

Measuring the Effect of Written Feedback on Writing

Dedra Demaree

Department of Physics, College of the Holy Cross, Worcester, Massachusetts, 01610

Abstract.

Members of the Physics and English departments at The Ohio State University (OSU) and Rochester Institute of Technology are involved in an ongoing study addressing issues related to writing activities in the physics classroom. Students in the physics 103 and 104 course sequence at OSU “The World of Energy” view weekly videos then turn in summaries as part of their homework grade. These summaries are given one point if turned in; they are not graded for the quality of their content. In winter quarter, 2006, some students were given substantial feedback on these summaries with comments aimed to improve their writing. Feedback-induced improvement in their video summaries is demonstrated in this paper.

Keywords: writing to learn

PACS: 01.40.Fk

INTRODUCTION

Writing within content areas is often promoted as a way to both improve specific writing skills and as a way for students to learn content [1, 2]. Members of the Physics and English Departments at both The Ohio State University (OSU) and Rochester Institute of Technology (RIT) have been studying writing to learn within physics to find ways to quantify claims in the literature and measure any effects of writing on learning [3, 4].

Writing is sometimes used in content areas as a formative assessment tool since it provides the instructor opportunity to give specific feedback to student ideas [5, 6]. In order for formative assessment to be effective, the learner must realize there is a gap between their present state and a desire goal, and take steps to close that gap [7]. Students must be able to either self-assess their gap, or instructors must be able to communicate to the students about how they are doing. In [7], Black and Wiliam warn that “there are complex links between the way in which the message is received... and the learning activity which may or may not follow.”

Traditionally, many writing to learn studies have been qualitative, and such studies seldom try to *quantify* the claimed benefits of writing [8]. In addition, there is sometimes little information given on the specific nature of the feedback, or more importantly, how the quality of the student writing was measured. It seems that at a minimum, if student learning is taking place based on feedback there should be some quantitative, measurable effect of that feedback. This paper uses Physics Education Research techniques to conduct a small study to look for a quantitative effect of feedback on student writing without external motivators, namely without giving students additional points toward their grade.

IMPLEMENTATION DETAILS

In the winter quarter of 2006, a study was conducted in order to see if student writing could be improved based solely on feedback given in the form of written comments on students’ papers. Students in Physics 104 (“The World of Energy”) at The Ohio State University (OSU) watch eight weekly videos on various topics related to energy issues, from pollution in Los Angeles to clean-up efforts at Chernobyl. After each video, students write at least a two paragraph summary of the contents. Students in this course are graded in a typical fashion based on homeworks, two midterm exams, and a final exam. The video summaries count for 1 point each out of a total of 140 course points. Students receive this point if they submit the video summary. The content of the video summary is essentially ungraded.

The Physics 104 course has been taught in some form for over 30 years at OSU. From the start, this was a studio-based course, with ample hands-on explorations. The videos are aimed at helping students see the connection between the technology discussed in the course and daily life. The video summaries primarily serve as assurance that the students watched the videos.

This course provides an ideal situation for this study because there is a situation where one set of students can be given additional feedback without disadvantaging the students who do not receive feedback. Since all students receive one point for submitting the summary regardless of the quality, giving feedback will not have any affect on student grades. There is also no concern if students who receive feedback learn more than the other students, because neither the content of the video summaries or their writing abilities are tested in the course.

There are three instructors for this course: A, B and C. Two of them teach one section each, while the third

teaches three sections. One section is taught on Saturday and contains a large percentage of graduate students (in-service teachers and pre-education majors) who have additional assignments. Instructor A typically does not give any feedback to the students when he grades; call his class N1. Instructor B does not give extensive feedback though she feels it is important to give positive encouragement and often wrote things such as “very nice” or “I agree” on students papers; call her class N2.

Instructor C (with three sections) typically only writes “okay,” “good,” or “very good” to indicate his level of satisfaction with the student work. In two of his sections I collected the summaries and gave extensive feedback on their writing; call them C1 and C2 (C2 is the above mentioned Saturday section). When he received the commented papers from me he also read them and wrote one of his usual three comments before marking down their point. In his third section, only his normal comments were given; call that section N3.

Sections N1, N2 and N3 are therefore the control sections, where no additional feedback is given, while C1 and C2 were given special treatment. These details are summarized in Table 1. I was not involved with the course in any way except to provide the additional summary feedback.

TABLE 1. This table shows the five sections involved in this research study.

Instructor	A	B	C
Type of Instructor Feedback	None	Minimal but positive	Minimal with some indication of satisfaction
No Additional Feedback (Control)	N1	N2	N3
Additional Feedback			C1 and C2

The students were not given any specific instructions from the course instructors about how to write good summaries. The instructions to the students were to “write at least two paragraphs summarizing the video and what you see as the major issues presented by the video.” The students were also given a list of questions specific to each video as a guide to help them take notes.

At the start of this study, a collaborator from the English Department gave me extensive information on what would normally be expected from a video summary, as well as what kind of feedback would be appropriate to give. These ideas were used to design both the feedback I gave and the grading rubrics for the purpose of this study. In general, a summary should have a clear order and there should be some indication as to why the information is important. The organization of the summary should reflect what is being summarized, for instance if the video

had two distinct parts the summary should too. The ideas in the summary should be related to each other and related to the overall point of the video and not just given in a list. The summary should reflect the balance of importance of points: more time should be spent on main ideas, and less should be spent on supporting stories.

DATA OBTAINED

During the course of the quarter, I collected summaries from sections C1 and C2 from the instructor the day they were handed in, and I promptly returned them to the instructor with my written feedback directly on the student papers. For all sections, after the instructors had a chance to look at the summaries and write their own comments, I collected them and made copies for my records. Students were given one point for submitting the summaries regardless of the quality of the content.

After the quarter was completed and the students’ grades were recorded, I used the copies of the summaries to grade them in detail based on the rubrics. From here on out, reference to the summary grade refers to this post-quarter grading for research purposes, and does not reflect the single point the students received during the quarter toward their course grade. I also recorded the number of instructor comments, if they were positive or negative, and the nature of comments I gave to each summary.

TABLE 2. This table shows the details of the five grading rubrics.

Rubric	Explanation
Balance	Does the emphasis placed on ideas in the summary reflect their importance?
Flow	How readable is the summary, and are the ideas related or just given as a list?
Content	Is the content correct, is terminology defined and are all important points included?

In line with the above information about what makes a good video summary, I graded the summaries based on three main parameters: Balance, Flow, and Content. The explanation for each of these is given in Table 2. For each of these parameters students received a grade ranging from 0 to 3 points, similar to the grading rubrics developed for the *ISLE* classroom by the Rutgers Physics and Astronomy Education Group [9]. 0 indicates that there is no attempt made by the student toward this parameter. 1 indicates that there is an attempt but it is barely perceivable. 2 indicates that the student made a decent attempt toward that parameter. 3 indicates that the student did a good job with that parameter. When I gave the written feedback to the students I mainly gave comments asking how ideas were related, what was

the importance of certain statements, or I asked for an explanation if something the student stated was not clear.

In addition, I also gave a voluntary feedback survey to the students in the C1 and C2 sections, asking them about the additional written feedback. Students were asked if they read the additional feedback, what they thought of it, if they felt it impacted how they wrote the video summaries, and were asked to comment generally on the usefulness of the videos and the video summaries. Since the survey was voluntary, I did not get many returned, but the range of student opinions is presented here.

RESULTS

Although there were roughly 120 students in the five sections of Physics 104, only 41 students were used in this study. 21 students were in the C groups and 20 were in the N groups. Almost none of the students in section N1 signed consent forms, and some students from other sections declined to participate as well. In addition, the graduate students were removed from the study since they did not have a comparable control group. Since the video summaries were only worth one point each, students from both the C and N groups who completed less than six of the eight summaries were also excluded from the study. When ranked by final course grade, students from the C groups are very equally mixed among students from the N groups, indicating that the two groups were similar in overall course content knowledge. Additional analysis described below shows that any effects of having a non-ideal selection of students for this study did not effect the results.

The remaining 41 students were given a grade for each video summary based on the three parameters as discussed above. These students were then separated into groups based on if they received the additional feedback, on their final course grade, and on their video summary grades. Summary grades were also analyzed based on three methods. First, I looked at the individual summary grades for each week, second, I looked at the difference between each student's initial and final summary grade, and third, I did a linear fit for the eight summaries of each student to see if students had consistent improvement in their video summaries throughout the quarter.

The strongest indicator would be seeing a consistent significantly higher summary grade for the later weeks for the C1 and C2 groups (with respect to the control groups, N1, N2 and N3). However, this would be difficult to achieve with such small numbers of students. Indeed, the students in the C groups have lower (5.7/9 compared to 6.5/9, not significantly with $p = 0.1$) summary grades the first week, and slightly higher summary grades (7.2/9 compared to 6.9/9) in the last week. It could be that the students in the N groups were initially

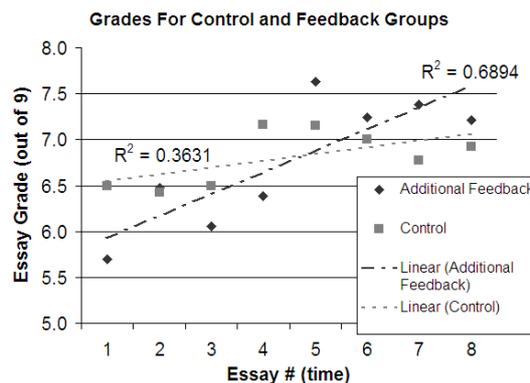


FIGURE 1. This figure shows the average grades for the Control groups and the Feedback groups, as well as a linear fit showing the higher slope for the students who received additional feedback.

better at writing the video summaries, and that this is masking any ability to see significantly higher grades in the C group at the end of the quarter.

In order to remove any initial writing level, I also looked at the students' final summary grade minus their initial summary grade. Students in the C groups had an average improvement of 1.1 point out of 9, while students in the N groups had an average decrease of 2 points out of 9. This significant ($p = 0.003$) result is sufficient to show that students in the C groups improved their video summaries throughout the quarter more than those in the N groups.

Another interesting way to look at the data is to look for signs of improvement throughout the quarter (not just comparing the final and initial summaries.) A positive slope as found from a linear fit to any individual student's summary grades indicates a net improvement. Linear fits may not be strong as grades can fluctuate based factors such as the video content and how much time the student devotes to a particular assignment. The average slope for students in the C groups is 0.24, indicating that over the course of eight summaries their grade would improve by nearly 2 points out of 9. However, for the N groups, the average slope is only 0.06, indicating that their grade would improve by only half a point out of 9. This result is statistically significant with $p = 0.005$. The average Correlation Coefficient for the linear fits is only $R^2 = 0.310$ for the C groups, and $R^2 = 0.180$ for the N groups. This reflects the difficulty of making a good fit to this type of data. However, when a linear fit is done for the average scores for all students in each group, a clear difference in the slopes can be seen (see Fig. 1), indicating that the improvement in the C group summaries was consistent throughout the quarter.

It is possible that only a certain subset of students

(for example, those with better grades, or those working with a particular instructor) would show improvement on their summaries regardless of feedback. I therefore also ranked students based on their final course grade. There was no statistical difference in the calculated slopes for students who did better in the course than others ($\rho = 0.38$). Students who had higher course grades did tend to have better video summary grades, but those grades did not change any more through the quarter than the lower scoring students. Similarly, students who started with high video summary grades were neither more nor less likely to improve than those who started with low grades ($\rho = 0.91$). These results indicate that the observed improvements in essay grades are dominantly based on the additional feedback given to students in the C group.

Students in the C groups improved more than the N groups in each individual parameter. However, the results for the individual parameters are not statistically significant. Students from the C groups improved most in the Flow and Content parameters (with an average of 0.5 out of 3, and 0.7 out of 3 point improvement through the quarter respectively). Effectively no improvement was observed in the N groups for these (with an average of -0.1 and 0.2 point improvement through the quarter respectively).

Out of the eight summaries, students in the C groups received on average 4.8 positive comments, or approximately one positive comment for every two summaries turned in. Those in the N groups received on average 6.5, so on nearly two thirds of their summaries. Recall that instructor B made a point to give positive feedback to the students. This shows that the simple presence of positive feedback does not explain the improvement in student writing; it is the specific feedback on how to write that made a difference. Students in the C group received on average 30 comments over the course of their eight summaries, while those in the N group received on average only 9.3. Students in both groups rarely received negative feedback, with only 0.5 comments for the C groups and 0.1 for the N groups (1 in 10 students).

Of the thirteen C group students who returned the voluntary surveys, nine of them said they always read the additional feedback, two sometimes did, and three never did. About three fifths of the students said the feedback was either clear, helpful, or good. Despite the lack of explicit negative feedback, about two fifths of them said it was annoying and/or unclear. Their reason was mainly that they were being criticized when they knew they would get full points anyway. Slightly less than half of the students thought the feedback impacted the way they wrote their summaries. Integrating the feedback with instruction about how to write the summaries would probably make the experience more positive for the students. The students overall enjoyed the videos but thought the summaries were not very useful for the course.

CONCLUSIONS

Students in this study who received additional feedback on their video summaries showed more improvement in the quality of their writing over the course of the quarter than those who did not receive additional feedback. They had significantly higher differences between their final summary grade and their initial summary grade, as well as significantly higher slopes as found by taking a linear fit of their summary grades through the quarter. These improvements were observed despite the fact that there was no motivator for students to improve their writing; they received full points toward their course grade regardless of the quality of their summaries.

Writing activities can provide the opportunity for formative assessment in the physics classroom, provided students reflect on the feedback provided to them. This study provides evidence that written feedback does impact student writing. This indicates that students did reflect on the feedback. This study also demonstrates one possible method for quantifying the quality of student writing, in lines with our desire to quantify writing to learn in physics.

ACKNOWLEDGMENTS

The author would like to thank Mary Wildermuth, James Bihari, and Bill Davis for their advice and for facilitating the collection of the summaries. The author would also like to thank Bernard Mulligan for his suggestion that I use this course in my studies, and for his useful discussion on writing in physics. The author would also like to thank Cat Gubernatis for her discussions about how to give feedback to the video summaries.

REFERENCES

1. C. W. Keys, *Science Education*, **83**, 115–130 (1999).
2. P. D. Klein, *Educational Psychology Review*, **11**, 203–270 (1999).
3. D. Demaree, C. Gubernatis, J. Hanzlik, S. Franklin, L. Hermsen, and G. Aubrecht., *PERC proceedings* (2006).
4. S. Franklin, and L. Hermsen, *In preparation for Educational Researcher* (2006).
5. T. Hein, *Eur. J. Phys.*, **20**, 137–141 (1999).
6. E. Etkina, and K. A. Harper, *Journal of College Science Teaching*, **31** (2002).
7. P. Black, and D. Wiliam, *Assessment in Education: Principles, Policy and Practice*, **5** (1998).
8. J. Ackerman, *Written Communication*, **10**, 334–370 (1993).
9. E. Etkina, A. V. Heuvelen, D. Brookes, and D. Mills, *The Physics Teacher* (2002).