

## AUTOMATED WRITING EVALUATION

Demeshko M.V.

Krasnova T.I. (scientific adviser)

National Research Tomsk Polytechnic University  
solarfreckle@gmail.com

### Introduction

Academic writing is a core aspect of university study. Writing is normally the last of the four skills acquired and is viewed by students and teachers as the most difficult area of second language use. In teaching, as well as in testing, much attention is given to students' writing efforts. Second language writing research has become increasingly sophisticated requiring researchers to investigate various theoretical and methodological perspectives as well as practical issues that arise in the process of research. Nowadays such new approaches as automated writing evaluation (AWE) or automated essay evaluation (AEE) are of great relevance.

Researchers point out one of the greatest problems in writing classes: teachers sometimes become frustrated by a large number of essays to be evaluated and limited time for this. As a result teachers rarely give such assignments to students depriving them of the possibility to develop writing skills. There came an understanding of a strong need for automated essay evaluation systems, which would facilitate evaluation process and save teachers' time for real-time personalized feedback during the learning process.

This technology was originally designed to reduce the heavy workload of grading a large number of student essays. The first automated essay scoring system (which is an ancestor of AWE systems), Page Essay Grade (1967), was developed by Ellis Page. He used multiple regressions to associate target essay with a set of essays on the same topic. This set had been scored by other English teachers. Pioneering work in the related area of automated feedback was initiated in the 1980s with the Writer's Workbench which worked in conjunction with Microsoft Word. A very successful application was created by Pacific Metrics in 2007. It was called constructed response automated scoring engine (CRASE®) and provided immediate and accurate scoring of essays.

Early AWE programs used simple style analysis of a text. Since the mid-1990s, the development of AWE systems has been improving rapidly. And now, newly designed systems can boast of the ability to conduct sophisticated analysis.

### Opposing Views

There are many views on such systems and sometimes they are completely opposite. On the one hand, it is known that teachers spend from 15 to 30 minutes checking an average essay and correcting mistakes in it. So, if the teacher asks the group of at least twenty students to write an essay, he or she will spend a lot of time checking them all. From that point of view, AEE systems can ease teacher's burden.

But on the other hand, the most popular argument against these systems is that they are supposed to perform robotic inspection or robograding, as the critics like to call it. Therefore AEE systems cannot replace the teachers' work.

At the same time, if we delve into the causes of emergence of these systems, we will find that everything revolves around the same points of view. Before the question about AEE had been actively discussed, there were questions about writing itself and these questions are topical today. This situation looks like a closed circle. First of all, students should have more practice in writing. Writing skills are extremely important in the era of global information systems. We all live with the information noise around us and the ability to convey your thoughts to the others is really useful.

But we can see that the amount of practice which is available to students is limited by the restrictions of interaction between teacher and student. Teachers do not have the time to respond quickly and thoughtfully to all students, particularly in case of large amount of essays.

Thus, we see the situation when students need practice and one teacher cannot provide them with it in sufficient volume.

And here comes the automated writing evaluation. The whole idea of these systems assumes solving the described problem. They save teachers' time and allow students to get a quick report on their essays. The distinguishing features of AEE are automatically calculated score and formative feedback.

### Human vs. Automated Evaluation

Traditional cycle of interaction between students and teachers includes: assigning an essay topic, students' writing and submitting essays, teachers' grading and commenting on essays and returning results to students.

Cycle with using AEE-systems should include slightly different steps:

1. Teacher assigns an essay topic from the given list. Sometimes teacher writes his or her own topic, but then scoring can be less accurate.
2. Students write and submit their essays with the help of AEE-interface and they are available to the teacher.
3. Special software scores the essay and offers feedback. Steps 2 and 3 may be repeated a few times at the discretion of the teacher. Moreover, teacher can add comments in addition to AEE-report.
4. Teacher grades students' essays and adds comments. All this information is available to students.

So, AEE systems not simply organize interaction, but at the same time take upon themselves some teacher's workload.

### **AWE and AES Systems**

As for AWE organization, these systems are usually described as consisting of two components. First component is scoring engine and second is feedback engine. They are separated because of different aims. Second component is the main difference between AWE systems and automated essay scoring (AES) systems. AES itself can only score some writing features such as grammar and mechanics. As you can see the AES can be scoring-engine-part of AWE. AES can provide feedback too, and it will include comments about spelling, grammar, mechanics, usage and style. AES is very accurate, but it means that with its help students can improve in writing mechanics and structures, but not overall quality.

And AWE systems inherit this disadvantage, because while facilitating practice and improving students' motivation, they can miss some rare non-mechanic issues, but for individual writer these issues may be very frequent.

How scoring engines can evaluate essays? Scores are generated with the help of artificial intelligence methods, for example: statistical modeling, natural language processing and latent semantic analysis. Generally, scoring engines are combinations of computational linguistics and statistical modeling.

There are different ways to provide formative feedback. The traditional way is to rely on linear multiple regression models between text features scores. But there is one way, which is more progressive. This way implies hierarchical classification. It affords the opportunity to provide feedback at different levels, concentrated on different linguistic features.

There are two sets of software tools in AWE that do not use artificial intelligence: a limited form of a learning management system (LMS) and a limited form of an online writing lab (OWL). With the help of LMS teachers can manage writing assignments, students can review their writing portfolios, and district administrators can track progress by reports on writing by teacher, student, grade, school, or other criteria. OWL features help to connect with writing aids (online dictionaries, graphic organizers, writing rubrics with sample essays).

### **Modern Applications**

Modern AWE systems (such as *Criterion* by Educational Testing Service and *MY Access!* By Vantage Learning) use sophisticated analysis tools: lexical complexity, syntactic variety, discourse structures, grammatical usage, word choice and content development. With the help of these tools AWE systems provide immediate scores and diagnostic feedback in various aspects of writing. AWE systems can be used as a source of auxiliary evaluation or summative assessment.

There are more and more online services that use AWE technology. For example, there are free tools

such as *Grammarly* (by Mark Fullmer), that allow you to check your essay and get feedback report about problem areas and number of errors in a particular area of writing.

### **Conclusion**

In this article we have given an overview of the challenges and opportunities in the area of automated writing evaluation. Despite all the challenges there is a growing research interest in this field because of the potential of real impact for language learners all over the world. Recent, innovative research in error detection and writing evaluation is of great value nowadays.

The existing AWE systems already show significant progress, they present realistic and convincing results but still they are far from human's proofreading. Researchers offer different solutions for the design of a solid evaluation method but they have little consensus on this field and there are open areas for research.

The main objective for AWE systems developers is to increase the efficacy of these systems for improving the writing of actual users. Some of AWE systems become not only an assessment tool, but also a writing assistance tool. It is very useful for those students, who want to improve their writing skills in complete absence of a teacher. They can benefit from the feedback and corrections that such systems will provide.

To sum it all up, the use of AWE systems is not a simple black-and-white issue. This issue involves a complex combination of factors concerning software design, pedagogical practices, and learning contexts. The pace of development of these systems is rising and sooner or later AWE systems can achieve the accuracy of verification, which will be indistinguishable from human examination.

### **References**

1. Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94-112
2. Crossley, S., Roscoe, R., McNamara, D. (2013). Using Automatic Scoring Models to Detect Changes in Student Writing in an Intelligent Tutoring System, 208-213.
3. Grimes, D., Warschauer, M. (2010). A Multi-Site Case Study of Automated Writing Evaluation. *The Journal of Technology, Learning and Assessment*, 8 (6), 4-43.
4. Roscoe, R., Kugler, D., Crossley, S., Weston, J., & McNamara, D. (2006). Developing Pedagogically-Guided Threshold Algorithms for Intelligent Automated Essay Feedback, 466-471.
5. Shermis, M. & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. New York: Routledge.