
Reviewed by Asao B. Inoue, California State University, Fresno

This collection begins (in the first three chapters) and finishes (in the final two chapters) with a familiar, if not obvious, refrain in writing assessment circles and generally in composition studies: machines and computers cannot read student writing, at least not in the meaningful ways humans can. Patricia Freitag Ericsson and Richard Haswell, in their introduction, place this collection next to Mark Shermis and Jill Burstein’s edited collection, Automated Essay Scoring: A Cross-Disciplinary Perspective, published in 2003, which generally supports computer assisted essay scoring, but Ericsson and Haswell insist: “This volume does not propose some countertechnology to jam the current industry software. It just questions the ‘truth’ that industry publicizes about automated essay scoring and problematizes the educational ‘consequences’” (2). Underneath this statement is the Shermis and Burstein collection. And if one isn’t careful, given the refrain that bookends the present collection (chapters by McAllister and White, Ericsson, Anson, Condon, and Broad) and can be found in several essays in between (the chapters by McGee, Jones, Herrington and Moran, Matzen and Sorensen, Ziegler, Maddox, and Rothermel), one might take the collection to be a simple rebuttal to the Shermis and Burstein collection; however, I argue that while the present collection is a counter to Shermis and Bernstein, there is an interesting undercurrent happening in the collection as a whole that stays true to Ericsson and Haswell’s initial claim: the computer scoring technologies currently under development (e.g. ACCUPLACER, e-Writer, and IEA) may still offer us something and help us understand how humans read student writing.

Certainly most, if not all, of the pilots and studies reported on in the heart of the collection provide evidence for computers as inadequate for writing placement and assessment purposes, but a few essays move away from indicting computers and assessment technologies outright and suggest understanding them as a part of any assessment. Essays by Haswell, Whithaus, and Brent and Townsend suggest we not simply throw out the electronic baby with the digitized bathwater. They each tacitly ask: are there ways we might understand computers in the processes of writing assessment?
as still important despite their drawbacks of not being human, of not being
able to “read” student writing in the ways humans do?

The first of these undercurrent essays is chapter 4. Haswell asks a
localized question that McAllister and White ask in the first chapter: how
did we get in a place where educational institutions are clamoring to buy
computer programs such as ACCUPLACER or WritePlacer when there is no
knowledge of their validity offered? The heart of Haswell’s critique is his
use of the concept of the “black box,” which he gets from Bruno Latour’s
1987 book, Science in Action: How to Follow Scientists and Engineers through
Society. In short, the black box of science, and by extension writing assess-
ment, explains Haswell, is “anything scientists take on faith.” In the realm
of computer-assisted writing assessment, it is “any construction, hardware
or software, that one can operate knowing input and output but not know-
ing what happens in between” (68). Reiterating Latour, Haswell explains
that writing assessment, from the human kind to the computer-assisted, is
built on a series of black boxes that should be examined carefully, but often
(traditionally) have not. The “crucial black box,” says Haswell, “the one that
writing teachers should want most to open, is the meaning of the final holistic
rate—cranked out by human or machine,” since all holistic scores account
for only about 9% of the total variance in any particular criterion (72). He
concludes by making a number of calls, which mostly amount to examining
the black box(es) of computer-assisted writing assessment—not to destroy
them, but to understand them and look for potential uses, or to reconstruct
them in more useful ways. In effect, Haswell calls for validation research on
computer-assisted writing assessment, something not done to date.

As a different reviewer of this collection states, most of the essays that
follow Haswell’s try to do just what he suggests: validate computer assisted
essay scoring and placement (Cumming 81). And as mentioned already,
no technology meets the claims made by their makers, nor do they meet
the minimum expectations placed on them by WPAs and administrators at
each site. The central validation studies in the collection come from McGee,
Matzen and Sorensen, and Maddox.

In chapter 5, Tim McGee draws conclusions from three experiments, or
“spins,” with the Intelligent Essay Assessor (IEA), an Internet-based program
touted by its producers as one that can “understan[d] the meaning of text”
(80). McGee explains: “I wanted to compare IEA’s notion of ‘meaning’ with
my own” (84). He comes up with three troubling findings: first, “global ar-
rangement is not part of IEA’s notion of ‘meaning’” (87); second, “meaning
appears to exist quite independent of any relationship to factual accuracy”
(88); and third, IEA did seem to account for “mechanics,” but it was unclear
what the construct meant since the sample essay and its nonsense variant
received the similar mechanics scores, with the nonsense essay getting only
one point lower in mechanics. McGee concludes that the meaning that IEA

92 Composition Studies
allegedly “understands” is not the same as the meaning most, if not all, humans understand (90). So one black box discovered and cracked open: meaning is different for computers.

In chapter 8, Richard N. Matzen Jr. and Colleen Sorensen report on a pilot study of writing placement based on ACT e-Write. They investigate what they term “fairness,” which appears to encompass both the validity and reliability of a test (131). In this short report, they offer e-Write and human rater correlations to several established tests of writing and reading used by the school, Utah Valley State College. They show that e-Write scores offer considerably weaker correlations to all tests than human raters. For example, when used to place students into writing courses, human raters have a .559 correlation to ACT English test scores and a .421 correlation to ACT Reading scores, while e-Write scores offer only a .290 correlation with ACT English scores and .192 with ACT Reading scores (137). While there would need to be more probing to understand how valid any of these placement decisions are, it’s clear—as Matzen and Sorensen conclude—that e-Writer is inadequate, at least in terms of its reliability and concurrent validity (with other ACT tests), for the job of placement. In chapter ten, Teri Maddox offers similar conclusions about the e-Writer’s use as a placement mechanism at Jackson State Community College. It failed all expectations, which were essentially along three dimensions: cost, time, and reliability (148). Another black box cracked: validity and reliability actually are in question.

The other two of the undercurrent chapters (Whithaus and Brent and Townsend), which come after Haswell’s chapter, are really about classroom applications and pedagogies. And this classroom focus may be why these two seem to create, with the Haswell essay, an undercurrent, an alternative inquiry that cuts against the refrain that’s constant in the rest of the essays.

In chapter 12, Carl Whithaus argues for a more nuanced stand than the one made by the Conference on College Composition and Communication, which is an unqualified rejection of computers as assessing or responding agents. Taken from Lee Cronbach’s and Brian Huot’s similar positions on validity as argument and the incorporation of assessment decisions’ uses to construct validity, Whithaus’s position is that the uses of computer program must be considered when deciding how valid they are for assessment purposes (166; 170). Computer-assisted writing assessment may have some uses after all. Whithaus offers two examples to back his argument, and makes a distinction between computers and software used as assessment “tools” and as assessment “media” (171), the difference being primarily in the uses to which the technology is being put. One interesting example he offers is the use of Microsoft Word’s grammar checker and the readability scores it produces. Whithaus’ students use these scores (produced by a Microsoft Word) in a metacognitive revision activity. Word is then an assessment tool that allows students to produce not just a revision of a paragraph but a
“metacommentary about the revised paragraph and the software’s reading of that paragraph” (173).

Similarly, in chapter 13, Edward Brent and Martha Townsend demonstrate and argue for a limited use of computer grading of student writing in a large-enrollment sociology course with a program that Brent develops. Ultimately, they show how a carefully and responsibly designed course can use computer-assisted essay grading, along with human readers, to encourage revision and writing-to-learn activities. Additionally, as Whithaus argues in the previous chapter, Brent and Townsend argue that the CCCC’s “Position Statement on Teaching, Learning, and Writing Assessment in Digital Environments,” especially its all or nothing language, should be reconsidered (197). Whithaus and Brent and Townsend offer a tactic, inquiry into uses, for looking more carefully into the black boxes of computer-assisted assessment.

These three chapters (Haswell, Whithaus, and Brent and Townsend) make the collection worth reading, but should be read next to a few other chapters in the collection. Doing this can help teachers and WPAs consider their own arguments for or against using computer-assisted writing assessment at their local sites: Anson’s discussion (chapter 3) of how computers and humans make meaning of text, and thus how each must read it; Condon’s discussion (chapter 15) of “systemic validity” (215) that accounts for all that an institution loses when its people no longer control writing assessment; and Broad’s closing chapter (chapter 16) on the distinctions made between artificial and human intelligence, which create the domains in writing assessment of each other (228-29). Overall, however, the refrain is clear. Machines cannot read student writing in ways we need them to, and even if they could, students don’t learn well by writing solely to computers. Yet the undercurrent is also clear, at least to me. There still may be some uses for computer-assisted writing assessment, and there may be no other choice for us in our futures as writing teachers and WPAs.

Finally, as useful as the previously mentioned chapters are, the collection seems to go no further than Brian Huot’s 1996 essay, “Computers and Assessment: Understanding Two Technologies,” does, which appears in the collection’s references, but I don’t recall any substantive discussion of it. This is a drawback, since surely much has happened since the mid-90s. More importantly, most in the present collection do not acknowledge or address (accept arguably Haswell, Anson, and Broad) a core premise of the book, that what is at issue is a paradox of technology. We already use and need technologies of assessment, yet we are fighting against certain kinds of technologies because they take us in different directions, shape our practices, assumptions, student arrangements, and working conditions in ways we do not value enough to pursue. Drawing on Andrew Feenberg’s “substantive” theory of technology, Huot argues that not only are both writing assessment and the computers and software that we use to do writing and assessment
tasks technologies, but each “constitutes” a “social system” (235). In other words, as teachers and WPAs, we are always using technologies to do assessment, always “assisted” in our writing assessments by technologies of various kinds—in fact, our assessments are technologies themselves—and that in turn these technologies are constructing social systems and black boxes that structure assessment, student arrangements, our jobs, our notions of our students’ competencies, our pedagogies, our classes, our world. While this insight is not identified in Ericsson and Haswell’s collection, it’s worth keeping in mind as one reads this necessary book. And given the overly enthusiastic collection by Shermis and Bernstein, this collection is an important counterweight to balance the scales.

_Fresno, California_

**Works Cited**


 Reviewed by Cassandra Parente, Texas Christian University

As we become more enmeshed in an information economy, the places of composition studies continue to expand beyond classrooms and into actual and virtual public spheres. Coinciding with this cultural transition, the direction of our work within our classrooms and our scholarship is also changing: academic discourse is losing ground to public discourse, studies leading to generalized findings are being replaced with studies focusing on the situatedness of writing, and there is a growing sense that compositionists need not only be informed pedagogues and engaged scholars, but also socially-aware, active citizens. Responding to these economic and disciplinary shifts, editors Christopher J. Keller, Christian R. Weisser, and the fifteen contributors to _The Locations of Composition_ offer the concepts of place, space, and location as a means of exploring these issues.