Professionals Against Machine Scoring of Student Essays in High-Stakes Assessment

We call for schools, colleges, and educational assessment programs to stop using computer scoring of student essays written during high-stakes tests.

Every year hundreds of thousands of students write essays for large-scale standardized tests. The scores are used in life-changing decisions. Students are accepted into, placed within, and rejected from educational programs. Graduates are hired or not hired. Teachers are qualified, evaluated, promoted, and fired. Learning institutions are compared, accredited, and punished. Yet in a major disservice to all involved, more and more of these essays are scored not by human readers but by machines.

Let's face the realities of automatic essay scoring. Computers cannot "read." They cannot measure the essentials of effective written communication: accuracy, reasoning, adequacy of evidence, good sense, ethical stance, convincing argument, meaningful organization, clarity, and veracity, among others. Independent and industry studies show that by its nature computerized essay rating is: *trivial*, rating essays only on surface features such as word size, topic vocabulary, and essay length; *reductive*, handling extended prose written only at a grade-school level; *inaccurate*, missing much error in student writing and finding much error where it does not exist; *undiagnostic*, correlating hardly at all with subsequent writing performance; *unfair*, discriminating against minority groups and second-language writers; *secretive*, with testing companies blocking independent research into their products (See Research Findings supporting these claims).

In sum, current machine scoring of essays is not defensible, even when procedures pair human and computer raters. It should not be used in any decision affecting a person's life or livelihood and should be discontinued for all large-scale assessment purposes.

THEREFORE, we ask

1. *Legislators and other policy makers:* STOP mandating essay scores generated by machines to make crucial decisions such as grade promotion, academic placement, graduation, school ranking, school accreditation, or teacher qualification, promotion, and pay.

2. *Directors of large-scale assessments:* STOP using invalid computerized scoring of student essays (assessments such as the Collegiate Learning Assessment, the Test of English as a Foreign Language, or those planned by the Partnership for Assessment of Readiness of College and Careers, Smarter Balanced Assessment Consortium).

3. *Governing boards, administrators, and registrars:* STOP buying or accepting machine scoring of essays until you can prove it is valid, equitable, and worth stakeholders' money.

4. *Schools:* STOP buying automated essay scoring services or programs in the counter-educational goal of aligning responsible classroom assessment with such irresponsible large-scale assessment.

5. *Colleges:* STOP making your applicants, many from lower socioeconomic backgrounds, spend substantial money on testing such as ACCUPLACER or COMPASS, whose automatic essay scores misplace students at unconscionable rates.

6. *Teachers:* STOP using the regressive data generated by machine scoring of student essays to shape or inform instruction in the classroom.