

# Automatically Score Essays with ETS's *e-rater*<sup>®</sup> Engine

The *e-rater*<sup>®</sup> engine is ETS's capability for the automated scoring of essays — the result of more than two decades of research supporting the development of construct-relevant and reliable scoring techniques. The engine has been in operational use since 1999 and is used to score millions of essays in a wide range of assessments.

The *e-rater* engine estimates the score for an essay based on features related to writing quality, including grammar, usage, mechanics, discourse coherence, sentence variety, vocabulary usage and organization, and development. The computational methodology underlying the system is natural language processing (NLP), which identifies and extracts linguistic features from stored electronic text. Measurement evaluations of its score predictions have been shown to be highly reliable; the *e-rater* engine often agrees with human ratings more strongly than human raters agree among themselves. It can also automatically flag responses that are off-topic or anomalous in other ways, so that they can be set aside for special consideration.

The *e-rater* engine is informed by an Evidence-Centered Design philosophy. This requires that the information used to assign a score should constitute valid evidence of the writing proficiency construct that an assessment aims to measure. Through research publications, ETS discloses the details of the engine's scoring methodology.

## Using the *e-rater* Engine in High- and Low-stakes Settings

In high-stakes settings, the engine is used in conjunction with human ratings for both the Issue and Argument prompts of the *GRE*<sup>®</sup> test's Analytical Writing section. Another example of the engine's high-stakes use is in conjunction with human ratings for the *TOEFL iBT*<sup>®</sup> test's Independent and Integrated Writing prompts. ETS research has shown that combining automated and human essay scoring demonstrates assessment score reliability as well as measurement benefits, while ensuring that a human reader stays in the loop.

In low-stakes settings, the *e-rater* engine is integrated into the *Criterion*<sup>®</sup> Online Writing Evaluation Service. This web-based, essay evaluation service is widely used as an instructional writing application in K-12 and higher education settings. Using the engine, the *Criterion* service offers immediate, individualized feedback about writing



errors; the presence and absence of discourse structure elements (e.g., thesis statement, main points, supporting ideas and conclusion statements); and style advice. This diagnostic feedback can be used by students, in conjunction with instructor feedback, to revise and resubmit an essay.

In addition, the engine's grammar, word usage and mechanics error detection features are used in ETS's *Writing Mentor*® tool, which is a Google Docs™ app. It is also the scoring engine behind practice assessments, such as the *TOEFL*® Practice Online tests and the *GRE*® *ScoreItNow!*™ service.

The *e-rater* engine is continually being developed and improved, with the aim of extending its ability to model important and challenging aspects of writing proficiency. Ongoing research aims to enhance the *e-rater* engine capabilities so that it can identify and evaluate the structure of an argument in an essay, as well as assess the creative use of language in student and test-taker writing.

#### Using NLP methods, the *e-rater* engine identifies features of essays related to:

- The presence of errors in grammar, word usage or the mechanics of writing
- The quality of discourse coherence in writing
- The presence of sentence variety in an essay
- The presence and development of essay-based discourse elements
- Style weaknesses
- Sophistication of vocabulary usage
- Idiomatic and appropriate use of collocations and prepositions

To learn more about the *e-rater* engine, visit [www.ets.org/erater/about](http://www.ets.org/erater/about)

