



How Are Essays Scored?

The PEG scoring engine automatically scores essays based on hundreds of student essays scored by humans. It seems like magic, but it is actually based on technology and human scoring.

A Single Set of Rubrics

To understand the scoring of essays in Utah Compose, it is important to remember that the characteristics of good writing are the same whether the writer is nine or twelve or eighteen years old. Sports are one example in which this is also true. Although the rules of basketball do not change as the players move from elementary school to high school, elementary school players are not expected to perform like those in high school. Still, coaches and spectators can identify excellence at each level. Essays in Utah Compose are the same.

Essays in Utah Compose are scored using models in our automated essay-scoring engine, PEG. The models are built using a multi-step process that begins with a committee of experts in writing. The committee chooses exemplars from among hundreds of essays for every grade band, every genre, every trait, and every score. One anchor set for style in narrative writing for the 3rd/4th grade band, for example, would include a high, middle, and low exemplar within each score point (1-5). When the essays for these anchor sets are being chosen, the committee focuses on interpreting the rubrics for one grade band at a time.

Once this process is complete, the anchor sets are used to train the expert scorers. These scorers receive training on each grade-band anchor set and concentrate on scoring one grade band at a time. Thus, an essay written by a third-grader is never compared to that of a fifth-grader.

Because the scoring is adjusted this way by grade band, students of all levels using Utah Compose can reference the same set of rubrics, making it easier for them to internalize the characteristics of good writing.

Human Scoring

Hand scoring hundreds of essays in each genre (Argumentative, Informative/Explanatory, and Narrative) for each grade band is necessary to build reliable models. Every essay is scored for each of the Six Traits of Writing by two highly trained, experienced scorers. Not only have these scorers been trained on the rubrics using the anchor sets, but they have years of experience scoring essays for many other projects as well.

How Does the Project Essay Grade (PEG) Scoring Engine Work?

Using these hand-scored essays, the PEG team builds models for each genre in each grade band (3-4, 5-6, 7-8, 9-10, 11-12). Grade bands are used because there is too little difference in student writing from one grade to the next to divide the models into separate grades. When a student submits an essay, PEG

compares the student's essay to the model for the grade band in which the essay is submitted. The Development of Ideas score, for example, is generated by comparing the development of ideas in the student's paper to the development of ideas in that particular model built using the human-scored essays.

Thus, PEG is really just matching up what is submitted to what the human scorers decided. When a student scores a 3, it means that the essay best matches the hundreds of original papers with scores of 3 for that genre in that grade band.

How Does PEG Score the Different Genres?

Originally, all essays were scored on one set of rubrics. Several years ago, the PEG team implemented scoring by writing genre. This allowed separate scoring of Argumentative, Informative/Explanatory, and Narrative papers using the specific rubric for that genre. Then models were built separately, as well. That means that when an argumentative paper is scored, it is ONLY compared to other argumentative papers for certain traits, and the same is true for informative/explanatory and narrative. The rubrics for the first four traits vary according to genre, but the rubrics for Sentence Fluency and Conventions are the same for all genres.

What Is Automatically Generated by PEG?

In addition to the scores for the six traits, PEG generates the spelling and grammar comments within the essay. This feedback is the result of "rules" that are incorporated into the engine. The PEG team endeavors to give enough feedback within the essay to support student practice and learning without overwhelming students. It is also important to balance the number of identified errors with the possibility of identifying errors that are not really errors. The English language is full of confusing and contradictory rules. For example, run-on sentences are particularly hard to differentiate from those that are simply longer and more complex. Word Choice is another area in which confusion can arise. You may notice that PEG has suggested changing a word in the essay to another similar sounding word. Sometimes the suggestions are appropriate, but when they are not, that is an example of a "false positive."

Since PEG is not human, it has to have very specific rules to follow. An "if, then, except when" is a difficult rule! For example, homophone errors, like using "bare" for "bear" and vice versa, are not always easily distinguished because some of those pairs can occur in similar contexts. If you have ever used the spelling and grammar check in Microsoft Word, you will see the same kinds of false positives. In general, the PEG team tries to limit the number of false positives, which sometimes has the unfortunate effect of allowing actual errors to escape detection.

PEG also provides Targeted Feedback, which is generated based on the specific student essay. Other feedback, such as recommended lessons, is more general, and appears based on grade band, genre, and trait score.

What if I Have a Question About Scoring?

The Utah Compose team is always happy to help you understand the scoring of an essay. We welcome questions as well as feedback from our users. This helps us ensure that our scoring is rigorous and reliable. Please contact us at any time. For more information about how essays are scored, see the [Utah Compose rubrics](#).