Measuring the Effectiveness of Revised Writing in Mathematics Classes

Lewis L. Lasser, Ph.D. †

Abstract

There are many studies that compare the usefulness of writing in mathematics classes. In this paper a study is presented in which all the mathematics classes contained writing assignments. Half had low to mid-stakes writing and the other half high-stakes revised writing. The central question is this: does revisiting writing assignments through the mechanism of revision have any effect on student performance as measured on tests and final exams?

Introduction

York College, located in Jamaica, New York, is one of eleven senior colleges of the City University of New York (CUNY). For a number of years York has offered Writing Intensive (WI) courses in almost every discipline. These classes feature low to high-stakes writing assignments with an emphasis on high-stakes revised writing that contributes a significant amount towards the final grade. York also offers a large number of General Education Requirement (GER) courses. Each of these classes should offer abundant writing opportunities that are usually low to mid-stakes. The term Writing Enhanced (WE) is applied to all GER classes to reflect this writing component. It is possible for a WE class to be designated WI. GER classes in the mathematics department at York are numerous, so it is not uncommon for a single instructor to have one or even two during a particular semester. From time to time an instructor may have one section of a particular course designated as WI and another section of the same course designated as WE. A rather natural question can arise in such a context. WI students revisit each assignment by receiving feedback and then revising the paper, so they reflect longer on each question than WE students. Does this additional reflection produce greater understanding of the subject matter?

In the Spring of 2005 this question was put to the test by examining two sections of an elementary probability and statistics class (Math 111), one WE and the other with a writing component equivalent to a WI class, and two sections of a second semester calculus class (Math 122), again one WE and the other WI equivalent. Since the WE classes had no graded writing assignments the only way to compare student performance across the section within a particular course was to use tests and finals.

For the most part the two courses (Math 111 and 122) had many common elements in terms of how the study was conducted. This paper will focus on the probability and statistics class with the understanding that there are many similarities with the calculus class.

How the Study was Conducted

A few days before the start of the semester one of the two sections of the probability and statistics class was randomly selected to be the one with revised writing. From this point onwards the section not selected would have the standard low stakes writing typical of a WE class and the selected section would have high-stake equivalents featuring revised writing. On the first day of classes a Writing Fellow (Ian Gardiner) associated with York's Writing Across the Curriculum (WAC) Program introduced the study; distributed, discussed, and collected the consent forms; and informed each section whether or not their writing assignments would be revised.

As the semester progressed the two sections were kept in sync with respect to the material covered each day. For the most part this was a routine process commonly applied when the same instructor has multiple sections of a single course. The primary difference for these two sections was compensation for time spent discussing the writing process, the actual assignments, and general feedback applicable for the entire class between the first and final drafts. Spread over an entire semester the difference was easily eliminated without altering the course content.

There were four tests throughout the semester. On a test day the first of the two sections was administered a test and no students were allowed to leave until the end of class. After a ten minute period between classes the second section was administered the same test. The risk of students exchanging information about the test was minimized by four factors: the relatively short time between the meeting times of the sections, the collection of all test related material (especially the question sheets) for the earlier section, the location of the rooms (on different floors two "wings" apart), and the relatively small chance of contact between the two sections. The last point can't be quantified, but many students who take this course are new students at York and, other than chance encounters, would have to actively seek out students in the other section (whose identity would be known only by searching the schedule, since neither section was identified to the other at any time by the instructor).

All four sections meet in a single room at a common time for the final. This guaranteed a single test for all students in a particular course and eliminated any chance of one section passing information to a later section.

With these precautions identical tests and finals for both sections was not only fair but allowed comparisons between the two groups without resorting to similar but different problems.

Writing Assignments

Students frequently express a mixture of disbelief and curiosity (in that order) when informed of the writing component in their mathematics class. If the writing is low-stakes (typical of a GER class) it is usually mixed together with calculations and soon enough appears natural. For high-stake writing (typical of a WI class) this process is an ongoing one that may last the entire semester for some. To speed this along, and to introduce the write-revise-write cycle for those students whose must revise their earlier work, I always start off with an algebra problem. For my first writing assignment I chose the evaluation of the formula $\Sigma(x-1)^2/10$ for the first five values of x: 1, 2, 3, 4, and 5. In this formula the symbol Σ means to sum (i.e. add). Once this is introduced and a few examples are done at the board this assignment can be given to the students at any time. It is independent of the material yet to be covered and at the same time is similar enough to formulas they will see (e.g. the variance of a data set) that it not only serves as a warm up exercise but also as a teaching exercise.

Probability and Statistics is an introductory class that is taken by students whose major is not one of the sciences such as mathematics, computer science, and chemistry and not business related such as business or accounting. The material tends to be problem orientated with a small, perhaps nonexistent, theory component. This has a strong influence on lectures, the homework problems assigned, and the questions that appear on tests and finals. It is only natural then that it should guide the writing assignments as well.

This course, like many others of similar name and similar intent, contains material on probability distributions both discrete and continuous. The focus for discrete distributions was the binomial, geometric, and Poisson. Many, but not all, of the homework problems are labeled in such a way that the distribution is effectively a given. When first introduced in lecture it is also clear that a given problem is an example of a particular distribution. After their introduction and before the test it was made clear to the students that the identification of the distribution necessary to solve a problem was the first step they would have to take before actually invoking formulas or looking up values in one or more tables. The fourth and fifth writing assignments provided practice along these lines by presenting a series of word problems that required (in part) the choice of a distribution. The students had to explain why their choice (binomial, geometric, or Poisson) was appropriate and why the other two were not.

A continuous probability distribution always covered in a class in probability and statistics is the standard normal distribution. The students were expected to find probabilities for three situations: for values less than a given number, for values between two given numbers, and for values greater than a given number. They also had to be familiar with a related concept called the Empirical Rule (also known as the 68-95-99.7 Rule). In both situations they were exposed to word problems that they had to solve and to the concept that the area under the curve is identical to the probability the random variable takes on the indicated range of values. The second writing assignment exercised both the Empirical Rule and the relationship between the standard deviation and the variability of the data. The sixth and last assignment examined the relationships between probabilities, areas, and looking up values in a table for the standard normal distribution.

For each assignment the students were told that the most important aspect of their paper is the impact it would have on their fellow students. A typical wording goes like this: “if you give your paper to another student who is taking a class similar to this one they will be able to understand how the techniques work and will be able to solve the specific questions asked and those that are sufficiently similar." Feedback stressed the content and clarity of the work but did not ignore issues of spelling and grammar. The grade on each paper was based on four criteria: clarity and effectiveness; understanding the subject and factual content; following instructions and completeness; and proper revision. Penalties for handwritten work, spelling, and grammar were minor.

Data Collection

The choice of what data to collect is one of the most important parameters in any study. The WE students wrote but their writing was not graded, unlike the WI students whose writing was graded. To compare the WE and WI students it was necessary to avoid the actual writing scores. This left tests and finals. The overall score on the final was considered the primary measure of the effectiveness of revised writing for this study. This was an easy decision that has been used before (Heid, 1988; Judson, 1990; Palmiter, 1991). In addition to raw test scores an effort was made to isolate errors within individual problems. Again, this is not a new idea (Porter & Masingila, 2000). In light of the writing assignments given, it was natural to look for the choice of probability distribution in the overall context of solving a binomial, geometric, or Poisson problem.

Three questions on the third probability and statistics test required a choice of discrete probability distribution in order to solve the problem. One was a binomial, a second a geometric, and the last a Poisson. Past experience indicates the proper choice of distribution is only a first step towards the solution; there are many things that could go wrong after the choice and before the problem was complete. For this study the choice of distribution rather than a complete correct answer was considered to be a better indication of the effectiveness of the fourth and fifth writing assignments. The choice of distribution was recorded for each problem and for each student.

One problem on the probability and statistics final focused on normal distributions. The first three parts featured a normal with mean and standard deviation given and asked for probabilities before, between, and after given values. The last three parts were similar but stipulated a standard normal. For this study either a student obtained a correct answer or failed to find a correct answer for each part of the question. The number (zero to six) of parts answered correctly by each student was recorded.

Two problems on the final focused on the choice of distribution necessary to solve a discrete probability distribution problem. For the same reasons as for the third test the overall correctness of the answer was not considered to be important. The choice of distribution was recorded for each problem and for each student.

For both courses (probability and statistics and calculus) the overall score on the final exam was recorded.

Analysis of the Data

There were five problems (three on a test and two on the final) that required a choice of probability distribution. A simple and effective way to study the responses is to divide the choices into two categories: the correct distribution was chosen and its opposite. Answers that could not be deciphered and missing answers were grouped together with the incorrect responses. For each section (revision and no revision) and for each question the proportion of correctly chosen distribution can be formed. For each of the five questions the proportions for the two sections can be compared to each other. This yields a total of five comparisons. If the fourth and fifth writing assignments helped the students in making these choices, any benefit from revisiting the material through revision should produce a statistically significant difference between the proportions.

The key question is whether or not the proportions are equal. Under these circumstances a pooled procedure is believed to be a more powerful test than one using separate estimators (Milton & Arnold, 1990). A two-tailed test was applied because a null hypothesis of equality implies an alternate hypothesis of unequal (smaller or larger) proportions. For problems one through three (those on the test) there were 24 students who were revising their work and 30 who were not. For problems four and five (those on the final) the numbers were 25 and 25. Rather than preset a significance level and performing a hypothesis test, the value of the test statistic was found and then a judgment of how common the test statistic was made (making this a significance test). The statistic for the pooled test is formed in two steps. First the weighted mean (denoted by p) of the proportions p_1 and p_2 (with weights equal to the corresponding population sizes n_1 and n_2) is calculated. Then the value of p is substituted into $(p_1 - p_2)/\sqrt{p(1-p)(1/n_1 + 1/n_2)}$. The results are shown in table one. The last row of this table (labeled "probability") shows the probability of seeing a value of the test statistic at least as extreme as its actual value. This probability is based on the standard normal distribution and a two-tailed test. None of the five values are unusual (although problem four comes close with a 25% chance), so the null hypothesis can't be rejected for the five comparisons.

The most natural way to compare the scores on the final exams is to compare means for students with and without revision within a particular course. For normally distributed populations the Smith-Satterthwaite procedure works well for populations with and without equal population variances, making a separate test for equality of variances unnecessary (Milton & Arnold, 1990). Significance testing rather than hypothesis testing was used. The results are shown in table two. Neither of the two values of the test statistic was unusual (using a t-distribution with 24 degrees of freedom), so equality of the means cannot be rejected.

As a precaution a non-parametric test was also applied to the final exam scores. The Wilcoxon Rank-Sum Test can test for equality of medians for two populations with any underlying distribution. For the probability and statistics class an equal number of students in each section took the final, so either sum could be W_m, the sum of the ranks corresponding to the smaller sample. Let W_m=W_{WI} (an arbitrary choice) makes W_{WI}=592.5 and W_{WE}=682.5. Applying a hypothesis test at the 5% significance level shows that equality of medians can't be rejected. For the calculus class the WI class was smaller, so the sum of the ranks for students in this class has to be the test statistic W_{WI}. With W_{WI}=258 a hypothesis test at the 5% significance level again shows the equality of medians can't be rejected.

Conclusions

For all seven tests the null hypothesis was equality between classes. This hypothesis means the WE and WI students were equally likely to identify a distribution as a first step to solving a word problem and they received, on average, equal scores on the final. The null hypothesis could not be rejected in any of the seven tests. This indicates that revision, as implemented in this study, did not increase the students understanding of the subject.

One possible reason for this similarity may be the revision process itself. The first draft and the second were typeset, printed, and handed in. This requirement was meant to simulate the true WI environment in which written works are viewed as professional works worthy of typesetting. Although this extra step might distract students from the actual content of the paper (the most important element) it may also be the case that a handwritten paper might be put off to the last moment (say, the morning it is due), again producing a paper with poor content. The feedback stressed content but did not ignore other elements such as clarity, length, spelling, and grammar. Given several things to concentrate on it is possible that revising the content may receive a lower priority that it should. Given a busy schedule with several classes, family issues, and the need to work it is possible that revision receives less time than it should. A fourth reason may be the resistance to change in any major way what was written in the first draft. For the most part students will leave many sentences unchanged except for correcting misspelled words or adjusting commas before they reach a single sentence that receives a substantial work over.

In the 2008-2009 school year a continuation of this study is planned that will examine these issues. A major concern will be the identification of factors that may lessen the impact of thinking twice about a particular problem. The precise form of these changes is not clearly defined at this moment in time. It may be the case that different assignments may stress particular techniques (especially if a small number of instructors participate) or that different instructors may choose techniques that work well for them (if there are many instructors).

Acknowledgements

Special thanks goes to Ian Gardiner. At the time of the study Ian was a Writing Fellow in the York WAC program. He started the study by recruiting its investigators and by relentlessly tracking down and handling all of the necessary paperwork and permissions. I would also like to thank Farley Mawyer, a Professor at York College, for teaching the two sections of calculus and for active participation throughout the study. Last, but not least, I would like to thank Victor Crome, a Writing Fellow at York College, for sharing his knowledge of WAC and writing studies in mathematics with the investigators.

	Problem One	Problem Two	Problem Three	Problem Four	Problem Five
p_1	0.5167	0.6667	0.6667	0.4800	0.4000
p_2	0.5000	0.6000	0.7000	0.6400	0.4800
p	0.5185	0.6296	0.6852	0.5600	0.4400
Test statistic	0.30	0.50	0.26	-1.14	-0.57
Probability	0.7642	0.6170	0.7948	0.2542	0.5686

Table 1

	Prob. & Statistics		Calculus	
	No revision	With revision	No revision	With revision
Sample size	25	25	20	13
Mean	74.32	68.96	70.55	78.69
Standard deviation	16.46	18.41	20.26	19.47
Test statistic	-1.085		1.158	

Table 2

†*Lewis L. Lasser, Ph.D.*, York College/CUNY, USA

References

Heid, M.K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education*, 19, 3-25.

Judson, P.T. (1990). Elementary business calculus with computer algebra. *Journal of Mathematical Behavior*, 9, 153-157.

Milton, J.S., & Arnold, J.C. (1990). *Introduction to Probability and Statistics: Principals and Applications for Engineering and the Computing Sciences*. New York: McGraw-Hill.

Palmiter, J.R. (1991). Effects of Computer algebra systems on concepts and skill acquisition in calculus. *Journal for Research in Mathematics Education*, 22, 151-156.

Porter, M.K., & Masingila, J.O. (2000). Examining the effects of writing on conceptual and procedural knowledge in calculus. *Educational Studies in Mathematics*, 42, 165-177.