

## What Is *Criterion* and *e-rater*, and How Can They Be Used in a Classroom?

Youngsoo Song

*Teacher's College, Columbia University*

It is not uncommon for teachers to worry about delivering effective instruction and employing good pedagogical methods to maximize student learning. However, this becomes more problematic with the growing number of students per class and by the shortage of teachers who can accommodate the students. An immediate concern that will be discussed in this short piece is assessment, feedback, and scoring on writing from ESL/EFL students. A teacher becoming fatigued by an overwhelming amount of work is one concern, but it is often followed by unreliable and inaccurate assessment and grading of the students' work. With little control over the shortage of teachers and the continuous enrollment of students, there must be a solution to alleviate the burden of teachers. This is where automated essay scoring (AES) tools such as *Criterion* by ETS may prove to be useful for teachers to effectively and efficiently provide meaningful feedback and grades. To elaborate, *Criterion* is an online platform providing convenience and practicality for its users, but it is the technology (also referred to as engine) of *e-rater* embedded in *Criterion* that provides automatic feedback on surface features of text such as grammar and scores

*Criterion* is an online tool that provides instructors a platform for their students to submit writing assignments. When submitted, *Criterion* will generate automatic feedback on mechanical errors, and will also provide a score with great reliability and accuracy. For students, they will be able to submit their writing assignments online through *Criterion* and receive immediate grades, within a few seconds and feedback on mechanical errors in addition to in-depth instructor feedback (*Criterion* provides a separate section for this). When research in AES became popular in the 1990's with rapid advancement in computer technology, *e-rater* was developed by the Educational Testing Service (ETS) in the 1990's (Burstein, Brade-Harder, Chodorow, Hua, Kaplan, Kukich, Lu, Nolan, Rock & Wolff, 1998; Burstein, Leacock, & Swartz, 2001) and was first commercially used in 1999. *e-rater* uses statistics and natural language processing (NLP) to extract and analyze linguistic features from essays. Not only can it provide descriptive feedback of the writer's mistakes and style, albeit formulaic and limited, but also a score (6 point scale) based on the writer's grammar, mechanics, word use, style, lexical complexity, length, organization and more. Ultimately, *e-rater* attempts to look at certain features and provide feedback, along with rating as a human rater would. As wonderful as *e-rater* sounds, it still has some shortcomings when it comes to discourse, topic content, linguistic idiosyncrasies, and insufficient input such as two to three sentences (Higgins, Burstein, & Attali, 2006). In fact, some researchers claim that *e-rater* is neither reliable nor valid as ETS claims it to be (McCurry, 2010; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002).

Despite mixed findings on *e-rater*, teachers and students can still benefit from using *Criterion*. To reap the benefits, it is important to note that AES tools cannot replace human feedback and rating, and therefore, *Criterion* should be used as an augmenting tool at the discretion of an instructor. With objectivity, speed, and automaticity, *Criterion* can lessen a teacher's workload by reducing repetitive and trivial error corrections, and allow both students

and teachers to focus on other important issues. In addition, *Criterion*'s integrated handbook and online references, platform for feedback, and digital recordings of student performance can be very useful in pedagogy. Aside from formative assessment, *Criterion* can also generate reliable holistic scores for teachers to use at their discretion. If teachers feel that they simply want their students to practice writing in a safe environment without the consequences of grades, *Criterion* provides the option to not generate any scores. In other words, ample training and experience will be required to make good use of *Criterion*.

However, one must be aware that *Criterion* still has limitations—it does not assess content or logic. For this reason, it is important for teachers to provide ample in-depth feedback in addition to *Criterion*'s rudimentary grammar error detections; this allows teachers to allocate their time to instruction or review sessions to cover more complicated and abstract topics. In addition, *e-rater* is still prone to 'test-wiseness'. In other words, with experience and strategy, students can take advantage of the technology, and earn scores that do not truly reflect their English proficiency. In sum, as long as both teachers and students use *Criterion* in a critical manner, it will prove to be a useful tool to enhance learning for students and raise practicality and better time management for teachers. With further development on *e-rater*, followed by empirical studies, *Criterion* may gain more popularity and utility in ESL/EFL classrooms.

## REFERENCES

- Burstein, J., Braden-Harder, L., Chodorow, M., Hua, S., Kaplan, B., Kukich, K., Lu, C., Nolan, J., Rock, D., & Wolff, S. (1998). *Computer analysis of essay content for automated score prediction: A prototype automated scoring system for GMAT analytical writing assessment essays*. ETS Research Report No. 98-15. Princeton, NJ: Educational Testing Service.
- Burstein, J., Leacock, C., & Swartz, R. (2001). Automated evaluation of essay and short answers. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.
- Higgins, D., Burstein, J., & Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. *Natural Language Engineering*, 12, 145-159.
- McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, 15, 118-129.
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: challenging the validity of automated essay scoring. *Computers in Human Behavior*, 18, 103-134.