INTRODUCTION

Patricia Freitag Ericsson and Richard H. Haswell

We're in the fifth year of the twenty-first century and the Parliament of India, several universities in Italy, and four Catholic churches in Monterrey, Mexico, all have bought cell-phone jammers. Meanwhile in the State of Texas, USA, the State Board of Education has decided that students who fail the essay-writing part of the state's college entrance examination can retake it either with ACT's COMPASS tests using e-Write or with the College Board's ACCUPLACER tests using WritePlacer *Plus*. Though dispersed geographically, these events have one thing in common. They illustrate how new technology can sneak in the back door and establish itself while those at the front gates, nominally in charge, are not much noticing. All of a sudden cell phones are disturbing legislative sessions and church services and allowing students to cheat on examinations in new ways. All of a sudden students can pass entrance examination essays in ways never allowed before, with their essays scored by machines running commercial software programs. How did this technology happen so fast?

And where were educators when it happened? We will leave the MPs in India and the deacons in Mexico to account for themselves, but as for automated essay scoring in the State of Texas, college educators can only throw up their hands. The decisions on e-Write and WritePlacer *Plus* were made by state government officials and industry lobbyists with no input from writing experts or administrators in higher education. The Texas step toward machine grading may not be typical so far. But in the near future there will be plenty of like steps taken everywhere in academe.

The analysis and scoring of student essays by computer—the history, the mechanisms, the theory, and the educational consequences—is the topic of this collection of essays. It is an understatement to say that the topic is rapidly growing in importance at all levels of the educational enterprise, and that the perspective on it has been, up to this point, dominated almost exclusively by the commercial purveyors of the product. Other than the notable exceptions of articles by Dennis Baron in the *Chronicle of Higher Education* (1998), Anne Herrington and Charles Moran in *College English* (2001), Julie Cheville in *English Journal* (2004), and Michael Williamson in the *Journal of Writing Assessment* (2004) and

the occasional sparsely attended conference presentation, the response on machine scoring from the academic community, especially from writing teachers and composition scholars, has been silence.

This book adds some voices from the academic community to the conversation, hoping thereby to jump-start a productive debate from educators about the machine assessment of student essays. With contributions by pioneers in writing assessment and by other scholars, the volume opens the discussion to broader audiences and to nonproprietary voices. It considers theory, practice, experiences, trials, lore, and data from the postsecondary side of machine scoring, especially from teachers and students. The essays focus largely on the postsecondary scene, but their implications move in all educational directions. Educators, administrators, and academic researchers provide background and understanding of machine scoring that will make productive sense of it to colleagues, students, administrators, legislators, and the interested public—to better shape, we hope, the way this new instructional technology will be used at all levels.

We even dare to hope that the entrepreneurs might benefit from these pages. Before the publication of this volume, the only book-length treatment of machine assessment came from the machine scoring industry. Mark Shermis and Jill Burstein's edited collection, *Automated Essay Scoring: A Cross-Disciplinary Perspective* (2003), advanced arguments supporting machine assessment technology—arguments from a group of authors almost all involved in the production and sale of the machinery. Despite the subtitle's promise of a "cross-disciplinary perspective," the voice of the academic world is almost completely muted, and largely omitted from discussion are many educational issues impacting higher education—historical, linguistic, social, ethical, theoretical, pragmatic, and political. Our volume seeks to fill in these lacunae. Our primary goal, however, is not to counter industry viewpoints, solely to cast a con against their pro. 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." 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.

The need for a wider audience is urgent because machine scoring programs are making rapid inroads into writing assessment. Teachers at every level are encouraged to use online writing software so their

students can pass standardized tests that will be graded, at least in part, by assessment machines. Web-based promotion for Educational Testing Service's Criterion (2004) promises student writers "an opportunity to improve skills by writing and revising in a self-paced, risk-free environment" and suggests that Criterion is a perfect tool for testing and assessing writing as well. Vantage Learning's Web site (2005a) touts its product benefits: from learning to write to assessing writing. Vantage "offers several solutions to aid educators in meeting NCLB [No Child Left Behind] requirements, from MY Access!, our online writing environment that has been proven to increase students' writing proficiency, to our customized solutions that are used in statewide assessments in the states of Texas, Oregon, Pennsylvania and Virginia." Claims by both these companies are accompanied by glowing praise from administrators and even from some teachers.

Two- and four-year colleges and technical schools are turning to online computer essay scoring in large numbers in order to place incoming students. One of the most popular of these placement machines is the College Board's ACCUPLACER, whose fact sheet (2005) claims its technology will "test students more accurately, with fewer questions in less time, and with immediate results"—and thereby make the whole campus happier!

> ACCUPLACER appeals to all members of your campus family.
> - Students find it less stressful and time-consuming, more accurate and immediate.
> - Faculty have more options, and find it more reliable, valid, and accurate.
> - Test administrators need ACCUPLACER because it is easy to use, accurate, reliable, and valid.
> - Institutional researchers appreciate the easy access to student and performance data.

ACCUPLACER's marketing campaign has been remarkably successful. In October 2004, Suzanne Murphy of the College Board told us that although she could not reveal which college and universities are using ACCUPLACER, she could tell us that "there are over 900 colleges that use ACCUPLACER in all of the U.S. states as well as a number of Canadian colleges and other colleges around the world." In addition, she was willing to let us know: "Last year over 5,000,000 tests were administered." Although ACCUPLACER once offered "traditional" (i.e., human) as well as electronic assessment of writing samples, that

option is no longer available, and all ACCUPLACER scoring is done electronically through the WritePlacer *Plus* technology.

University teachers across the curriculum are using commercial computer evaluation programs to grade essay exams and term papers. With funding from a Pew Grant in Course Redesign, Florida Gulf Coast University (2005) recently implemented an "automated essay assessment capability" for a general-education course, Understanding the Visual and Performing Arts. This automated assessment is provided by Knowledge Analysis Technologies (the creators of the Intelligent Essay Assessor) and allows for what the Pew Grants demanded: a redesign of "instructional approaches using technology to achieve cost savings as well as quality enhancements." These grants were focused on "large-enrollment introductory courses, which have the potential of impacting significant student numbers and generating substantial cost savings." This Florida university is not alone in its use of the Intelligent Essay Assessor or other programs like it in content area courses.

Two-year schools especially have been attracted to the technology. In 2000 Elisabeth Bass, a college teacher in New Jersey, informed writing program administrators of her intuition that "virtually every community college in the state has moved to ACCUPLACER." Bass's guess proved not far wrong. In a 2004 study of the placement practices in twenty-four New Jersey colleges and universities, Ramapo College testing coordinator Wanda G. Kosinski (2003) found that 70 percent of the institutions used computerized assessment for placement. Of those 70 percent, ACCUPLACER was used by fifteen (or 62.5 percent). Of the institutions in her study, Kosinski found that 69 percent of the two-year schools were using ACCUPLACER, Criterion, or COMPASS (all testing batteries containing computerized assessment machines). But many four-year schools have also been attracted to the "substantial cost savings." Of the four-year schools Kosinski gathered data from, 54 percent were using ACCUPLACER, giving ACCUPLACER over 62 percent of the computerized testing share in this study. Hers is not the only state where this is happening, and ACCUPLACER is not the only product available. ACT's COMPASS tests will probably be outselling ACCUPLACER in a few years.

Machine scoring no longer has a foot in the door of higher education. It's sitting comfortably in the parlor. In K–12 schools, machine scoring has become even more of a permanent resident, heavily promoted for use in grade-promotion examinations, in graduation examinations, in practice for mandated state testing, and in grading and responding to

course writing as a relief from the "burden" put on teachers. But that is a topic for another book.

At whatever academic level, machine scoring of student essays has been admitted with hardly any questioning of its true academic credentials. We stress *academic.* The entrepreneurs have thoroughly validated their software in terms of instrument validity, test equivalency, interrater reliability, and cost efficiency. What they have not done and what educators have little done is validate the software in terms of instructional adequacy, viability, and ethics. We have intentionally avoided providing much "how-to" advice, believing that this volume will serve as a sourcebook and a springboard for those interested in a critical look at the educational impact and implications of machine assessment of essays. Specific institutional circumstances will impact how administrators and teachers will respond to machine scoring. Our hope is that this book provides the critical resources for those responses. It will make a start in answering hitherto unbroached questions about the history of machine scoring, its consequences in the classroom, the ease with which writers can fool it, the reaction to it by students and teachers, and the authenticity of its "reading" of student writing. But we think this volume will do more. Even a quick glance through the chapters reveals that they raise questions that will help set the future agenda for debate and action on automated essay scoring.

- Who are the stakeholders, what are their interests, and why have some—teachers and students—largely been left out of the conversation? (McAllister and White)
- Can machine analysis programs understand the meaning of texts? (Ericsson)
- What are the capabilities and limits of a computer's ability to interpret extended discourse? (Anson)
- Why have composition teachers been so blasé about computer analysis of writing? What has been their complicity? (Haswell)
- How easy is it to fool the machine? (McGee)
- How closely do the score results of grading software match the careful evaluation of writing teachers? (Jones)
- How do students react when they find out their placement essays are being graded by a computer? (Herrington and Moran)
- What is the actual success, the adequacy, of automatic scores in placing students? (Matzen and Sorensen; Ziegler; and Maddox)
- How does use of machine scoring for placement affect the role of the writing program coordinator? (Corso)

- Are there ways that computer analysis of student writing can be effectively used in composition classes? (Whithaus)
- Are their legitimate spaces for machine analysis in a curriculum devoted to teaching higher reasoning and critical thinking? (Brent and Townsend).
- What is the educational or language ideology promoted by the machinery? (Rothermel)
- What do we lose if we let computers score student writing? (Condon)
- Where will machine scoring lead the teaching profession: to greater or less control over our courses, to more or less success in instruction? (Broad)

Although we have presented these essays without topic partitions, the book moves from historical and theoretical issues through concrete applications and problems to future ramifications. It ends with two pragmatic tools that we think will help everyone move toward a productive continuation of the discussion: a bibliography of the machine scoring literature, 1962–2005, and a glossary of terms. We think the questions that our authors raise and answer wrestle with the main truths (theoretical and empirical) and the main consequences (instructional and ethical) of machine scoring of student essays.

It is worth asking, finally: what have been the consequences of the lopsided production of discourse seen so far on automated scoring of essays? Our authors' questions are most readily asked by teachers and students, not by politicians or business people. None of these topics are explored by Shermis and Burstein's *Automated Essay Scoring* (2003), which, for instance, reports not one completed study of the instructional validity of machine scoring. In fact, on the issue of automated essay scoring teachers and students have not been encouraged to ask questions at all and sometimes have been systematically excluded from forums where their opinion should have a voice and might have an appeal. In *The Neglected "R": The Need for a Writing Revolution* (2003), authored by the National Commission on Writing in America's Schools and Colleges and published by the College Board, teachers and students are excluded from the agenda for "revolution." Instead it is recommended that "the private sector work with curriculum specialists, assessment experts, and state and local educational agencies to apply emerging technologies to the teaching, development, grading, and assessment of writing" (30). The essays in this volume stand as a response from the very stakeholders

that the National Commission on Writing omits. We hope they mark the start of a different revolution, in which the people most affected by this particular "emerging technology" have a say in that future.