

Out of the box: A review of Ericsson and Haswell's (Eds.) *Machine Scoring of Student Writing: Truth and Consequences*

by Elliot Knowles

"What we need are new tools for thinking with, new frames in which to place things, in which to see the old and the new, and see them both newly." Gunther Kress: *Literacy in the new media age*

Overview

Technology, like life, will out. Much like life finding a way--through persistence or sheer genetic determinism--even the most mundane technologies will either be deemed useful through their intended use, or through some unexpected purpose derived by a savvy technician. Once the box is opened, any possible users of technology can find a use in practice. Whether the technology be as simple as banging two rocks together to produce a spark or as complicated as the biomechanical apparatus of a pacemaker, the intricate social construction of symbol systems such as languages, the codification of the human genome, or the spiraling production of knowledge, techniques and artifacts alike hinge on the manner in which they are interpreted and used. By taking this social constructivist view of technology, a technology cannot exist as a neutral, autonomous entity. Langdon Winner (1986) expands this argument by stating:

what matters is not technology itself, but the social or economic [or political] system in which it is embedded. This maxim, which in a number of variations, is the central premise of a theory that can be called social determinism of technology. It serves as a needed connective to those who focus uncritically upon such things as 'the computer and its social aspects' but who fail to see the social circumstances of their development, deployment, and use. (p. 20)

Like any other technology, writing assessment and the techniques and artifacts associated with it fall prey to missing the forest for the trees when it comes time to evaluate how such technologies are put into use. Most notably this plays out in the fact that computers cannot as of yet read and interpret writing.

In the introduction to their anthology, *Machine scoring of student writing: Truth and consequence*, Patricia Ericsson and Richard Haswell (2006) warn that "new technology can sneak in the back door and establish itself while those at the front gates, nominally in charge, are not much noticing" (p. 1). Yet, as the technology of machine scored essays enters through the back door, Ericsson and Haswell report that with the exception of writing teachers and composition scholars such as Dennis Baron, Anne Herrington and Charles Moran, Julie Cheville, and Mike Williamson, "the response on machine scoring from the academic community ... has been silence" (p. 2). The anthology stands in opposition to this silence: If you are interested in writing assessment and the emergent technologies involved in automated means of scoring student essays, and you wish to take part in the larger conversation of whether or not programs such as ACCUPLACER--a program used by the College Board to rate student writing samples--are viable and appropriate assessment tools, this book is a necessary read. It contains sixteen essays written by both scholars and educators: McAllister and White, Ericsson, Anson, Haswell, McGee, Jones, Herrington and Moran, Matzen Jr. and Sorensen, Ziegler, Maddox, Corso, Whithaus, Brent and Townsend, Rothermel, Condon, and Broad. The essays in this collection range from Haswell's extensive bibliography of over 40 years of work on machine-scored student essays to the theoretical and interpersonal implications of machine scoring to research collected by teachers involved in programs using various machine scoring platforms to projections of where the advent of computer assisted rating and response to student writing might lead us as assessors and educators.

Highlights

The first essay in the collection, "Interested complicities: The dialectic of computer-assisted writing assessment" by Ken McAllister and Ed White sets the stage for the larger discussion of machine scoring of student writing. They state,

our purpose here ... is to offer readers a broad perspective on how computer-assisted writing assessment has reached the point it occupies today, a point at which the balance of funding is slowly shifting from the research side to the commercial side. (p. 9)

They cite the various players in computer-assisted assessment as English departments, researchers, entrepreneurs, adopters, and finally users--students. Regardless of the various stakeholders, they conclude that

writing teachers need to adopt a model of praxis--a process of critical (including self-critical) reflection and informed

practice towards just ends--as they pursue their interests concerning computer-assisted writing assessment. This means that all complicit parties, but most particularly the faculty (which ultimately owns the curriculum), need to be aware of the history and profundity of the issues behind computer-assisted writing assessment. (p. 27)

The awareness and reflection called for by McAllister and White is necessary for the stakeholders most affected by this technology--teachers and students--to have a place in the conversation they have so far lacked.

While McAllister and White don't adopt a specific stance on computer-assisted writing assessment, Chris Anson unequivocally states, "the processes humans use to read, interpret, and evaluate text can't be replicated by a computer Machines are incapable of reading natural discourse with anything like the complexity that humans read it" in his article, "Can't touch this: Reflections on the servitude of computers as readers" (p. 39). However, having made these claims, Anson goes on to argue that the fact that computers cannot "read" a text should not hinder the "continued exploration of digital technologies both to analyze human prose and possibly to provide formative information that might be useful to developing writers" (Ibid.). The notion that computers cannot read texts the way a human in no way suggests that computers will be incapable to one day process and interpret texts in viable ways.

Richard Haswell's essay, "Automatons and automated scoring: Drudges, black boxes, and *die ex machine*" argues that computers cannot offer formative feedback to student writing spelling and grammar corrections. Tracing the history of programs geared around computer language analysis starting in the mid-1950's through programs such as ACCUPLACER introduced in the early 2000's, Haswell probes the question of how we arrived at the point where universities use programs such as WritePlacer to make placement decisions for students (pp. 58-59). Haswell's history warns against the danger of teachers and administrators blindly accepting claims that computer-assisted writing assessment makes assessment easier and faster. He argues,

in terms of curricular potential there is more here than the computer algorithms of sentence length and topic token-word maps, and also more than faculty alarms over spelling ... and comma splices. Writing faculty, as well as machines, need the skill to diagnose such subtleties and complexities. (p. 78)

In "Computerized writing assessment: Community college faculty find reasons to say 'Not yet'," William Zeigler reports on the E-Write pilot at J. Sergeant Reynolds Community College in Virginia. The pilot study found that "both the overall scores and the five analytic scores tended to cluster in a midrange [And] no scores corresponded closely with instructor ratings of the samples"; follow-up surveys showed that E-Write scores did not predict student success in composition courses any better than faculty reader scores, and "more than 25 percent of the pilot samples stumped the scoring engine and required human assessment" (p. 140). Yet, Zeigler reports that this trial was not without usefulness despite its lack of success:

Writing faculty see placement through a lens that finds usefulness in the work of creating and maintaining a placement instrument [C]onducting writing placement forces faculty to revisit vital questions: what are the basic skills of writing? What traits do we agree to recognize as demonstrating competence in these skills? (p. 146)

While the current platforms of computer-assisted writing assessment are not yet living up to the hype of the companies that sell them or meeting the levels of validity and reliability that are demanded in high-stakes decisions such as placement, this does not mean we should abandon hope that machine scoring can one day meet these standards.

Closing the anthology, Bob Broad's essay "More work for the teacher?: Possible futures of teaching writing in the age of computerized assessment," directly attacks the claims of companies such as ETS that "students and teachers of writing ... ought to accept [the view that assessment is time taken away from teaching] and endorse it by purchasing products that help to separate teaching from assessment" (p. 224). Broad, leaning on Huot (2002), argues that teaching and assessment cannot and should not be separated (p. 225). Yet, computer-assisted writing assessment technologies are here and in use, leaving educators and administrators with the responsibility to examine, explore, and fully consider the consequences of these technologies. Broad explains,

[W]e have the solemn responsibility to study and predict the impact on rhetorical learning of these various applications. If we, as professional educators, determine that a particular use of artificial intelligence helps students and teachers meet established learning goals, then we should support and invite that use of technology. Where we determine that use of computerized evaluation would trivialize and denude rhetorical instruction and experience, we must fight it and prevent it from being used. (p. 232)

Moving Forward

Broad's call for responsible inquiry regarding machine scoring of student essays permeates throughout the collection. At the very start, Ericsson and Haswell explain,

[o]ur primary goal ... is not to counter industry viewpoints, solely to cast a con against their pro. This volume does not propose some countertechnology to jam the industry software. It just questions the 'truth' that industry publicizes about automated essay scoring and problematizes the educational 'consequences.' It takes the discussion of machine scoring to a broader level and a wider audience, to the kind of polyvocal discussion and critical analysis that should inform scholarly study and civic discourse. (p. 2)

The anthology challenges the notion that assessment can blindly be purchased off the shelf and out of the box to make our lives as educators and administrators magically easier. As responsible educators and administrators, it is our duty to evaluate the needs of our students and faculty on a local level and provide them with whichever means best meet those needs with the least negative consequences. This anthology both challenges the notion that machine scoring of student writing is inherently evil and that we are powerless to resist technologies forced upon us. By critically evaluating the development, deployment, and use of assessment technologies such as machine scoring of student writing, we can continue to satisfy our responsibility to our profession and our students to provide the most useful environment for learning, be it human or artificial.

References

- Anson, C. (2006). Can't touch this: Reflections on the servitude of computers as readers. In P. Ericsson and R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 38-56). Logan, UT: Utah State University Press.
- Broad, B. (2006). More work for teacher?: Possible futures of teaching writing in the age of computerized writing assessment. In P. Ericsson and R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 221-33). Logan, UT: Utah State University Press.
- Ericsson, P., and Haswell, R. (Eds.). (2006). *Machine scoring of student essays: Truth and consequences*. Logan, UT: Utah State University Press.
- Haswell, R. (2006). Automatons and automated scoring: Drudges, black boxes, and dei ex machina. In P. Ericsson and R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 57-78). Logan, UT: Utah State University Press.
- Huot, B. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge.
- McAllister, K. S., and White, E. (2006). Interested complicities: The Dialectic of computer-assisted writing assessment. In P. Ericsson and R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 8-27). Logan, UT: Utah State University Press.
- Winner, L. (1986). Do artifacts have politics? In *The whale and the reactor: A search for limits in an age of high technology* (pp. 19-39). Chicago: University of Chicago Press.
- Ziegler, W. W. (2006). Computerized writing assessment: Community college faculty find reason to say 'Not yet'. In P. Ericsson and R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 138-46). Logan, UT: Utah State University Press.

Bio

Elliot Knowles is an advanced graduate student at Kent State University in the Literacy, Rhetoric and Social Practice Program. His research interests include feminist research methodologies, psychometrics, and classroom writing assessment. He is currently collecting data for his dissertation on the role of assessment in the college writing classroom.