Criterion Online Writing Evaluation

J. Elliott Casal

PRODUCT AT A GLANCE

<table>
<thead>
<tr>
<th>Product Type</th>
<th>Online Automated Writing Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language(s):</strong></td>
<td>English (Manuals available in English, Bilingual Writer’s Handbooks available other languages)</td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td>Beginner through advanced</td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td>Automated Writing Evaluation (scores and feedback), prewriting activities, essay writing, peer and teacher review</td>
</tr>
<tr>
<td><strong>Media Format:</strong></td>
<td>Web browser based</td>
</tr>
<tr>
<td><strong>Operating System(s):</strong></td>
<td>Windows® XP through Windows 7 supported</td>
</tr>
<tr>
<td></td>
<td>OS X® 10.6 or higher supported</td>
</tr>
<tr>
<td></td>
<td>iPad® with iOS 5.1 or higher supported</td>
</tr>
<tr>
<td><strong>Hardware Requirements:</strong></td>
<td>PC: Pentium® III or greater</td>
</tr>
<tr>
<td></td>
<td>Mac: G3-G5® supported (Intel chip used by reviewer)</td>
</tr>
<tr>
<td></td>
<td>Tablet: iPad (external keyboard recommended)</td>
</tr>
<tr>
<td><strong>Supplementary Software or Hardware</strong></td>
<td>Web-Browser (IE 8-10, Firefox® 18-22, Google Chrome® 28, Safari® 6.0 or higher supported)</td>
</tr>
<tr>
<td><strong>Printed Documentation</strong></td>
<td>Quick Access Guides (.pdf) for instructors and students</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td>ETS does not publish pricing information, but such information is available via direct communication with an ETS agent.</td>
</tr>
</tbody>
</table>

General description

Criterion® Online Writing Evaluation software, created by Educational Testing Service (ETS), is a feature rich online Automated Writing Evaluator (AWE). At the core is the e-rater® holistic scoring system, which was designed to be used for evaluating writing done under testing conditions’ (http://www.ets.org/criterion/
The e-rater scoring engine calculates holistic scores on a scale from 1 to 6 or 1 to 4. It is trained on a database of scored samples and analyses structural and linguistic elements of new submissions to calculate scores. Scoring functions are available for instructor created topics as well as writing tasks chosen from the extensive topic library, which includes hundreds of essays for grades 4–12, university level, or test specific conditions. Another central feature of the e-rater engine is that it provides learners with indirect (correct forms not provided) and coded (categorized) feedback for their writing.

Criterion software offers users an extensive set of features and choices. It allows instructors and students to manage multiple courses, writing tasks, and submissions. Instructors have a wide variety of options available in designing and customizing writing tasks (e.g., time limit, availability of resources, categories of feedback offered), and can input electronic feedback via comments or group students for peer review. From a student perspective, Criterion software offers basic word processing, spellcheck, and thesaurus interfaces for text construction, has built in prewriting and planning activities, and allows for chat-like dialogue with peer reviewers and instructors. Many of these options can be disabled at the instructor’s discretion. Indirect feedback, as mentioned above, is provided across multiple categories and coded further into error types. In addition, writer’s handbooks are accessible at multiple skill levels and in bilingual Spanish/English, Japanese/English, Korean/English, and Simplified Chinese/English versions. All planning activities, drafts, comments, peer interactions, and dialogues are stored in electronic portfolios and can be exported for later use. Additionally, a variety of reports can be generated to track progress of individuals or groups across multiple variables over time.

On their product website, ETS states that their Criterion software can positively impact learning processes. They note that use of Criterion software can increase the number of opportunities for students to write, receive feedback, and integrate feedback. At the same time, ETS believes that use of their software will increase efficiency and allow instructors to ‘teach higher level writing skills’ (https://www.ets.org/criterion/about/why/). They have other claims regarding its suitability for use in standardized test assessment and by administrators for making ‘data driven decisions’ (https://www.ets.org/criterion/about/why/), but this review will evaluate the usefulness of Criterion Online Writing Evaluation software for purposes related to learning rather than assessment.

**Evaluation**

**Technological features**

Criterion software is easy to access due to the browser-based nature of the application. No updates, plugins, or setting changes were required on any of
the tested browsers or machines during this review. Webpage load times are relatively short except for the brief, expected delay when submitting or saving content.

Similarly, navigating Criterion software is easy due to the generally open, uncluttered design. Most tasks, such as creating assignments or accessing reports, are as simple as ticking checkboxes or selecting from dropdown menus. In the same way, the word processing interface, as Figure 1 shows, offers the features necessary to accomplish most basic tasks while not distracting from the writing. The software is designed to mediate. There is an elegant, utility-based simplicity to Criterion software.

At the same time, some of these features, such as the peer-review and comment modules, could benefit from expansion. While Criterion software is easy to use, those who are not familiar with computer or Internet technologies may require aid. For assistance, ETS offers free quick-access guides, which offer step-by-step instructions for accomplishing tasks related to the central features, and a training course for three hundred US dollars. As noted on the Criterion home page, ETS also offers support via a toll free number and support mailbox.

**Figure 1:** The Criterion software interface places the focus on texts. Copyright © 2013 Educational Testing Service. All rights reserved.

The first of two primary features of Criterion software is the e-rater holistic scoring engine. The algorithms on which the e-rater engine runs are described in detail through research presented on the ETS website. Essentially, the scoring engine is trained on large numbers of previously scored essays and conducts a linguistic and structural analysis of new submissions. When essays display unusual characteristics (e.g., abnormal length or paragraphing), the
Learning Technology Review     149

user is notified that scores may be somewhat less reliable. It is fast and consistent, but ETS explains that scoring is not based on content and therefore ‘can be fooled by an illogical, but well-written, argument’ (https://www.ets.org/criterion/faq/). Naturally, the e-rater engine does not assess essays based on the same criteria as human readers. Li, Link, Ma, Yang, and Hegelheimer (2014), for example, found that the e-rater scores only ‘moderately and weakly correlated with the instructors’ rating’ (p. 76) in their longitudinal analysis of how instructor and e-rater scores correlated on essays from 67 students in a classroom context. The authors indicate that this finding supports similar classroom correlation studies, but it is important to note that human raters do not always correlate highly either. Nevertheless, Criterion software does include an explanation of scores, as seen in Figure 2. However, as is often the case with holistic rubrics, it does not indicate which parameters were met and which were not. Instead, as participants in Li et al. (2014) indicate, the explanation is ‘always the same’ (p. 75) for scores at each level.

Figure 2: The e-rater engine in Criterion software provides holistic scores and explanations

The other primary feature of Criterion software is the category based indirect feedback that the e-rater engine provides (Figure 3). After an essay has been scored, the user has access to an extensive yet digestible feedback report. Categories (i.e., organization & development, grammar, usage, mechanics, style) are displayed across the top with dropdown lists of subcategories. Clicking on a subcategory highlights each instance of the error in a salient shade...
of yellow. As such, feedback is indirect in that correct forms are not provided, but they are categorized according to error type. In addition to grammar, the software is capable of identifying repetition, spelling, and sentence boundary issues, among others, with a high degree of accuracy. Nevertheless, there are limitations. References and citations often bait Criterion software into identifying sentence fragments, for example, because the scoring models are not trained on source-based responses. Likewise, when grammatical errors occur with high density or create ambiguity, Criterion software cannot always correctly categorize the issues. Often these cases end up in the ‘Garbled Sentences’ category which may require more effort to resolve. To its credit, Criterion software can reliably signal issues and draw attention to patterns and trouble spots. Finally, the development and organization feedback is helpful but not as consistent. Criterion software can effectively mark single and even multiword transitions, for example, but may miss content oriented or phrase based transitions.

**Figure 3:** Feedback in Criterion software is indirect and coded

**Pedagogical features**

A considerable part of any writing course will be spent writing. Criterion software does little to get in the way of this process, and in some ways has the potential to enhance it. Depending on the course and level, the ETS topic bank may be of varying use to an instructor, and other options are worth noting. First, Criterion software comes with powerful and easy to use prewriting activities, as seen in Figure 4. These planning templates range from empty free-writing boxes to organized concept maps. Instructors can allow all prewriting activities or even assign a specific choice. By making prewriting activities digital, accessible, and easy to use, perhaps instructors can encourage students to make greater use of them. Criterion software offers instructors numerous options to customize the writing experience. Options exist to disable spell-check and thesaurus features, disable or limit automated feedback, or to link external resources. Through careful planning, instructors can customize resource sets available to learners, but successfully selecting the right settings
for a desired outcome requires careful thought. While Criterion software may lack some of the more advanced word processing or collaborative features of other technologies, the ease with which the writing experience can be controlled to fit learning objectives or contexts is a major advantage.

Perhaps the greatest contribution Criterion software can make to a writing course comes through its automated feedback. While scores may not correlate with students’ progress towards all course goals, they may have a motivating effect on learning. The indirect corrective feedback that Criterion software provides, explored in detail later, is immediate and can be used to set personal goals or track progress. As correct forms are not provided, learners must resolve issues themselves or through interaction with individuals or resources. EFL/ESL instructors can exploit this through collaborative activities. During this review, for example, students were asked to collaboratively review feedback reports. This must be done in person because feedback reports are not available through the peer review interface. Additionally, instructors can elicit particular structures or encourage learners to review error types which they are capable of fixing themselves. However, the holistic nature of score reports and lack of content assessment limits the long-term usefulness of learning via score reports alone.

![Image](image_url)

**Figure 4:** An ‘Idea Web’ is one of eight prewriting activities available in Criterion software

The communicative features of Criterion software are rather limited. While the inclusion of a chat-based dialogue system is noteworthy, there seems to be no easy way to choose the recipient of a message (instructor or peer-reviewer). In a similarly limiting fashion, instructors and peer-reviewers can only attach comments to single words in the text. There is no option to highlight, strike-through, attach a comment to multiple words, etc. These are valuable feedback
options available in other contexts which may be missed. In addition, it is surprising that, given the iterative nature of text submission in Criterion software, instructors cannot attach grading rubrics or feedback forms. In the end, Criterion software based writing activities can be highly controlled and include extensive feedback, but are typically single-user events and are focused on local rather than global issues. Instructors may welcome additional interactive and communicative features from Criterion software in the future.

Learner fit (design)

Criterion software can be appropriate for a wide variety of educational contexts. It is important to note that it is only available for English language writing. Nevertheless, it remains highly adaptable. The simple menus and limited text within the application make it accessible to beginning level students, while the feedback provided in some categories will remain useful throughout a student’s study of English. The quick-access guides are available only in text-heavy English, though, which may create an initial barrier for some low proficiency users. Criterion software, however, is a tool for classes. Students cannot directly subscribe and must be enrolled in courses connected to teacher accounts. Successful implementation of technological resources in learning contexts often depends on the effectiveness of learner training. Hubbard (2004) emphasizes the importance of ‘formal, sustained training in how to take operational competence in a given computer application and transfer that into learning competence’ (p. 51). The distinction between knowing how to use a tool and knowing how to use it for learning is important. In terms of Criterion software, students should be trained to be critical users who understand the strengths and limitations of Criterion software scores and who are willing to explore its recommendations. This, logically, requires that instructors themselves become critically aware.

Teacher fit (approach)

Broadly speaking, Criterion software’s role as a learning tool is based on its potential to provide student writers with useful feedback in useful ways and its potential to foster healthy and robust writing processes and practices.

As such, it is important to explore the potential usefulness of the two types of feedback the e-rater engine provides: a holistic score and a feedback report. ETS acknowledges, as previously stated, that the software can be tricked into giving a high score to a well-structured, nonsensical essay. On the one hand, this reminds us that the e-rater engine cannot read, but rather analyses texts linguistically and structurally based on statistical models. On the other hand, it highlights the importance of these structural and lexical elements in effective writing. Important as these features of language may be, it would be difficult to
argue that they are *more* important than the message itself. Ware (2011) effectively captures this sentiment in her concern that ‘focus on the more observable mechanistic and formulaic aspects of writing may be counterproductive if they lead teachers and students further away from writing purposefully for real audiences’ (p. 771). Across contexts, the emphasis of writing courses typically lies in effectively communicating with an intended audience. Consequently, the holistic score should be viewed with caution.

However, the usefulness of the categorized feedback report has a clear basis in scholarship. Written corrective feedback plays a significant role in student-teacher interactions in writing courses. While scholars have explored the benefits of both direct and indirect feedback, Ferris (2011) summarizes classroom error correction studies as demonstrating ‘the overall long-term superiority of indirect feedback’ (p. 32). However, she later acknowledges limitations, such as student ability and level of explicitness. Nevertheless, indirect feedback such as that provided by Criterion software may have potential benefits to engagement and learner autonomy. Regardless of the benefits instructors and researchers attach to these forms, Haupt and Bikowski (2014) highlight the importance of learner preferences and perceptions and that they may change over time. Importantly, they explain that instructor actions may be able to influence these evolving perceptions (p. 299). This may mean that learners can develop the ability to learn from and appreciate indirect coded feedback if instructors create the proper conditions. As such, the indirect corrective feedback report, though more targeted to local level issues, may contribute to learning.

With regards to developing healthy writing habits, there is reason to believe that Criterion software may have a positive impact on learner behavior, notably concerning prewriting and proofreading. Instructors and manuals often emphasize the importance of these practices, yet learners often struggle to integrate them into their own behaviors. The digital form and simplicity of Criterion software’s built-in prewriting options may positively impact student motivation. Generally speaking, Healey *et al.* (2011) note in a summary of CALL findings, that technology use often has a positive effect on motivation in language learning (p. 9), and these effects are likely to be present here. In terms of proofreading, the iterative ‘multiple drafting’ (Cotos & Huffman, 2013: 77) nature of Criterion software submissions can be expected to have a positive impact as well.

**Summary**

Choosing to use Criterion Online Writing Evaluation software is a major commitment due to the substantial annual cost and extensive in-class training time involved. In addition, improper use or overemphasis of the linguistic
elements the feedback highlights may be counter to course goals. However, with proper care and training, Criterion software offers a wide variety of features that may lead to more tailored writing conditions, more autonomous second language writers, and healthy writing processes. If the instructor and students remain aware of what an AWE is and is not, Criterion Online Writing Evaluation can be a powerful learning tool in ESL/EFL contexts and beyond.

Note
1. Criterion and e-rater are registered trademarks of Educational Testing Service (ETS).

References


Producer details
Educational Testing Services
660 Rosedale Road
Princeton, NJ 08541 USA
Phone: 1-609-921-9000
highered@ets.org
http://www.ets.org/criterion/
Reviewer information

J. Elliott Casal has a Masters in Applied Linguistics from Ohio University. He currently works with the English Language Improvement Program (ELIP) at Ohio University as an EAP Term Instructor, and the Coordinator of the ELIP Writing and Critical Reading Labs. He is also a Term Instructor in the intensive English program at Hocking College.