process, viewing writing assessment as a summative rather than formative practice (Dikli, 2006, p. 24). In light of this criticism, many applications using AES have tried to integrate more process-oriented features into their programs, such as prewriting exercises, more substantive feedback, and the ability to save and revise drafts (Dikli, 2006, p. 24). This raises the question about the utility of feedback provided by applications using AES systems. If the computer could process text and provide specific, qualitative feedback, it could indeed be a valuable aid in the development of writing skills.

To investigate this question, our research team submitted a short essay in response to a prompt asking the writer to discuss a favorite artist. We received a score of 4.3 out of 6. We then tacked onto our initial essay several unrelated paragraphs and received a score of 5.4 out of 6. Our score was boosted one full point simply by making the essay longer using unrelated material. In another instance, our research team wrote and submitted a nonsense essay in response to a prompt about student employment. Here are the opening lines:

> Working part-time while going to school is a stupendous experience. It teaches all of the categorical imperatives available in the game of life, while also very pedagogically sound and irrelevant. I have been working as a store attendant for twenty years while in middle school and I now have the experience needed to succeed in the corporate world. There are several reasons why this is true, which I will explain in this short essay. . .

The essay was well constructed in a formulaic manner with correct grammar, syntax, punctuation and spelling, and with a liberal smattering of vocabulary related to the prompt topic. The result was a score of 6 out of 6 points. It was clear that the program assigned a high score because the text possessed surface indicators that it was a well-written piece despite the lack of underlying coherence or meaning. The apparent over-reliance on formal and narrow textual features for scoring, and the suspicion that AES programs simply correlate the presence of certain markers with good writing, are a frequently mentioned concern by researchers in the field (Chen & Cheng, 2008; Wang & Brown, 2007). Other concerns include insensitivity to creativity and context, vulnerability to cheating, the effects of not writing for a human audience, the tendency of AES programs to appreciate formulaic approaches to composition structure and wording, a focus on "counting rather than meaning making," and the inability of AES tools to appreciate stylistic differences and compositional complexities (Wang & Brown, 2007).

In our experience, these programs are not able to address meaningfully the semantics of texts and were generally unable to construct valid targeted feedback on text beyond the sentence level. In fact, the narrative accompanying holistic scores provided by the AES systems we observed, whether overall or focused on specific traits, is consistently generic and tentative: it describes a "typical" essay receiving a given score and offers tentative feedback such as "may have organization in parts, but lacks organization in other parts." This kind of response might cause the writer to wonder which parts were organized and which were lacking, leaving them with little information on how to proceed. For example, a score of 4.3 out of 6 yielded the following:

> A response that receives a score of four adequately communicates its message. Typically, a response at this level has a reasonably clear purpose evident, though the presentation may stray from the controlling point, and may

[2] See GRE Guide to the Use of Scores at <http://www.ets.org/s/gre/pdf/gre_0910_guide.pdf>.

> lack focus and cohesion. A response at this level typically shows reasonable support of ideas with adequate use of examples, but may lack supporting detail or specificity. The response shows reasonable organization with an apparent pattern of ideas presented in a logical order. The sentence structure is generally free from errors and shows some variety in presentation. Usage and word choice are generally effective. The response is generally free from mechanical errors such as spelling and punctuation, though some errors may be present that do not significantly interfere with the communication of the message (*MY Access!*).

An independent study on the effectiveness of *MY Access!* in EFL classes in Taiwan suggested that there may be a place for AES programs if used in conjunction with well placed human feedback. It showed that "although the implementation of [AES] was not in general perceived very positively by the three classes [that used *MY Access!*], it was perceived more favorably when the program was used to facilitate students' early drafting and revising process, followed by human feedback from both the teacher and peers during the later process" (Chen & Cheng, 2008, p. 94). Though most AES programs do not necessarily facilitate human interaction, neither do they preclude it. In fact, many AES programs allow the teacher to type feedback into the student essay, similar to Microsoft Word's "track changes" feature. It is, therefore, incumbent upon the teacher to exercise discretion over how AES programs are used and when to supplement them with more socially interactive activities.

Another group of automated essay scoring products is designed to give prompt and targeted sentence and word level feedback. This includes programs such as *Sentenceworks*, *Writer's Workbench,* and *EssayRater*. Some of these programs provide supplementary instructional features; for example, *Writer's Workbench* contains 29 self-guided modules that offer practice in grammar, punctuation and style, and *Sentenceworks* has plagiarism detection capabilities. Their primary purpose is to identify mechanical, grammatical, and syntactical errors and to suggest word choice alternatives. Similar to the computer-scored programs discussed earlier, the feedback is immediate and targeted to a specific text. They also have some of the same liabilities, including vagueness and the tendency to over-correct or identify problems that do not exist. For example, one program was confounded by footnotes and flagged every page with a footnote as lacking proper punctuation. Another program reacted to a previously published text with 168 suggestions for correction or improvement. A scan of the flagged items showed that only one of the suggestions (dealing with word choice) might have merit. Examples of false positives included the following: suggesting a hyphen between high and school; flagging every word preceded by a bracket or an apostrophe as an error; recommending colorful synonyms and an active voice in a context that would have been inappropriate for the research paper that was written; and suggesting that any sentence over 20 words in length was too long and should be broken up (this sentence contains 69 words!). When we pasted into one program Lincoln's *Gettysburg Address*, it highlighted one-third of the text as too wordy and stated "mistakes are made by using too many words, or by repeating words and ideas."

At least two companies that use automated error identification describe their products as "Microsoft Word on steroids;" however, our team's experience with these hyperactive programs is that less may be better. Over-correction is not only time consuming—deciding whether or not an error is valid requires a high level of expertise and patience from the user. Also, the language used to deliver the feedback can be technical and difficult for students to understand (such as *split infinitives* and *dangling modifiers*). More concerning is the view of writing that these programs' feedback presents to students. The emphasis tends to be on a narrow, formalistic view of writing that is "correct" at the word or sentence level rather than a broad view of writing that focuses students on the social function of writing.

Our team also reviewed programs for their potential to facilitate collaboration. None of the programs previously discussed provide a workspace that allows multiple student input beyond a simple tracking feature where peers may enter corrections and comments. These programs appear to ignore the insights that might be gleaned from social network writing spaces such as Facebook, Twitter, and so on, as well as emerging collaborative writing tools such as Google Docs and Etherpad. The proliferation of these social networking and collaborative writing technologies in the real (out-of-school) world provide further evidence that writing is more than ever a socially situated, interactive meaning-making activity.

3.2. Function, form, and genre

A common feature of writing assessment tools that rely on automatic essay scoring is their routine use of writing prompts to delineate the parameters of the essay for purposes of computerized scoring (Burstein, 2003; Elliot, 2003; Landauer, Laham, & Foltz, 2003). Scoring models are built for each specific prompt, using data from large numbers

of human-scored essays (Attali & Burstein, 2006). Teachers using these programs can select from libraries containing hundreds of prompts and several genres in order to create assignments. With this many options, teachers find prompts relevant to their courses, and it is possible that students would find topics that interest them. Teachers are permitted to create their own prompts, but teachers who do so are warned that the scoring engine will not be able to give trait feedback as it can with prompts from the program library. Programs that require a pre-formulated prompt limits teachers and students who want to select topics relevant to their own needs and interests. Students who are invested in their topic may be more motivated than those who are assigned a topic or select it from a prescribed list. Also, despite the large number of prompts, teachers may be frustrated to not find ones that link with the curricular materials used in their classes;. For example, one of the programs has 72 persuasive essay prompts available for middle school grades, but only 2 of those prompts are on art-related topics. Additionally, our team searched one middle school library for the topics *Shakespeare*, *poetry*, and *children*, and could find no prompts for any of these topics.

Learning how to write to a prompt may be of value when the prompt resembles certain real-life activities, like filling out an application or taking a test; however, studies of real-world writing indicate the varied nature of writing, where writers compose in many different forms for a wide range of functions and audiences (Bazerman & Paradis, 1991; Beaufort, 1999; Luff, Hindmarsh & Heath, 2000; Odell & Goswami, 1985). Over-reliance on prompts is problematic, particularly when those prompts not only constrain topics but also delimit genre, as the majority of the programs we examined do. In addition, if the scoring engine is the only one "reading" the work, the student may learn to focus on surface features of writing that correlate with a good score but not necessarily a good essay (Chen & Cheng, 2008), for example vocabulary and essay length (Attali & Burstein, 2006,); absence of mechanical errors (Jones, 2006); and structure and grammar emphasized over content and meaning (CCCC, 2006). Students may come to think of writing as "not the art of saying something well but rather of saying something new using a set of preexisting rules" (McAllister & White, 2006, p. 27). Thus, the view of "good writing" is equated with a narrow definition of "correct writing," and genre is merely connected to form rather than, in Bakhtin's (1986) words, "spheres of human activity" (p. 65).

In addition, we examined the computer-based writing programs in terms of their capacity to provide students with options, not only in the tools and resources provided but also in how projects are conceived, planned, and constructed. Apart from topic selection, prewriting is, by definition, the earliest point at which students may begin to develop a map of their writing project. Prewriting engages students in activities meant to help them generate or organize ideas for composition, and it can be effective in preparing students for writing. A 2007 meta-analysis by Steve Graham & Delores Perin, 2007 found prewriting exercises to have a positive mild to moderate impact on improving the quality of student writing. Prewriting activities, which may include individual or group activities, typically involve early information gathering, visual representation of ideas, organizational plans, or a combination thereof.

In our survey of writing software programs, we found a dearth of prewriting aids. Most programs we looked at do not address prewriting at all, and those that do typically provide a collection of templates that students may use to develop visual organizational representations. These templates include Venn diagrams, outlines, and a variety of charts into which students may enter text. Our research team found the available charts and diagrams to be very limited with no opportunity to adapt them to individual purposes; the programs gave little, if any, information on how to utilize them effectively. Rather than engage and expand students' creative powers, these templates serve to mold and harness student work into simple and formulaic structures based on specific genres.

As we saw earlier, most writing programs use prompts to generate writing topics for students; the frequent use of prompts may diminish student creativity, motivation, and ownership of the writing process. Our team was interested to know how well these same programs support student individuality and creativity during the rest of the writing process. Of particular interest to us were programs such as *Criterion*, *WriteToLearn*, and *MY Access!,* both because they are widely used and because they evaluate more holistic aspects of writing such as organization, style, and voice—aspects that are only apparent by taking a larger view of the written work. Descriptions of the criteria used to assess these aspects of writing seem promising; for example, one program that evaluates the author's "voice" describes it as "the sum of all decisions an individual makes when writing. It's what makes writing unique." Another program sent us the following response to our question about how voice is defined in their essay evaluations:

> [Our program] defines voice the same way that teachers define and score voice. Looking at aspects such as tone, using the correct tone that suits the topic, audience and purpose. Also the writer's ability to connect with their audience, in that they anticipate the audience's response and connect directly with the reader. (A quote taken from

> a written communication from the company's representative in response to our query on how voice is defined in essay evaluations.)

We understood these descriptions of voice to mean that the scoring mechanism is able to appreciate not only a variety of compositional elements but also variations in the way these parts come together to form a unique and meaningful whole. At the same time, based on our interaction with these tools, we questioned the ability of the AES systems to detect meaning and evaluate the essay in a more holistic manner. We found, in fact, that the computer generated a stock response based on the holistic score it assigned to individual sub-traits such as "voice." In one instance, the program gave the following advice for improvement of voice:

> Revision Goal 1: Use words effectively
> Revision Goal 2: Use well-structured and varied sentences

While this kind of vague information may not be particularly useful or inspirational to students in their writing process, it does not necessarily inhibit the student from pursuing a unique strategy for writing. We then looked at the kind of feedback generated by these programs on "organization and development." In this case the computer evaluated our essay by identifying traditional elements such as introductory material, thesis statement, main ideas, supporting ideas, and conclusion. It identified and underlined the text signifying each of these structural elements, then posed a tentatively-phrased question and offered stock advice. For example,

> Is this [underlined] part of your essay your introduction? In your introduction, you should capture the reader's interest, provide background information about your topic, and present your thesis sentence. Look in the Writer's Handbook for ways to improve your introduction.

The program appeared to select structural elements based on position (for example, it invariably selected the first sentence of each paragraph as the main point) and based its diagnosis on the presence of key transitional and ordinal words and phrases (*such as*, *although*, *for example*, *therefore*, *first*, *second*, *third*, and *in conclusion*). We found the program's assumptions about organizational elements often to be incorrect. In one instance it misidentified our thesis statement, which appeared at the end of the first paragraph:

> In my essay I would like to focus on several reasons why students' robust participation in curriculum selection might not only increase student achievement, but also student motivation and ownership of his or her educational experience, thus having a net positive effect.

It identified this sentence as "other" (not a valid structural element) and responded with the following perplexing query:

> Is this material a title, class name, section number, opening, closing, signature, or name? This material does not seem to be part of your essay.

This kind of incorrect identification and confusing response could leave the writer with more questions than answers. Not only might it be difficult for the student to know how to proceed in situations like this, but there is also the vague sense that the student has done something wrong that needs to be fixed. Counter to developing confidence in one's own voice and approach to writing, the take away message here is that the best way to obtain "approval" in the future is to stick to standard writing structures that include traditional organizational features and stock transitional phrases. In fact, one of our team's more sophisticated essays was flagged as "off topic" and targeted for human scoring in addition to the machine score. Although we appreciate the fact that a human expert would be needed to validate the computer score, we fear that the persistent underlying urge towards conformity may stifle individual creativity.

The pedagogical effectiveness of these programs may hinge on the purposes and the extent to which they are used. Students do need to learn how to write clear and concise short essays in response to a question or prompt. They need to learn the nuts and bolts of writing conventions and the basic structural elements of a variety of genres. To the extent that students are able to use these products in support of these purposes, and in doing so separate the wheat from the chaff of computerized feedback, these programs may be helpful. However, we do have concerns about the conforming effects of the persistent or inappropriate use of them for broader writing activities.

Several AES programs specifically seek to give students practice in writing effective short essay responses, particularly in venues (large lecture halls) and disciplines (business, sociology, etc.) in which multiple-choice testing tends to dominate. These programs primarily evaluate content, with feedback on writing form being a secondary concern. Two such programs are *MarkIT* and *SA Grader*. *SA Grader* gives students feedback on short essay content written to a series of prompts on a particular course-related topic. The program checks for key terms, concepts, and the relationships between them. Feedback is immediate, suggesting to students which parts of answers are correct and giving hints about missing content so that students may revise and resubmit. Students may also challenge the program's assessment, and the program's assessment can be overridden by the teacher. The makers of *SA Grader* refer to the pedagogy employed by their product as asynchronous collaboration because the teacher is interacting indirectly with the student through use of the teacher-designed activity (Brent, et al, 2009). *SA Grader* has been shown to improve students' final grade (content knowledge) by an average of 20%, but no claims are made for the improvement of writing skills (Brent et al., 2009). The program operates on the assumption that it is important to provide essay writing opportunities and exercise students' writing skills in a variety of disciplines.

Our team wrote essays for *SA Grader* in response to three questions on "operant conditioning." We found that the specific terms had to be spelled precisely to register as correct by the program; for example, it recognized "1. Fixed ratio" as a correct answer, but it did not recognize "1.Fixed ratio" as correct because it lacked a space between the 1 and the F. We also found that the program only looked for terms and concepts without attention to grammar, sentence structure, or mechanical errors. After we submitted a grammatically and factually correct essay that earned a score of 10 out of 10, we deleted much of the non-content-related text and resubmitted it. Again, the essay received a 10 out of 10 score. The program only scored for key terms and phrases and could not detect grammar errors or incoherent sentence structure. Programs like *MarkIT* and *SA Grader* neither enlist students' creative writing skills nor evaluate writing form; however, they do provide opportunities for students to practice writing in a specific genre in disciplines and venues that otherwise might rely entirely on item-based testing. We question the usefulness of writing practice, however, when it is not accompanied by formative feedback.

Other AES-based products like *Criterion*, *WriteToLearn*, and *MY Access!* provide assignments in a variety of genres depending upon the grade level. For example, at the primary grade levels, *Criterion* offers writing exercises in descriptive, persuasive, process, cause and effect, compare and contrast, informative and narrative genres. At the high school level, *Criterion* offers exercises in expository, narrative, descriptive, and persuasive genres. Although it may be necessary for students to learn the basic characteristics of these genres, it is also important to understand that writing approaches within each genre can widely vary. One concern our team had about the way AES systems evaluate writing is their adherence to a rigidly conventional view of genre that may make it difficult for the scoring engine to appreciate variations in writing style. For example, it was our experience in writing persuasive essays that those constructed in a more formulaic manner, containing standard transitional and ordinal words and phrases, were frequently scored more favorably than much more complex essays that did not adhere to a simple formula. A steady diet of this kind of feedback might eventually dissuade students from being inventive, using different writing strategies, and taking the occasional risks that can make their writing intellectually adventurous or creative.

By way of contrast, the *Choices* program offers students a variety of writing projects, each connected to a different genre or function. The students are given flexible assignment parameters with plenty of room for the writer to take an individual approach to the task. The assignments utilize genres that require the students to do research, think and write critically, and sometimes challenge social norms; for example, one assignment is a persuasive paper meant to shape public opinion on an issue that impacts the community. Some of the other assignments include an informative essay that requires students to conduct an investigation, an academic project that involves research, and a media critique:

> Project 3: Media Critique - For this project, you will produce a media critique that analyzes how a specific magazine advertisement uses cultural myths to make its appeals to the target audience for the ad. You may choose to write about an advertisement that appears in any type of magazine, and you should take into account the type of magazine in which the advertisement appears and how that helps to define the target audience. As an alternative, you could choose to examine a television or radio advertisement, but if you do, you will need to have some way to save or copy the ad so that you can view it as many times as you need to. As an alternative, you might pick an ad that is repeated frequently and take very good notes when you see it. Your job is to make the connections between the text, its target audience, its use of cultural myths, and its appeals.

This assignment encourages students to find their own voice on issues of social importance, to develop skills that will equip them to effectively critique social and political institutions, and to "[link] the practice of schooling to democratic principles of society and to transformative social action" (Darder, 2007, p. 113). Assessment of this kind of writing assignment is highly complex; fortunately, this program's peer review capabilities can accommodate some of those complexities.

3.3. Multimodal meaning-making

This section is extremely limited as we found few examples of multimodal meaning-making in these programs. Ironically, the promoters of these programs clearly understand the power of multimodality. You will find a full range of multimodal meaning-making practices on the writing assessment software, websites, and all of the programs' websites combine text and images to convey information. Many of the sites, including *Criterion, Mark IT, MY Access! MyCompLab, Semi-Automatic Grader, Writers Workbench,* and *WriteToLearn* also provide video for program demonstration and promotion, including many video testimonials. In addition to textual feedback, many of these same programs also provide visual feedback to students in the form of graphs and charts. But what about student opportunities for multimodal meaning-making? We discovered that only one of these programs, *Choices 2.0*, enables students to add images and upload video to their compositions. Despite wide recognition that "writing in the 21st century" is "made not only in words" (Yancey, 2004, 2009), these writing assessment programs privilege text over sound, image, and video. In fact, text eclipses all other meaning-making practices.

4. Summary and Conclusion

The programs that we have described possess many appealing features: quick feedback, reliability, plagiarism detection, the capacity to connect with state standards, and assessment rubrics. These features are particularly attractive in today's test-driven educational environments.

In fact, many of these programs emulate standardized writing tests. It is therefore unsurprising to find efficacy reports (MY Access, 2007; Pearson, n.d.) that demonstrate that using such programs have the effect of increasing test scores. What is surprising, however, is that many of these programs largely neglect the potential that technology has to offer in terms of our three foundational understandings of writing: that it is a socially-situated practice; that it is a functionally and formally diverse activity; and that it is increasingly multimodal. In terms of our first principle, we found that some of the programs connected students to real readers and human generated feedback—but many did not. Similarly, interaction through collaborative writing, an area where technology has made huge strides, was absent in these programs. We also found the programs inadequate in absorbing or linking ancillary aspects of the writing process such as pre-writing and ongoing extra-textual dialogue of various kinds. Finally, the vast majority of the programs that we investigated were not equipped to enable multimodal meaning making practices.

Instead, we found evidence of formulaic approaches, non-specific feedback, incorrect identification of errors, a strong emphasis on writing mechanics such as grammar and punctuation, and a tendency to value length over content. We found writing programs that assumed that successful student writers would reproduce conventional, purely written-linguistic generic structures. The problems that we found were particularly prevalent in programs that relied primarily on automated assessment. When focusing on these findings, it is easy to see why literacy educators take umbrage when faced with the use of such programs as they appear to trivialize the complex process of writing assessment and undermine teacher professionalism. At the heart of opposition to machine scoring is the fervent conviction that writing is social in nature, and that as such, "all writing should have human readers, regardless of the purpose of the writing" (CCCC, 2004, para. 11).

Is it simply the computers that are the problem? After all, computer technologies in the wider world facilitate broader and more diverse communicative practices—and we all use a form of automated assessment and feedback every time we click on the spelling and grammar checker. We want to suggest that the fault is not with the technology, but with the vision of writing and assessment that these programs generally assume and promote—a narrow view that conforms to systems requirements in an era of testing and accountability. New technology is being used to reinforce old practices.

The challenge for new writing technologies then is not entirely technological. In computer supported learning environments–as in any classroom–ways must be found to ensure that writing is recognized and valued as a socially situated, diverse, and multimodal phenomenon. Whether in the classroom or online, it is important to consider the

message being sent to students about what matters in writing. Expanding our view of writing beyond the mechanics of generic form also means expanding our vision of assessment. Emerging writing technologies might help, but only if these online assessment environments are embedded within a context where students are provided with the opportunity to explore the social contextuality, diversity, and multimodality of what it means to write in the digital age.

Appendix A. Writing Programs Analyzed in this Research

Bayesian Essay Test Scoring System: <http://echo.edres.org:8080/betsy/>
Calibrated Peer Review: <http://cpr.molsci.ucla.edu/>
Choices: <http://www.choicesportal.com>
Criterion: <https://criterion2.ets.org/cwe/> and <http://www.ets.org/criterion>
DIGIMS Online: <http://www.digimsonline.com/>
EssayRater: <http://www.essayrater.com> (now Grammarly: <http://www.grammarly.com/?er>)
Glencoe Online Essay Grader: <http://www.glencoe.com/sec/glencoewriting/>
Mark IT: <http://www.essaygrading.com/>
MY Access!: <http://www.myaccess.com>
MyCompLab: <http://www.mycomplab.com/>
Project Essay Grade: (Page, 2003).
SAGrader: <http://www.sagrader.com/>
Semi-Automatic Grader: <http://sio.midco.net/jblessinger/>
sentenceworks: <http://www.sentenceworks.com/> (now Grammarly: <http://www.grammarly.com/edu/>)
Webgrader: <http://sourceforge.net/projects/indautograder/>
Writer's Workbench: <http://www.writersworkbench.com>
WriteToLearn: <http://www.writetolearn.net/>

Appendix B. Categories for Data Collection and Analysis

Program name
Original author(s)
Publisher/Owner
Website address
Development date
History of ownership
Target audience/Grade level
Primary purpose
Underlying primary algorithm
Use of corpus training
Technological paradigm
Plagiarism detection capability
Number of institutions using the program
Free trial availability
Individual user cost
Site use cost
User interface type
Program description and features
Multimodal capabilities
Facilities for interaction and collaboration
Feedback mechanisms
Genres/Forms of writing promoted
Opportunities for engaging in writing process
Primary strengths
Primary weaknesses

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

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

Sarah McCarthey 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

Mary Kalantzis is Dean of the College of Education at the University of Illinois, Urbana-Champaign. She was fomerly 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>

References

Attali, Yigal, & Burstein, Jill. (2006). Automated essay scoring with e-rater V.2. *The Journal of Technology. Learning, and Assessment*, *4*(3), 1–29.

Bakhtin, Mikhail. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.

Barton, David, & Tusting, Karin (Eds.). (2005). *Beyond communities of practice: language, power, and social context*. Cambridge: Cambridge, UK University Press.

Bazerman, Charles, & Paradis, James (Eds.). (1991). *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities*. Madison, WI: University of Wisconsin Press.

Bazerman, Charles. (2004). Speech acts, genres, and activity systems: How texts organize activity and people. In Charles Bazerman, & Paul Prior (Eds.), *What writing does and how it does it: An introduction to analysing texts and textual practice* (pp. 309–339). Mahwah, NJ: Lawrence Erlbaum.

Bazerman, Charles (Ed.). (2008). *Handbook of research on writing: History, society, school, individual, text*. New York Lawrence: Erlbaum.

Beaufort, Anne. (1999). *Writing in the real world: Making the transition from school to work*. New York, NY: Teachers College Press.

Black, Paul, & Wiliam, Dylan. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7–74.

Brandt, Deborah. (1990). *Literacy as involvement: The acts of writers, readers, and texts*. Carbondale, IL: Southern Illinois University Press.

Brent, Edward, Atkisson, Curtis, & Green, Nathaniel. (2009). Time-shifted online collaboration: Creating teachable moments through automated grading. In A. Angel Juan, Thanasis Daradournis, Fatos Xhafa, & Santi Caballe (Eds.), *Monitoring and assessment in online collaborative environments: Emergent computational technologies for e-learning support*. Hershey, PA: IGI Global.

Burstein, Jill. (2003). The E-rater scoring engine: automated essay scoring with natural language processing. In Mark D. Shermis, & Jill C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum Associates.

Chen, Chi-Fen Emily, & Cheng, Wei-Yuan Eugene. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language, Learning & Technology*, *12*(2), 94–112.

Chung, Gregory K. W. K., & Baker, Eva L. (2003). Issues in the reliability and validity of automated scoring of constructed responses. In Mark D. Shermis, & Jill C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary assessment* (pp. 23–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Conference on College Composition and Communication (2004). CCCC position statement on teaching, learning, and assessing writing in digital environments. Retrieved from http://www.ncte.org/cccc/resources/positions/digitalenvironments

Cope, Bill, & Kalantzis, Mary. (1993). *The powers of literacy: Genre approaches to teaching writing*. London, UK and Pittsburgh, PA: Falmer Press (UK edition) and University of Pennsylvania Press (US edition).

Cope, Bill, & Kalantzis, Mary (Eds.). (2000). *Multiliteracies: Literacy learning and the design of social futures*. New York, NY: Routledge.

Cope, Bill, & Kalantzis, Mary. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal*, *4*, 164–195.

Darder, Antonia. (2007). What is critical pedagogy? In William Hare, & John P. Portelli (Eds.), *Key questions for educators* (pp. 113–117). San Francisco, CA: San Francisco Press.

Dikli, Semire. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, *5*(1), 1–35.

Elliot, Scott. (2003). IntelliMetric: From here to validity. In Mark D. Shermis, & Jill C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71–86). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, Patricia F., & Haswell, Richard H. (Eds.). (2006). *Machine scoring of student essays: Truth and consequences*. Logan, UT: Utah State University Press.

Gee, James. (2005). Semiotic social spaces and affinity spaces: From the age of mythology to today's schools. In David Barton, & K. Karin Tusting (Eds.), *Beyond communities of practice: Language, power and social context*. Cambridge: Cambridge, UK University Press.

Graham, Steve, & Perin, Delores. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. New York: The Carnegie Corporation of New York.

Halliday, Michael. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Baltimore, MD: University Park Press.

Heath, Shirley B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.

Hillocks, George. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.

Hodge, Robert, & Kress, Gunther. (1988). *Social semiotics*. Cambridge: Polity.

Huot, Brian, & Neal, M. (2006). Writing assessment: A techno-history. In Charles MacArthur, Steve Graham, & Jill Fitzgerald (Eds.), *Handbook of writing research* (pp. 417–432). New York: Guildford Publications.

Hull, Glynda A., & Nelson, Mark. E. (2005). Locating the semiotic power of multimodality. *Written Communication*, *22*, 224–261.

Hymes, Dell. (1972). In Courtney Hazden, P. John Vera, & Dell Hymes (Eds.), *Introduction to Functions of Language in the Classroom*. New York: Teachers College Press.

Jones, Edmund. (2006). Accuplacer's essay-scoring technology: When reliability does not equal validity. In Patricia Freitag Ericsson, & Richard Haswell Ericsson (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 93–113). Logan, UT: Utah State University Press.

Kress, Gunther. (2009). *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.

Kress, Gunther, Jewitt, Carey, Ogborn, Jon, & Tsatarelis, Charalampos. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London: Continuum.

Landauer, Thomas K., Foltz, Peter W., & Laham, Darrell. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Landauer, Thomas K., Laham, Darrell, & Foltz, Peter. (2003). Automatic essay assessment. *Assessment in Education*, *10*(3), 295–308.

Lave, Jean, & Wenger, Etienne. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lemann, Nicholas. (1999). *The BigTest: The secret history of the American meritocracy*. New York: Farrar, Straus, and Giroux.

LeFevre, Karen B. (1987). *Invention as a social act*. Carbondale, IL: Southern Illinois Univeristy Press.

Lemke, Jay. (2002). Travels in hypermodality. *Visual Communication*, *1*, 299–325.

Luff, Paul, Hindmarsh, Jon., & Heath, Christian (Eds.). (2000). *Workplace studies: Recovering work practice and informing system design*. Cambridge, UK: Cambridge University Press.

McAllister, Ken S., & White, Edward M. (2006). Interested complicities: The dialectic of computer-assisted writing assessment. In P. F. Ericsson, & R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 8–27). Logan, UT: Utah State University Press.

MY Access! (2007). MY Access! efficacy report. Retrieved from http://www.vantagelearning.com/docs/myaccess/myaccess.research.efficacy.report.200709.pdf.

National Governors Association & Council of Chief State School Officers. (2010). Designing common state assessment systems. Retrieved from http://www.edweek.org/media/ngaccsso_assessmentdesignpaper.pdf.

Newkirk, Thomas. (2009). *Holding on to good ideas in a time of bad ones*. Portsmouth, NH: Heinemann.

New London Group (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review* 66, 60-92.

Nickle, Todd. (2006). Using calibrated peer review (CPR) to improve student communication skills. In: M. A. O'Donnell (Ed.), *Tested studies for laboratory teaching,* Volume 27 (pp. 329-333). Proceedings of the 27th Workshop/Conference of the Association for Biology Laboratory Education (ABLE).

Odell, Lee, & Goswami, Dixie (Eds.). (1985). *Writing in nonacademic settings*. New York: Guilford Press.

Page, Ellis Batten. (2003). The Project Essay Grade. In Mark D. Shermis, & Jill C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 262–271). Mahwah, NJ: Lawrence Erlbaum Associates.

Pearson. (2007). *Evidence for reliability, validity and learning effectiveness*. Retrieved from http://www.pearsonkt.com/papers/WTLReliabilityValidity-082007.pdf.

Pearson. (n.d.). *Demonstrating reading and writing performance gains: WriteToLearn efficacy report.* Retrieved from http://school.writetolearn.net/downloads/WTL_EfficacyReport.pdf.

Prior, Paul. (2009). From speech genres to mediated multimodal genre systems: Bakhtin, Voloshinov, and the question of writing. In Charles Bazerman, Adair Bonini, & Debora Figueredo (Eds.), *Genres in a changing world* (pp. 17–34). Fort Collins, CO: WAC Clearinghouse and Parlour Press.

Shermis, Mark D., & Burstein, Jill C. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum.

Shipka, Jodie. (2009). Negotiating rhetorical, material, methodological, and technological difference: Evaluating multimodal designs. *College Composition and Communication*, *61*, 343–366.

Street, Brian V. (1984). *Literacy in theory and practice*. Cambridge: Cambridge University Press.

Vygotsky, Lev S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wang, Jinhao, & Brown, Michelle Stallone. (2007). Automated essay scoring versus human scoring: A comparative study. *The Journal of Technology, Learning, and Assessment*, *6*(2), 1–27.

Wenger, Etienne. (1998). *Communities of practice: Learning, meaning, and identity*. New York: Cambridge University Press.

Yancey, Kathleen B. (2004). Made not only in words: Composition in a new key. *College Composition and Communication*, *56*(2), 297–328.

Yancey, Kathleen B. (2009). Writing in the 21st century. Retrieved from http://www.ncte.org/library/NCTEFiles/Press/Yancey_final.pdf.