Chapter 14
Can Software Grade My Students' Papers?
Do I Want It To?

Catherine F. Flynn
https://orcid.org/0000-0002-0732-7976
Purdue University Global, USA

ABSTRACT

Writing remains a staple of academic evaluation instruments, and for good reason. Few other tools come close to writing in assessment of the student's ability to formulate coherent arguments, demonstrate critical thinking, and present explanations. Conscientious instructors mark-up and comment on student writing, evaluating its success in meeting content goals, structure, relevance of evidence, grammar, mechanics, and style. It is an exhausting and time-consuming process but remains the single best way to support developing writers and thinkers. Technology to detect patterns of errors and to offer feedback has evolved over time, but can the programs do what professors do? This chapter provides a status report on automated writing evaluation and its role in higher education. A balance can be struck between the efficiency of tech tools and professor judgment. Recommendations for automating and expediting the review of student writing are offered, with a focus on remote learning environments.

Author's Note: *Numerous phrases for the same concept of automated writing review have emerged over time. Automated Essay Scoring, Automated Essay Evaluation, Automated Writing Evaluation, Automated Essay Grading, Automated Essay Assessment, and Automated Essay Scoring are all synonymous. Automated writing evaluation (AWE) is the preferred term in this chapter as it encompasses all writing rather than the limited genre of the essay form.*

DOI: 10.4018/978-1-7998-7653-3.ch014

INTRODUCTION

Computer programs capable of myriad tasks have entered our daily lives, often with a warm reception. Improved vehicle safety, automated health monitoring, streamlined financial transactions and on and on, we increasingly relinquish tasks to programs that are more efficient, more resourceful, and less likely to become distracted than we are. It certainly seems like a quick step from a computer conducting major surgery to one assigning a numerical value to a student's writing assignment. But do professors assigning writing necessarily want such a program? Do students?

With the exception of personal journals or diaries, the purpose of writing is to communicate with an audience. Adopting widespread use of automated grading software would effectively remove the audience from the equation, asking students to write something that will not be read. Such a model would potentially suggest the professor's role is perhaps moot. Do we want to step over that threshold?

Assessment is and needs to be a central element in the educational process. Development and use of computer-based assessment systems has grown exponentially in the past few decades in response to both the number of students attending university, and the increasing numbers of courses being offered using eLearning applications. Automated tools can and should be embraced as complements to the process of supporting student learning through feedback, with the emphasis on "complement." Professors who evaluate the subtle nuances and developing voice of the writer must always remain central to the review of student writing.

AUTOMATED WRITING EVALUATION BACKGROUND

Research supports the role of automated writing evaluation as a practical solution to the onerous task of manually grading student writing. Despite this potential value, over the past half century, many products have emerged and disappeared, defeated by the many challenges of determining validity and reliability, evaluating semantic content, and providing useful feedback (Zupanc & Bosnic, 2015). Writing assignments that best fit the automated review system are prompt-specific, meaning all are on the same topic. Human assessors score the writing to construct a learning set that is then used to develop the scoring model of the automated evaluation. The model is then applied to score new, ungraded writing. The performance of the model is validated by determining how well it replicated the scores assigned by human graders (Parra & Calero, 2019; Zeide, 2019). As the pool of writing samples increases, scoring efficiencies also, theoretically, become more accurate and thereby more useful. Inherently problematic in application of automated writing evaluation is the prompt-specific requirement. Student writing remains the single best way to assess student learning because of its demands on presentation of complex functions. Unlike objective questions, good writing assessments demand the ability to recall, integrate, organize, and present, all challenging actions. The measure of these outcomes, corresponding to the higher levels of Bloom's (1956) taxonomy - evaluation and synthesis – brings the greatest value to the student written response. Prompt-specific, defined answer questions do not provide the depth and value of complex writing assignments (Yu, 2019).

Writing review is a complex process that demands analysis, summary, and judgement of expertise. By definition, automated writing evaluation (AWE) is the task of building a computer-based system that reduces the demands on human raters as much as possible (Parra & Calero, 2019). AWE remains a challenging tool to develop because of its reliance on not only grammar, but also semantics discourse, and