

Students Improve Learning by 20 Percentage Points with Essay Grading Program, SAGrader™

Edward Brent, Theodore Carnahan, and Jeff McCully
Copyright © Idea Works, Inc., 2006

An analysis of 1,595 student essays submitted in a college-level introductory course found that essays revised based on feedback from the SAGrader program achieved final grades 20 percentage points higher than the grades assigned to first drafts of those essays. The average score for the first drafts of revised essays was 70% (a grade of C-), while the average score for the final drafts of those same essays was 90% (a grade of A-). The final scores for revised submissions were higher than the scores for students who quit after their first draft, even though the initial scores for revisers were much lower.

This study examined the performance of 78 students enrolled in an introductory sociology course at the University of Missouri. This was a writing-intensive course with 43 different writing assignments over the semester. Assignments varied from several short-answer questions to moderate sized (1 or 2-page assignments) to a full-blown 15-page term paper. Somewhat more than half of the assignments were required of all students, while the remaining assignments were optional and could be completed for extra credit. The SAGrader computer program graded all of these assignments. Using this program, students submit their essays online using a standard web browser and, in a matter of seconds or minutes, receive detailed feedback and comments telling them their score, the learning objectives they got right in their essay, and the types of items (but not the specific items) related to learning objectives that should have been included but were not. Students were permitted to revise their essays as many times as they wished based on this feedback from the SAGrader program. For many assignments students used this opportunity to revise their essays and improve their grades. Some students chose to stop with their initial grade on specific assignments and did not take advantage of SAGrader's feedback to improve their score. There were a total of 1,595 student essays and 3,791 separate submissions, for a mean of 2.4 submissions (one first draft and 1.4 revisions) per essay.

How Much Do Revisions Improve Scores?

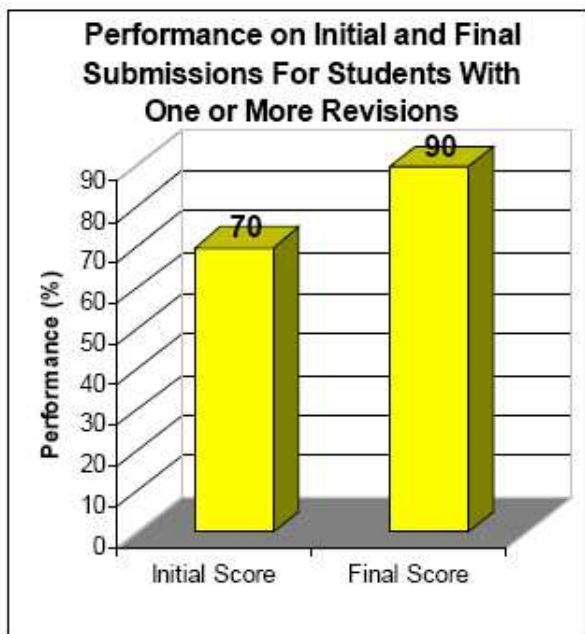
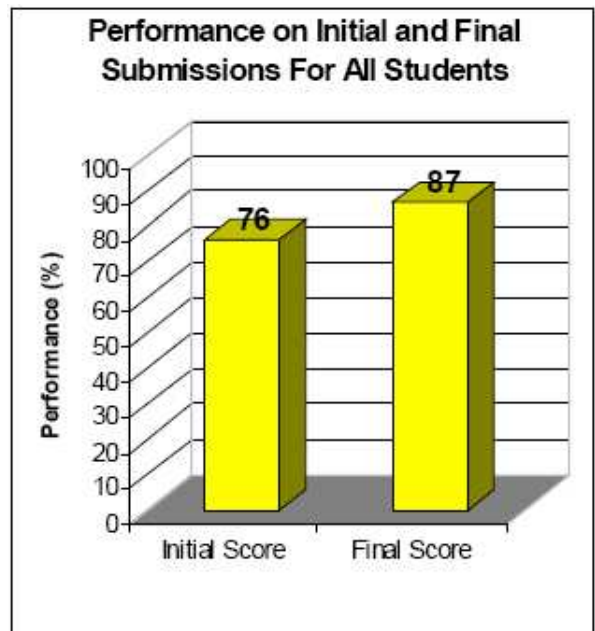
The table below compares initial score and final score means for all essays, essays having first drafts not followed by revisions, and essays including one or more revisions.

	Number	Initial Score	Final Score
All Essays	1595	76%	87%
Essays with first drafts only	670	84%	84%
Essays including one or more revisions	925	70%	90%

For all essays across all assignments the average performance for the first draft is 76% and the average for the last submission is 87%, for a change of 11% or approximately one letter grade. This is an impressive improvement.

However, in 42% of the cases (N=670) first drafts were not followed by revisions. The initial and final performance for those essays is the same - 84%. So those 42% of essays that show no change make the improvement due to revisions appear lower than it really is.

To better understand the impact of revisions we look only at essays having one or more revisions. For essays having one or more revisions, the average performance on the first draft is 70% and for the last submission, 90%, resulting in a change of 20%, or a two-letter grade jump from a C- to an A-. That is, by using feedback from SAGrader and revising their work, students improved their grade by 20 percentage points or two letter grades—even before the instructor examined their essay.



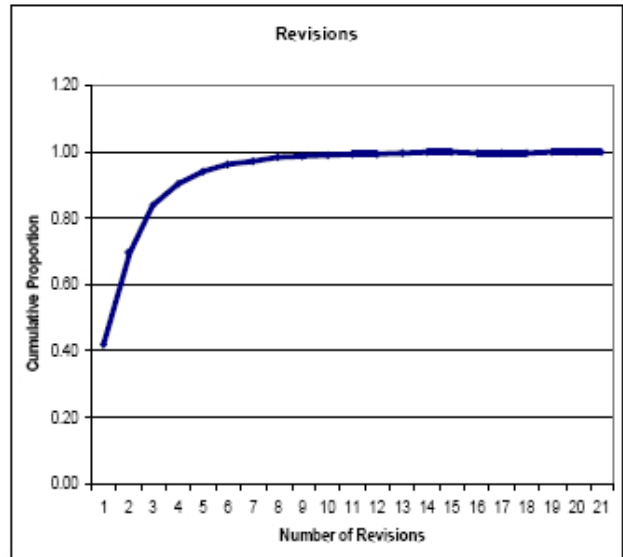
In fact, *students who revise, on average, do better than students who do not revise*, even though the non-revisers began with an average of 84% compared to a beginning average of 70% for revisers.

This improvement in scores from 70% to 90% for essays that are revised based on SAGrader's feedback is impressive. But could it be overestimated? It might be argued that the improvement by students from first to last drafts is overestimated to the extent that students submit very poor first essays and then use feedback from SAGrader to modify their essay. That is to say, perhaps they are letting SAGrader do more of the work for them. It is possible that students do sometimes exercise less care with their first drafts and rely more heavily on SAGrader's feedback to improve their final drafts. However, if this was widespread, one would expect most early drafts to be poor ones. Instead, the average score

on first drafts (including students who do not do subsequent drafts) is 76%, a good solid C grade. The case study to be described below in which the first draft score was 50%, shows that getting even a 76% on an essay is not a trivial matter and requires some effort and knowledge. The fact that for 42% of essays students stop after the first draft is also consistent with the view that students make a reasonably good effort on their first drafts and do not rely exclusively on advice from SAGrader to improve their scores. Finally, it is also possible that the improvements made by using SAGrader are underestimated to the extent that original essays receive good scores. For essays that are less difficult where initial average scores are already high, there is little room for improvement, no matter how helpful the SAGrader program's advice might be. So any biases that might lead to overestimating learning with SAGrader may be countered somewhat by biases that underestimate learning.

How Many Revisions Do Students Make?

The cumulative proportion of essays by the number of submissions is displayed in this chart. In ninety percent of the essays, students submitted three or fewer revisions for an assignment. Eighty-four percent make two or fewer revisions. Seventy percent make one or no revisions. In 42% of cases, students make no revisions. The mean is 2.4 submissions, or 1.4 revisions, per essay.



How Do Students Decide When to Stop Revising?

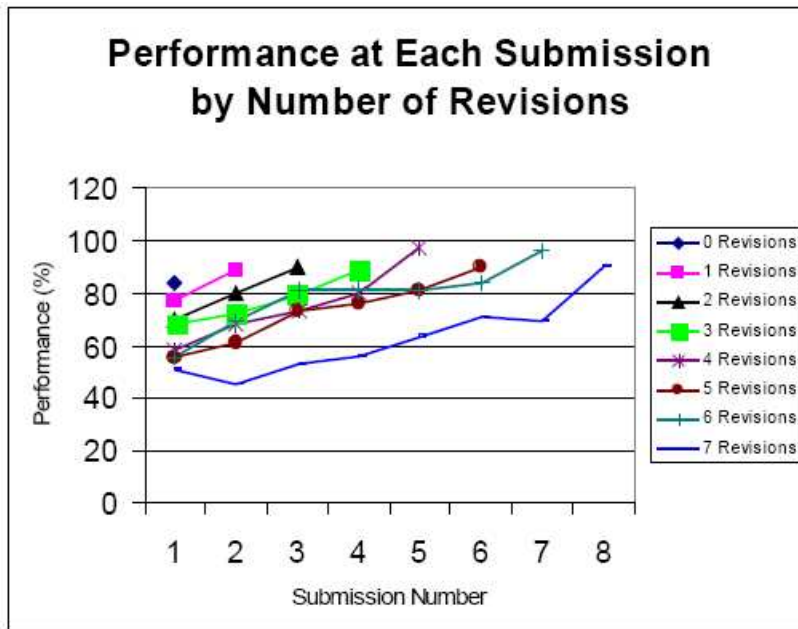
In the table below, the performance scores for students for each essay are broken down by the number of revisions made by that student for the assignment. For example, there were 670 essays where students submitted no revisions and had an average score of 84%. There were 443 essays where students submitted one revision. Their initial score was 77% and their final score was 89%.

Mean Scores by Number of Revisions and Submission Number

Number of Essays	Number of Revisions	Submission Number							
		1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00
670	0	84*
443	1	77	89*
223	2	70	80	90*
104	3	68	72	79	89*
61	4	58	68	73	80	97*	.	.	.
34	5	55	61	73	76	81	90*	.	.
14	6	55	69	81	81	81	84	96*	.
19	7	51	45	53	56	63	71	69	90*

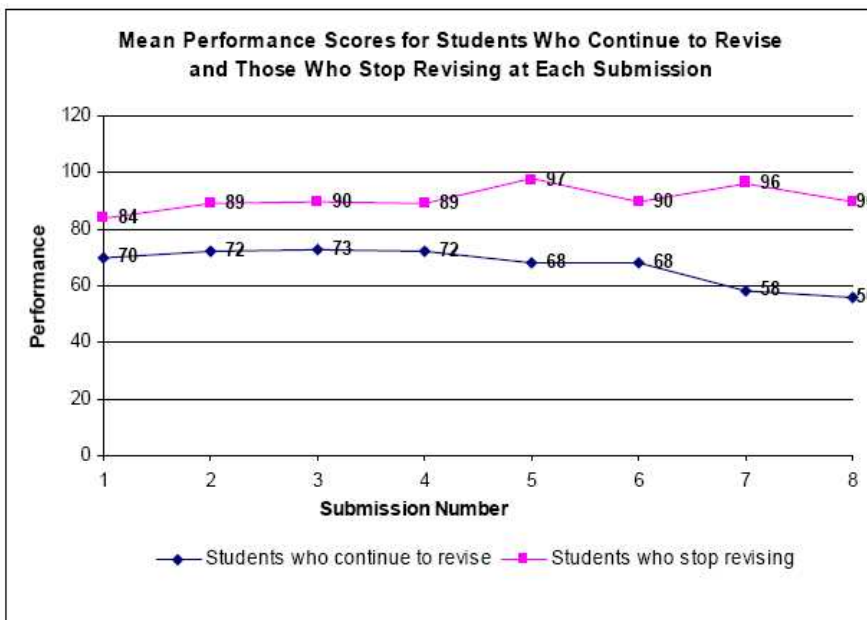
*last submissions

These same data are displayed in the following graph.



Notice that the average of scores for the last submission, whether it is the second, the eighth, or any in between, is essentially 90%. Thus it appears that for most essays, *students seem to quit revising when their average approaches 90%, regardless of how many revisions it takes to achieve that average.* Some students achieve a personally acceptable score with their first draft; others take one, two, or more revisions to reach that score. It is interesting that this acceptable score appears to be roughly the same regardless of how long it takes them to achieve it. If final scores for essays with more revisions were substantially lower than the scores for essays with few revisions, this would suggest some of the students were giving up on some of the essays rather than being satisfied with their score. In fact, this does not appear to happen for most essays.

While the final scores on essays are very much the same regardless of numbers of revisions, for each submission number there is a large gap between scores for essays where students stop revising and those essays where students continue revising. This gap between performance scores for essays students continue to revise and those where students stop revising at each submission can be seen clearly in the figure below.



Generally, there is roughly a 20% difference. Essays where students continue revising have mean scores around 70% or lower, while essays where students decide to quit have mean scores approaching 90% or higher. These results suggest students are both rational and tenacious, continuing to work on their papers when they are not satisfied with their score, quitting only when they achieve what they regard as a “good” or at least “acceptable” score. The decision about when to stop working on their papers appears to be determined primarily by their current grade and has little or no relationship to the number of revisions they have submitted.

A Case Study

We can illustrate this general trend of 20% improvement from first draft to final draft when essays are revised by examining a specific case. This student submitted his essay three times. The first submission was at 10:08 PM and the student received a score of 50%. The second submission occurred 9 minutes later and the student received a score of 52%. This did not show much progress and the student apparently then spent more time looking back at the lectures and book until making the third and final submission for this assignment at 10:37 PM and receiving a greatly improved score of 90%. In the course of three submissions and less than an hour this student went from a score of 50% to a score of 95% -- all in the middle of the night and all occurring before the instructor saw any of the essays.

This case illustrates the general trends found above. In this case the student submitted three times (the average is 2.4 submissions). The student's initial grade was less than 70% (50% in this case). The student stopped revising the paper when they achieved a grade of 90% or higher (95% in this case).

But does this dramatic increase in a student's grade reflect real learning and true improvements in performance, or was the student able to "game" the program, increasing his score through some trick or by taking advantage of some flaw in the program, without any genuine learning? That question can be answered very specifically for this case. Examining the first and last drafts for this case we can see clear improvements in the essay. The student's initial draft had 322 words in 6 paragraphs. The third draft was considerably longer, with 424 words and 7 paragraphs. Three paragraphs in particular were longer and more developed. The coverage of important topics was greater. The language is more precise. This can be seen by the side-by-side comparison of excerpts from the first and third submission in the following figure.

Figure 1: Comparison of First and Third Drafts of a Student's Essay

Submission 1	Submission 3
<p>Material culture is everything that belongs to culture that is tangible. Nonmaterial culture would be the values, beliefs, and behavior accepted in culture. There are two types of norms in culture. Folkways govern everyday behavior but are not strictly enforced. Mores are more serious, carrying greater moral gravity and are strictly enforced.</p>	<p>Material culture is art and material objects that belong to a culture. Nonmaterial culture would be the symbols, language, knowledge, beliefs, values, attitudes, and norms accepted in culture. Norms are the expected behavior in a society. There are two types of norms in culture. Folkways govern everyday behavior, are not morally important, and are not strictly enforced. Mores are more serious, carrying greater moral importance, and are strictly enforced. Values are standards of importance and rightness in society. Language is a abstract system that allows people of a society to communicate. Symbols are arbitrary signs that stand for something.</p>
<p>Social change happens by cultural diffusion where ideas of one culture pass to another culture. Cultural Integration is the consistency of elements found in a single culture. Culture changes following the cultural lag theory by which technology changes culture and other elements</p>	<p>Social change happens by cultural diffusion where cultural elements spread from one culture to another culture. Cultural Integration is the consistency of elements found in a single culture. Culture changes following the cultural lag theory by which technology</p>

<p>lag behind. Cultural leveling evens out the good and bad of two cultures to make them equal taking away the uniqueness of each.</p>	<p>changes culture and other elements lag behind. Cultural leveling reduces the differences of two cultures to make them equal taking away the uniqueness of each. Cultural universals are elements that appear in all cultures. Cultural relativism judges a culture by the standards of the culture. One response to cultural diversity is multiculturalism. Multiculturalism recognizes the contributions of all groups and does not hold one group better than another. Some do not accept cultural diversity and oppose cultural change. Other accept the cultural diversity.</p>
--	--

The third draft expands the last two paragraphs of the first draft and adds a new paragraph. Among other changes, it adds several more examples of nonmaterial culture (norms, values, language, and symbols) and their definitions. The new paragraph covers a topic left out of the first draft altogether—responses to cultural diversity. This third draft, while not perfect, is clearly much better than the first one.

SAGrader is designed to insure that the correspondence between scores and knowledge found in this example is likely to be found in general. To begin with, constructed responses or essays are much harder to “fool” than fixed-choice tests, and the process of constructing essay responses itself often fosters learning. The way in which SAGrader assesses essays makes it hard to “game” the system and get a high score. It is not sufficient for students to simply list key concepts the program is looking for to get a good grade. Good essays must not only include appropriate concepts and terms, but must also express relationships among them consistent with the knowledge underlying the learning objectives. In the student excerpts above, for example, mentioning concepts such as language and symbols alone is not enough to get full credit. Students must also indicate that those are both examples of nonmaterial culture. Definitions must be clearly linked to the correct concepts, authors must be associated with the correct theories, and so on. Another check on validity provided by SAGrader is the built-in ability for instructors to easily review any essay and override the program’s score if they felt it was incorrect.

The feedback SAGrader provides students also encourages learning. SAGrader identifies in detail items they got correct in their essay, but does not tell them the explicit items they missed. Instead, it indicates the number of additional items of various types that should be present in a good answer. Students who improve their papers must go back and reread the text, more carefully think about the issues, or more precisely draft their essay to improve their score. SAGrader gives them direction for improving their work, creating a learning environment in which they can continue to learn and improve their score, but does not hand them the answers on a silver platter.

Discussion

When instructors grade essays, each submission incurs a substantial commitment of time, money, or both to grade. This encourages instructors to limit writing assignments and to restrict the number of times students can revise their work or prohibit revisions altogether. As a result, writing assignments are few and far between and when they do occur they are used primarily to evaluate students rather than as a learning experience. Writing-across-the-curriculum programs at many universities try to

overcome this by making more resources available for selected courses to make it possible to have more writing assignments and to permit multiple revisions and learning through writing.

In contrast, SAGrader automates the grading of essays, creating an environment in which assessments of student revisions are nearly free. Once the initial essay assignment is constructed, the cost of grading multiple drafts by each student is hardly more than the cost of grading a single draft (see “SAGrader Benefits Exceed Costs After One Semester”). Thus, SAGrader provides both an assessment tool and a learning environment where students can submit essays, receive detailed personalized feedback immediately while the issues are still fresh in their minds, and revise and resubmit their work.

This analysis of students’ use of SAGrader finds it to be a very effective learning environment in which revised essays improve by an average of 20 percentage points or two letter grades. Students display considerable tenacity and rationality, taking advantage of this learning opportunity by continuing to revise and resubmit their essays until they achieve an average grade of roughly 90% or higher, even when their initial grades may have been quite low. This dramatic improvement does not appear to be an artifact of artificially low first draft scores or artificially high last draft scores. The case study illustrates how this large change in scores chronicles a true improvement in understanding and communication that is reflected in increased length, quality of writing, coverage of topics, and precision.



www.sagrader.com
www.ideaworks.com