Why You Shouldn't Use Utah Compose Scores as Grades

Utah Compose does an excellent job of scoring essays and giving students feedback on their writing. The scores are based on thousands of essays scored by highly trained hand-scorers, and our scoring engine, PEG, has actually won awards for its ability to match the scores of humans. So why is it a bad idea to use its scores for student grades?

First, PEG is not human. It's a computer program. It is designed to match essays to human-scored essays to come up with scores for six writing characteristics. However, it cannot adjust to nuances or those essays that do not match the norm. When faced with something different from the training set of essays, PEG is stumped. Suppose a student copies and pastes whole pages from a website, for example. His teacher could immediately understand that the student has plagiarized, but PEG would not, because PEG does not 1) know the student, 2) relate the writing to the student's past writing, or 3) differentiate between what is appropriate for a seventh grader and what is clearly professional writing. PEG would compare the essay to the training set of essays for seventh graders, find that it best matches a five, and score the essay accordingly.

The same thing happens when a gifted writer takes certain writing risks. While the essay may be amazing, PEG may score the essay lower than it deserves because the essay's characteristics do not fit the training set for a four or a five.

Because of PEG's inability to "track," we often joke that PEG is like Drew Barrymore's character, Lucy, in the movie *50 First Dates*. In the movie, Lucy has a type of amnesia that causes her to remember nothing and no one from her past. Thus, every day, she makes the same discoveries she made the day before. When PEG "sees" an essay, it is as if it is the first time ever that PEG has encountered an essay, whether it is actually a first or a sixth draft.

The writing sites team often fields questions and concerns about spelling and/or grammar errors that PEG misses. We are constantly updating PEG's error identification capability, but because the English language is a complex one, PEG will never be able to identify one hundred percent of the errors found in essays. This is because there are contradictory rules and because PEG is not capable of discerning meaning. A homophone example will make this clearer. A proficient reader is able to realize that the "I" in "wind" is a short one in this sentence: "The wind was bitter today," but that it is a long one in this sentence: "I need to wind the yarn into a ball." PEG does not have the ability to understand context, so cannot make these types of distinctions.

Another problem is that PEG does not recognize content. An essay written for social studies is the same to PEG as one written for science, language arts, art, or calculus. Thus, it cannot look for content-specific vocabulary, nor can it make sense of content concepts.

Thus, while PEG is amazing, it is important to remember that it is a machine, not a brain, and can only do what it is trained to do. It would be very unfair to assign a grade to a student based on its evaluation of an essay. One student might benefit from transferring PEG's scores to grades, while another might be penalized.

Utah Compose was designed to help students practice their skills in writing and to improve them based on feedback. This is a great tool, but it is not the only tool for teaching writing nor for assessing it. Just as a chef uses many different tools in preparing a meal, teachers need to use different tools to teach and assess writing. Substituting a mixing bowl for a stove would not make a very good cake. Substituting Utah Compose for a teacher-assessed essay does not make a dependable summative grade.