What Do Students Think About Using SAGrader™ For Essay Grading?

Edward Brent, Theodore Carnahan, and Jeff McCully

To assess how well students like the use of SAGrader for automatic essay grading, we asked them a series of questions after they completed their essay and received the program's feedback and their scores. Virtually all students like the immediate feedback, the detailed comments, and the opportunity to revise and improve their paper. Most think it is fair even in its first assessment. They prefer automatic essay grading over traditional multiple-choice tests by almost two to one.

WHAT DO STUDENTS LIKE ABOUT SAGRADER?

Students like the IMMEDIATE feedback
Students loved the immediate feedback (92% liked it, with 60% liking it a lot). Most SAGrader essay grading exercises produce detailed comments for the student in less than 2 seconds. A very long answer to an extremely complicated assignment might take as much as 15 or 20 seconds, but we expect we can reduce most assignments to less than 10 seconds in response time. In contrast, grading of essays by the instructor or students requires at least a few days and sometimes two or more weeks before students learn their grade.

Students like the DETAILED feedback
Students overwhelmingly liked the detailed comments provided by SAGrader (88% liked the detailed comments, with 43% liking them a lot). SAGrader grades essays by assessing whether students can identify important concepts, theories, researchers, studies, and findings in the discipline, and can reason about them as required by the discipline. Hence, it is possible to provide detailed feedback to students telling them what they got right and the kinds of things (but not the specific items) that are missing from their essay. Below is a sample portion of the feedback the program provides in its automatic mode based on an early version of the deviance assignment.

> **Evaluation Result: 6 out of 10 points.**
> You were asked to write about responses to cultural diversity. You only identify the responses to cultural diversity 'ethnocentrism.' You failed to identify 2 other responses to cultural diversity. You were asked to describe ethnocentrism. For ethnocentrism, you successfully describe the feature 'using one's own culture as the standard for judging others.' You could also describe 1 other feature of ethnocentrism. Your grade for this submission is 6.7 out of 10 points (67%).

Students like the opportunity to revise their paper
Students overwhelmingly like the opportunity to revise and resubmit their paper (92% liked it, with 66% liking it a lot). Because SAGrader automates the essay grading process it is possible to let students revise their paper and resubmit it multiple times to improve their score. Instructors can permit unlimited revisions, a specific maximum number of revisions, or can permit no revisions. When

multiple revisions are permitted this facilitates student learning. In contrast, when essays are graded by hand, multiple revisions may require far too much effort to grade, and are often prohibited.

Students prefer automatic essay grading over multiple-choice tests
When asked, students indicated they prefer evaluation based on automatic grading of essays over multiple-choice tests by almost 2:1 with (47%) preferring automatically graded essays, 26% preferring multiple-choice tests, and 26% undecided. This is important because in large introductory courses it is often not possible to have students write many essays and hand-grade them. The only realistic alternative for large courses is often multiple-choice tests.

Most students thought the grading was fair
Most students thought the initial grading was fair (65% thought the initial grading was fair, 35% disliked the initial grading). This result is from the first times we used the program. We are still in the process of evaluating how fair they believe the program is after further improvement and our efforts to address their concerns. Of course we are very concerned about fairness, so we build in the opportunity for students to criticize the program or to raise questions about particular points they believe the program missed in student challenges. In most cases, their concerns are unjustified and the program is doing what we want it to do. However, in some cases, students have correctly identified mistakes that we have addressed or are in the process of addressing. Those are discussed in some detail in the section below.

Other advantages
Students also like other aspects of the program. They like that it is unbiased and consistent, treating everyone the same regardless of who they are. The program does not give better grades to more popular or better looking students, but instead grades everyone exactly the same. They also like that it is reliable. They can depend on getting the same result regardless of when their paper is graded, unlike human scorers who may inadvertently grade harder after just reading a very good essay. Perhaps most importantly for students who seem to live on a different clock than teachers, they like that the program is available 24 hours a day and can immediately grade their paper just as easily at 3AM as it can during the day.

WHAT DO STUDENTS DISLIKE ABOUT SAGRADER?

Undetected Items
The most common student criticism of the program is that it failed to detect an item the student believes is expressed in their answer. SAGrader records each version of an essay students submit along with any challenges. Those challenges can be reviewed by the instructor who can revise the program or override its score. This quality control is built directly into the program.

Most missed items fall into the following categories. The first are really student errors, not program errors. The last group are program errors that can be corrected.

Student errors
- **Incorrect terminology**. This occurs when a student uses the wrong term (e.g., one student said "differentiation of labor" instead of the widely recognized sociological term, "division of labor"). This is a mistake by the student, and the TAs or instructor would mark it wrong as well. Most disciplines expect students to learn the names of key terms.
- **Misquotes**. For example, in one exercise the program requires students to include quotes from an article to show how that article reflects important sociological concepts. Sometimes students identify a passage that reflects a concept we hadn't recognized in our model

answer and we add that quote. Far more often though, it turns out they misquoted the source and they are incorrect.

- **Spelling errors**. Sometimes when students complain that the program missed a key concept, they have misspelled it. We encourage students to use a spell-checker on their work before they submit it. Currently, SAGrader does not check spelling. This feature is planned for future versions.

Program errors

- **Knowledge gaps in the program**. This occurs when the SAGrader knowledge base does not include a key concept. This can happen when instructors switch texts and students are expecting a topic to be covered that is not included in the program. To fix this the program should be carefully coordinated with the course text.
- **Failure to recognize an acceptable expression of an item**. For example, "understanding things from the other person's perspective" might not be recognized as equivalent to "taking the role of the other." This is corrected by adding other key terms and their weights to enable SAGrader to recognize phrases using fuzzy logic. Carefully chosen key phrases and weights can enable SAGrader to recognize literally thousands of possible expressions of a concept.

Resistance to Deception

Some students were concerned that it would be possible to get a good grade on the paper by simply listing key words. However, SAGrader requires more than just key words, so this is not true. It also looks at whether students can make an argument consistent with the knowledge required to answer the question. Typical assignments require them to perform tasks such as relate some concepts to others, define key terms, summarize a theory, and show how quoted material relates to substantive concepts for that discipline. For all of these reasons, just entering key terms in a list would not earn students a good grade.