**VALUE ADDED: ENGLISH 1010 VS. ENGLISH 2010 YEAR II**

**Craig Petersen, Director**

**Office of Analysis, Assessment, and Accreditation**

**and**

**Kathryn Fitzgerald, Associate Professor**

**Department of English**

**Utah State University**

**March, 2006**

Note: The results that follow are based on data collected Fall, 2005. The methodology used for data analysis and the format used to present the results are similar to those of an April, 2005 report based on data collected Fall, 2004. That earlier report can be accessed at http://aaa.usu.edu/assessment/pdf/ValueAddedEnglish1010andEnglish2010.pdf

# **INTRODUCTION**

As part of the university’s overall assessment of instructional effectiveness, USU’s Office of Analysis, Assessment, and Accreditation and the Department of English cooperated on a study to evaluate the collective improvement in student writing and research skills from the time they start the department’s freshman-level writing course (English 1010) to the time they complete the sophomore-level course (English 2010). The objective of the study was to measure the “value added” by the two courses.

English 1010 (Introduction to Writing: Academic Prose) is a three credit course in which students learn “skills and strategies for becoming successful academic readers, writers, and speakers, such as how to read and write critically, generate and develop ideas, work through multiple drafts, collaborate with peers, present ideas orally, and use computers as writing tools.” English 1010 satisfies part of the Communication Literacy requirement of the university’s general education program. However, this requirement can also be satisfied (and English 1010 waived) based on a student’s performance on the relevant AP, CLEP, ACT, or SAT tests. Most students take English 1010 during their freshman year, but it is also available as a high school concurrent enrollment course.

English 2010 (Intermediate Writing: Research Writing in a Persuasive Mode) is also a three credit course. To enroll, students must have successfully completed English 1010 or have satisfied one of the waiver options. The course is described as “Writing of reasoned academic argument supported with appropriately documented sources. The course focuses on library and Internet research, evaluating and citing sources, oral presentations based on research, and collaboration.” Completion of English 2010 satisfies the USU’s general education Communications Literacy requirement.

# **METHODOLOGY**

The preferred approach to measure value-added would be to conduct a longitudinal study that compares the work of freshman students as they begin English 1010 with their work after they have completed English 2010 as sophomores. Under ideal conditions, such a study could be finished in a year. Unfortunately, USU is in a unique situation where many of its students (especially male students) interrupt their education for a two year period between their first and second year of college. In addition, many of the better students are able to waive English 1010 and move directly to English 2010. Thus, the longitudinal analysis would not include a broadly based sample of USU students.

The alternative approach used for this study was to compare performance of a random sample of student just starting English 1010 with that of a random sample near the completion of English 2010.

During Fall Semester, 2005, about 2,000 students were enrolled in English 1010 and about 800 took English 2010. During the first week of class, all English 1010 students were assigned to write a 750-1000 word essay based on one of two prompts:

1. Derek Bok’s article, Protecting Freedom of Expression on Campus. [[1]](#endnote-1)

For several years, universities have been struggling with the problem of trying to reconcile the rights of free speech with the desire to avoid racial tension. In recent weeks, such a controversy has sprung up at Harvard. Two students hung Confederate flags in public view, upsetting students who equate the Confederacy with slavery. A third student tried to protest the flags by displaying a swastika. These incidents have provoked much discussion and disagreement. Some students—especially minorities—have urged that Harvard require the removal of symbolic displays, arguing that they are insensitive and unwise because any satisfaction they give to the students who create them is far outweighed by the discomfort they cause to many others. On the other hand, I think that the displaying of such symbols falls within the protection of the free-speech clause of the First Amendment. Rather than prohibit such communications, I feel that it would be better to ignore them, since students would then have little reason to create such displays and would soon abandon them. If ignoring these acts is not possible, the wisest course would be to talk to those responsible, seeking to educate and persuade rather than to ridicule and intimidate, recognizing that only persuasion is likely to produce a lasting, beneficial effect.

Or

2. Rich Lowry’s article “Boys Will be Boys” [[2]](#endnote-2)

Your 8-year-old son who has trouble reading or little interest in picking up a book could benefit from the Larry Summers controversy. That’s because from out of the ashes of the Harvard conflagration is rising a nugget something valuable. The Harvard president, as everyone now knows, speculated that men might be overrepresented for genetic reasons in the top jobs in science and engineering at universities.

In response to the flap, *Time* magazine ran a cover story featuring the work of Leonard Sax, author of the new book, *Why Gender Matters: What Parents and Teachers Need to Know About the Emerging Science of Sex Differences*. Sax might simply have been dismissed as a Neanderthal not too long ago.

As Sax explains, at the heart of the debate about gender is a paradox: To ignore the hard-wired differences between boys and girls is to perpetuate gender stereotypes. That’s because ignoring those differences means we will continue to fail to teach many boys how to read and many girls how to do math and science.

As it happens, the gender-insensitive American education system hurts everyone. Take boys and reading. According to a National Endowment for the Arts survey, between 1992 and 2002 the gap between young women and young men in reading widened considerably. High school seniors who are girls score on average 16 points higher than boys on a reading test given by the National Assessment of Education Progress.

The flip side of this is girls when it comes to math and science—they develop more slowly. They will suffer the same discouragement as boys if they are pushed too soon, or in the wrong way. Sax says that at age 12, for instance, girls are less interested in “pure math” than boys, so problems have to be presented with practical applications.

It is obviously difficult to be mindful of these differences in coed classrooms, let alone coed classrooms devoted to the proposition that gender is a meaningless social construct. The institution of single-sex education, long ago tossed in the ash bin of history, would better serve both genders. Girls who go to all-girl schools are six times more likely than girls in coed schools to major in math or science in college.

The first step to overcoming gender, it turns out, is admitting how much it matters.

Most English 1010 and 2010 students responded to the Bok prompt. The Lowry topic was used in about one-fourth of the sections in both English 1010 and English 2010.

The students were told they were writing for an academic audience and were encouraged to use other sources. Two copies of their completed essay were to be accompanied by a questionnaire that asked about how much they had read and written in recent years and what types of outside information they had used in the process of writing their essay. The questionnaire also included questions designed to provide demographic data about the students. The same essay and questionnaire assignments were given to all English 2010 students during the last week of the course during Fall, 2005. In both courses, the class instructor graded the essay and used the score as a factor in determining grades.

From the population of all essays/questionnaires submitted, 200 were randomly chosen from English 1010 and 200 from English 2010. Each essay/questionnaire was given an identification number and then all information that could identify the students or the course level was removed from the essays. The questionnaire information was entered into a data base. In the previous year of the study, the high school GPA and ACT Composite and ACT English scores were also included in the data base. That was not possible with the 2005 data because Utah State University was in the middle of a transition to Banner software and the information was not readily and consistently available.

To reduce possible inconstancies in grading, the essay grades assigned by the course instructor were not used in this study. Rather, a team of fourteen graduate students who had taught one or both of the courses was assembled to re-grade the essays. This team was given a prescribed grading scheme and carefully trained through practice grading of essays not included in the sample. The graders were instructed to assign a score of 1 (low) to 6 (high) for each of the following nine categories:

1. Purpose
2. Audience
3. Presentation of multiple/oppositional views
4. Organization and flow
5. Appropriateness/quality of concepts
6. Pertinence of evidence
7. Citation of evidence
8. Writing style
9. Surface mechanics

The re-grading was done in one day and each essay was independently scored by two readers.

# **RESULTS**

Using the identification numbers, the total and component scores for each essay (as assigned by the raters) were appended to the data base. The result was an entry for each student in the sample that included demographic data, average essay scores, and information on sources used by the student to write the essay. These data form the basis for the findings reported in this section.

## **Demographic Data**

The information from the questionnaire allows the demographic characteristics of the English 1010 and English 2010 students in the sample to be compared. Fifty-two percent of those in 1010 were female, compared to 49% in 2010. The median age for the 1010 students was 19.3 years, while that of the 2010 students was 21.1 years. The nearly two years difference reflects the choice of many students to stop out between their freshman and sophomore years.

One of the questionnaire items asked how many college credits the student had completed. The distribution is shown in the table below. Nearly three-fourths of 1010 students had completed eighteen credits or less. This may reflect the fact that many students with AP or concurrent enrollment credit are able to waive English 1010.

**College Credit Completed**

| **Credits** | **English 1010** | **English 2010** |
| --- | --- | --- |
| 0-9 | 38% | 0% |
| 10-18 | 33% | 0% |
| 19-27 | 16% | 1% |
| 28-36 | 4% | 5% |
| 37-45 | 6% | 29% |
| 55-63 | 1% | 20% |
| > than 63 | 3% | 44% |

Students in English 2010 were asked how they met the university’s English 1010 requirement. The table below indicates that about half took a traditional 1010 course at USU or at another college or university. About one-fourth were able to get the requirement waived because of AP credit or their ACT/SAT score and about one-fourth took a high school concurrent enrollment course.

**Met English 1010 Requirement**

|  |  |
| --- | --- |
| **Method** | **Percent** |
| USU English 1010 or comparable | 53% |
| Waived by AP, CLEP, or ACT | 23% |
| Concurrent Enrollment | 25% |

The questionnaire asked how many papers of at least two pages in length the students had written during their junior and senior years in high school. As shown below, the English 1010 students reported slightly more papers written, but there was not a large difference between the two groups.

**Number of Papers Written Jr/Sr High School Year**

|  |  |  |
| --- | --- | --- |
| **Number** | **English 1010** | **English 2010** |
| 0 | 3% | 1% |
| 1-4 | 9% | 12% |
| 5-8 | 22% | 27% |
| 9-12 | 23% | 21% |
| 13-16 | 15% | 17% |
| 17-20 | 12% | 6% |
| > than 20 | 16% | 16% |

The students were also asked how many books (excluding textbooks) they had read during the last two years. In the case, the timeframe for the English 1010 students was the last two years of high school, while for the English 2010 students it included part of their college time. The table indicates that the 1010 students had done more recreational reading.

**Number of Books Read Last Two Years**

|  |  |  |
| --- | --- | --- |
| **Number** | **English 1010** | **English 2010** |
| 0 | 2% | 2% |
| 1-4 | 18% | 28% |
| 5-8 | 28% | 24% |
| 9-12 | 20% | 17% |
| 13-16 | 12% | 7% |
| 17-20 | 6% | 4% |
| > than 20 | 15% | 18% |

## **Use of Outside Sources to Write Essays**

One of the objectives of the study was to determine how the research habits of students were affected by going through English 2010. The questionnaire asked if, in writing, their essay, the student had looked for outside sources. It also asked if they had used the library to find sources and what kinds of outside sources were located. The comparative results are shown in the following tables.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Looked for Outside Sources** | | | **Used Library to Find Sources** | | |
|  | **English 1010** | **English 2010** |  | **English 1010** | **English 2010** |
| Yes | 66% | 66% | Yes | 9% | 20% |
| No | 34% | 34% | No | 91% | 80% |

**Types of Sources Found**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Books** | **Websites** | **Journal Articles** | **Newspapers & Magazine Articles** |
| English 1010 | 8% | 55% | 8% | 12% |
| English 2010 | 10% | 49% | 16% | 16% |

The data indicate that the two groups were equally likely to have looked for outside sources, but the 2010 students were about twice as likely to have used the library. However, neither group had much contact with librarians. The data also demonstrate the importance of the Internet as a research tool for today’s college students. Those at both course-level were far more likely to have used websites for information than any other source.

## **Total Essay Scores: Comparison of English 1010 and English 2010**

The essay scores used in this study were the average of the scores assigned by the two graders (Variation in grading is discussed later). The maximum score on the essay was 54 points. Aggregating the scores for both of the prompts, the mean score for the English 1010 essays was 28.24, while the mean for the English 2010 sample was 34.52—a gain of 6.28 points of 22.2%. Possible scores on each of the nine components used to grade the exam ranged from 1 to 6. As shown in the table below, the mean score for the 2010 students was higher than for 1010 students for each of the nine components. The largest gain was in citation of evidence.

**Average Scores on Essay Score Components**

|  |  |  |  |
| --- | --- | --- | --- |
| **Grading Component** | **English 1010** | **English 2010** | **Gain** |
| Purpose | 3.65 | 4.40 | 0.75 |
| Audience | 3.44 | 4.10 | 0.66 |
| Multiple Views | 3.32 | 3.94 | 0.62 |
| Organization | 3.51 | 4.09 | 0.58 |
| Quality of Concepts | 3.09 | 3.71 | 0.62 |
| Pertinence of Evidence | 2.67 | 3.38 | 0.71 |
| Citation of Evidence | 1.81 | 2.80 | 0.99 |
| Style | 3.30 | 4.04 | 0.74 |
| Surface Mechanics | 3.46 | 4.06 | 0.60 |
| Total Score | 28.24 | 34.52 | 6.28 |

As previously mentioned, the 2005 sample included essays based on two different prompts—Bok and Lowry. On average, students scored higher on Lowry than on Bok as shown below.

**Average Scores by Prompt**

|  |  |  |
| --- | --- | --- |
|  | Lowry | Bok |
| English 1010 | 30.99 | 27.44 |
| English 2010 | 35.73 | 34.13 |
| Gain | 4.47 | 6.69 |
| % Gain | 15.3% | 24.4% |

Note that the gain of Bok was substantially greater than for Lowry. The difference in gain do not distort the overall average scores and gain because the proportions of Lowry for both groups were both about one-fourth. Dividing the nine score components based on the two prompts, it was determined that the English 2010 students had higher average scores on each component, but the gain was less for each component for those who responded to the Lowry component.

The gains for English 2010 shown in the above table may be misleading because they do not take into account other factors that can affect scores. Specifically, students with good writing skills are more likely to have waived English 1010 using the ACT, AP, or CLEP options previously mentioned. Consequently, there are almost certainly differences in ability between the English 1010 and English 2010 samples used for the study. Part of the gains may reflect these ability differences rather than improvement that came as a result of taking the courses. In addition, students in 2010 are, on average, older than those in 1010 and their greater life experience and maturity may be positively affect their scores.

## **Total Essay Scores: Regression Analysis**

A more sophisticated approach to analyzing the effect of having completed the two courses is to use multiple regression analysis. This approach is a statistical technique than can quantify the influence of various factors that are expected to affect essay scores.

The starting point for using regression analysis is to identify the important factors that might affect the students’ essay scores. Some are intangible, such as personal motivation and effort and subjectivity in the grading of the essays, but others can be measured. For this study, it was assumed that the measurable factors that affected essay scores were age, gender, whether or not the student looked for outside information in writing her/his essay, whether the student was in English 1010 or English 2010, and whether the student was responding to the Bok or the Lowry prompt.

Multiple regression was then used to estimate the impact of each of these measures on total essay scores. The results are presented as an equation whose coefficients are a quantified estimate of the impact of each factor. However, because of the nature of statistical analysis, some of the coefficients may not represent actual impacts on scores. Remember that the study relied on taking a random sample of essays written by students in the two courses. If a different random sample had been selected, there would have been some differences in the data. It is possible that using multiple regression to analyze the effects of, for example, gender in one random sample might indicate a substantial difference between scores of men and women, while applying the same methodology to another sample would show a smaller difference or even no difference.

To assess whether a factor really does affect scores or whether the result is a “statistical aberration” based on the particular sample selected, the traditional approach is to use a t-test. This test estimates the probability that a given coefficient is not zero, meaning that the factor really does have an impact on scores. The t-test relies on the size of a measure called the “t-statistic.” For the sample sizes used in this study, a general rule of thumb is that if the t-statistic for a particular factor is greater than 2.0 or less than -2.0, then there is a 95% probability that the factor has an impact on test scores. The size of the coefficient of each factor is an estimate of how much an impact the factor has.

The first question to be considered using multiple regression analysis is whether essays written by students near the end of English 2010 are better than those written by students at the start of English 1010. The regression analysis allows the impact of having taken the two courses to be estimated once the influence of the other factors has been taken into account.

The “2010” variable takes on a value of “1” if the essay was written by and English 2010 student and “0” if by an English 1010 student

“Age” is measured in years.

“Female is a “1” for women and “0” for men.

“Info” is “1” if the student reported using outside sources and a “0” otherwise.

The prompt is accounted for by assigning a “1” if the student was responding to Lowry and a “0” if Bok.

Results are shown below. The t-statistics for the regression equation are the numbers below the coefficients.

Score = 20.76 + 5.46(2010) + 0.42(Age) + 0.84(Female) + 4.64(Info) + 3.71(Lowry)

(5.59) (2.05) (0.86) (4.72) (3.44)

Although 400 essays were sued in the sample, data for all the variables in the equation were only available for 355 students. Note that the equation has a constant term of 20.76. This number represents the combined influence of other factors that affected scores. One result of regression analysis is a number called the R2 or coefficient of determination. It is a measure of the percentage of the variation in essay scores that can be explained by the variables (i.e., 2010, age, female, info and prompt) in the equation. For the above equation, the R2 is 20%. This means that about 80% of the variation in the student essay scores resulted from factors that were not measured, such as motivation and effort. The low R2 may also reflect the grading problems discussed later in this report.

The regression results suggest that scores of females were, on average, higher than those of males, but the t-statistic of 0.86 suggests that there is no real difference based on gender. Age is positively correlated with scores and the t-statistic is greater than 2.00. Students who searched for outside information to write their essays and those who got the Lowry prompt had higher scores, and the large t-statistics for each variable indicate a high probability that these factors really did make a difference.

Of particular interest to this study is the “2010” variable. The estimated coefficient is 5.46 and the t-statistic is 5.59. The interpretation of these numbers is that, once other factors that could influence essay scores are taken into account, on average, the students who were completing English 2010 had higher scores. The t-statistic suggests that the difference is real and not just a statistical aberration—writing proficiency is improved by completing the two courses.

Another question of interest is whether writing ability is influenced by the way in which students met the university’s English 1010 requirement. How do scores compare for students who took traditional 1010 courses with those for students who took the course in high school as concurrent enrollment or who used ACT, AP, or CLEP scores to waive the basic course? To answer this question, the influence of other factors that affect scores, such as age, gender, and experience (as measured by college credits taken) must be accounted for. The regression equation below is based on 172 essays written by students in English 2010. The last two variables reflect how the student met the English 1010 requirement. The “Waived” variable compares how students who met the English 1010 requirement by having high AP, CLEP, or ACT scores scored versus those who took a traditional English 1010 course. The “Con En” variable compares scores of students who took English 1010 in high school versus those who took the traditional course.

Score = 18.21 + 0.36(Age) + 1.09(Female) + 3.57(Lowry) + 0.89(Credits)

(0.44) (1.59) (2.25) (0.44)

+ 6.25(Waived) + 1.43(Con En)

(3.32) (0.82)

As expected, students who were writing to the Lowry prompt had higher scores. The only other coefficient whose coefficient was statistically significant was for students who had waived the traditional English 1010 course. Their estimated score was about six points higher than those who took traditional English 1010. This result is not surprising. Students with high enough ACT, AP, or CLEP scores to skip English 1010 would tend to be high ability students. The R2 for this regression equation was only 12%, indicating the presence of other important factors that determined essay scores.

## **Components of Essay Scores: Regression Analysis**

Thus far, statistical analysis has only been used to evaluate total essays scores. It is also possible to use multiple regression analysis to gain insights into the factors that affected scoring of the nine components used for grading. Each row in the table below shows the results for the grading component listed in each column. The results are based on 355 essays written by students in the two courses. In each case, scores on the component were regressed on the previously-described variables—2010, age, gender, info, and the prompt to which the students were responding. The numbers in the table are the estimated impact of the variable. The size of the coefficient can be placed in context by recalling that allowable scores on each component ranged from 1 to 6. Thus, a coefficient of 0.50 for the purpose row and the 2010 column indicates that the estimated gain in the purpose score for English 2010 students compared to English 1010 students was 0.50. The asterisks indicate that that t-statistic for the coefficient was greater than 2.00 or less than -2.00, indicating a high probability that there is an actual impact.

**Estimated Impact on Essay Component Score**

| **Component** | **2010** | **Lowry** | **Info** | **Female** | **Age** |
| --- | --- | --- | --- | --- | --- |
| Purpose | 0.71\* | 0.44\* | 0.17 | 0.03 | 0.03 |
| Audience | 0.60\* | 0.30\* | 0.36\* | 0.24 | 0.06\* |
| Multiple Views | 0.48\* | 0.58\* | 0.35\* | 0.07 | 0.06\* |
| Organization | 0.51\* | 0.38\* | 0.16 | 0.08 | 0.03 |
| Concept Quality | 0.49\* | 0.43\* | 0.41\* | 0.00 | 0.06\* |
| Pert of Evidence | 0.56\* | 0.42\* | 1.29\* | 0.00 | 0.05 |
| Cite of Evidence | 0.91\* | 0.44\* | 1.51\* | 0.16 | 0.03 |
| Style | 0.67\* | 0.41\* | 0.19 | 0.11 | 0.05 |
| Mechanics | 0.53\* | 0.30\* | 0.20 | 0.16 | 0.04 |

Again, the effect of having taken English 2010 is of special interest. All of the estimated coefficients for “2010” are positive and all have asterisks. The interpretation is that, when other factors that affect scores are taken into account, completion of English 2010 tended to increase scores for each of the nine grading components. The largest estimated effect was for citation of evidence.

## **Discrepancies in Grading by Essay Readers**

One other factor affecting the results presented in this report needs to be considered. As previously noted, the scores used for each essay in the sample were the average of the scores assigned by the two readers. But grading is an art rather than a science. Although the graders were carefully trained, there is an element of subjectivity in the process. This was verified by analyzing differences in the essay scores each reader assigned to each essay. The maximum possible score on an essay was 54.

The average difference in the two grader’s scores was 8.42 points for the English 1010 papers and 8.53 points for the English 2010 essays. In one case, the two readers disagreed by 36 points. For about one-sixth of the essays the difference was 16 points or more and the average scores of the 14 raters involved in the project varied from 25 to 43. The problem seems to have been with a small number of raters and not the entire group. Analyzing those instances when the two scores differed by 16 points or more, it was determined that two of the raters were involved in about 40% of the cases. One of those raters had a very high average score, which helps explain the disparity, but the other was near the middle of average scores of raters.

The low R2 value for the regression equations may partially be due to the subjectivity in grading the essays. That is, a large component of the variation in scores may reflect disagreements among the raters.

# **SUMMARY AND CONCLUSIONS**

In education assessment, the most persuasive evidence of success is usually considered to be the value that is added to students as they progress through a curriculum. The objective of this study was to evaluate the value added in writing skills as a result of students taking USU’s two introductory composition courses. The challenging part of the analysis was to separate factors that may be associated with higher essay scores such as age, gender, acquisition of ideas from outside sources, and innate ability from the benefit that students receive from taking the courses. This task was accomplished using multiple regression analysis.

Holding the effects of other factors, it was determined that total scores on a sample of 400 essays were about five points (20%) higher for English 2010 students than for those in English 1010. It was also determined that scores on each of the nine components used to grade the essays were significantly higher for those students who were nearing completion of English 2010. The implication is that there does seem to be value added associated with the university’s introductory writing courses. It should be noted, however, that the results reported in this report are not as conclusive as those from the first year of the study because ACT scores were not available this year to account for differences in students’ ability.

An additional finding of the study pertains to grading of papers in composition classes. Even with a structured grading scheme, extensive training, and experienced graduate students reading the essays, the initial scores assigned by the two readers of each paper varied significantly in larger number of cases. The inconsistency in grading was similar to that which occurred when the study was undertaken using Fall 2004 data. It implies that there is a substantial element of subjectivity in the evaluation of students and that student course grades may, to some extent, reflect who they are assigned to as an instructor and not just how well they write.

1. REFERENCES

   Boston Globe 25 May 1991. [↑](#endnote-ref-1)
2. National Review. [↑](#endnote-ref-2)