The Effects of Grammar Testing on the Writing Quality
and Reduction of Errors in College Freshmen’s Essays

Wes Davis, Ph.D.
Associate Professor of English

and

Kelley Mahoney, M.A.T.
Associate Professor of English

Division of Humanities
Dalton State College
213 North College Drive
Dalton, Georgia 30720
Phone: (706) 272-4444
E-Mail: wdavis@em.daltonstate.edu
ABSTRACT

The Effects of Grammar Testing on the Writing Quality and Reduction of Errors in College Freshmen’s Essays

This experimental, statistical study investigated the effects that the testing of grammar and writing mechanics would have on the overall quality and reduction of errors in college students’ essays for freshman composition. In the experimental group of 42 students, the professor assigned several exercises in grammar and mechanics as a review related to composing skills and then gave two major tests on proofreading essays for grammatical errors. However, the other professor did not give these grammar tests to the 41 students in the control group. The study used “T-tests” for statistical analysis on pretest and posttest essays, which each of the 83 students had written.

On overall writing quality, the faculty raters holistically scored the students’ essays, using a scale from 1 (failing) to 4 (superior). Since the two raters scored each student’s pretest and posttest essay, each essay had a combined score resulting in a scale from 2 (failing) to 8 (superior). The results showed that the 42 students of the experimental group who tested on grammar had a mean pretest essay score of 2 and a mean posttest essay score of 4.53, showing a gain of 2.53. Statistically, these students made very highly significant gains in overall writing quality. The 41 students of the control group not tested on grammar had a mean pretest essay score of 2.66 and a posttest score of 4.49, showing a gain of 1.83. These students’ also made very highly significant gains in overall writing quality, although the experimental group’s posttest essay scores were still significantly higher than the control groups.
On traditionally serious grammatical errors, the experimental group had a mean number of 1.5 serious errors on the pretest essay and a mean number of 0.93 error on the posttest, a reduction of 0.57 which was statistically significant. The control group’s mean number on the pretest essay was 1.23 serious errors and a mean posttest number of 0.64 error, a reduction of 0.59 which was also significant; however, there was no significant difference between both groups in the reduction of serious errors. For the less serious “minor” errors, the experimental group had a mean number of 7.4 minor errors on the pretest essay and a mean number of 5.12 errors on the posttest essay, a reduction of 2.28 errors which was highly significant. The control group had a mean pretest essay number of 12.5 minor errors and a mean posttest essay number of 7.42 errors, a reduction of 5.08 errors which was very highly significant. Moreover, statistics showed a significant difference between both groups’ posttest essays in the reduction of these minor errors, with the control group making more significant reductions in the minor errors than the experimental group. The researchers concluded that the two major grammar tests on proofreading two essays for errors may have had some effect on the experimental group’s gains in overall writing quality for correctness. However, these tests appeared to make no difference between both groups, by having very little if any effect on the students in the experimental group to reduce the number of errors significantly in their essays.
Introduction

Is there any correlation between good grammar and good college writing? Does testing in grammar, writing mechanics, and punctuation really affect the overall quality of freshman college students’ essays? Patricia Bizzell (2000) has stated, “Correctness is a perennial issue in basic writing instruction,” which “has focused on the problem of how to enable under-prepared college students to write correct academic discourse” (p.4). As far back as 1963, research has examined and seriously questioned just how effective teaching grammar and mechanics is on high school and college students’ writing quality, usually reporting a “negligible effect” (Braddock, Lloyd-Jones & Schoer, 1963; Elley, Barham, Lamb & Wyllie, 1976, 1979; Applebee, 1981; Morrow, 1984; Hartwell, 1985; Hillocks, 1984, 1986: Williamson, 1986). However, very few studies have statistically assessed how college freshmen’s writing quality has improved from instruction in grammar and mechanics, by examining the correlation between a reduction in grammatical errors and overall writing quality. This experimental, statistical study investigated the effects that testing students in grammar and writing mechanics would have on the overall quality and number of errors in first-year college students’ essays for freshman composition. In the experimental group of 42 students, the professor assigned several exercises in grammar and writing mechanics as a review and then gave two major tests on proofreading for grammatical errors in actual essays. However, the other professor did not assign these grammar proofreading tests to the 41 students in the control group. By analyzing the 83 students’ pretest and posttest essays, the researchers hypothesized that the students in the experimental group would show more significant
gains and outcomes in overall writing quality and more reductions in error counts than what the control group would demonstrate in these essays.

**Review of the Research Literature**

Grammar has always been an integral part of composition instruction, even in college. Virginia R. Monseasu (2002), editor of the *English Journal*, writes:

No matter what is said about alternative forms of language study, English teachers are still concerned about the teaching of grammar in their classes. Should we continue to teach the traditional grammar that has had such a strong foothold over the years? Should we teach grammar only as the need arises in student writing? Should we ignore grammar instruction completely, concentrating instead on . . . writing content? (p. 9)

In the article “Why Revitalize Grammar,” Dunn and Lindblom (2003) “know there are many effective writing teachers who understand that grammar is a tool for making meaning and not an end in itself” (p.43). In her article Patricia J. McAlexander (2000) writes:

Rei Noguchi, in his 1991 *Grammar and the Teaching of Writing*, argues that style is “just as global. . .as organization and content” (p.13) and that teaching grammar and mechanics can help students improve their style. Further, correctness is important, Noguchi points out, since “many readers, particularly in business and other professional settings, perceive [errors] as major improprieties” (p.14). A reason for the “negligible” effect of much grammar instruction, Noguchi speculates, is that “students,
though possessing sufficient knowledge of formal grammar, fail to apply that knowledge to their writing” (p.7). His conclusion advises moderation between extremes: composition faculty should teach grammar—but not at length and for its own sake. Rather, they should integrate grammar instruction with writing instruction and teach the most vital terms and the most frequently made errors (pp.17-18). (pp.124-125)

Noguchi mentions that teaching grammar and mechanics can improve students’ style, which mainly involves the grammar of syntax or sentence structure. In their study on grading grammatical errors, Briggs and Pailliotet (1997) discovered that when given the opportunity to respond to anything in students’ texts, most teachers (71% for Applebee, about 75% for Anson) chose “surface features” (Anson, p. 344). According to Briggs and Pailliotet, “these studies indicated that teachers often chose errors as the focus of their commentary” (pp.48-49). Briggs and Pailliotet’s “examination of samples of 50 pre-service [student] teachers’ written discourse about grammar and conventional English revealed that they had largely negative attitudes toward writers who made conventional errors” (p.46). Moreover, several articles in the English Journal show a revival of teaching grammar and conventional English at the secondary level of education (Blasé, McFarlan & Little, 2003; Sams, 2003; Perrin, 2003; Vavra, 2003; Ehrenworth, 2003; Johansen & Shaw, 2003; Doniger, 2003; Carroll, 2003; Dean, 2001).

Even for communicating on the job, employees see the importance of good grammar. In a survey of the workplace (Craig, 2001), 98% of the respondents claimed that correct spelling, grammar, and mechanics were important in their writing on the job. Moreover, Perry (1996) presents the results of a national study in which members of the
Society for Technical Communication rated 20 selected English-usage principles in grammar as important to professional writing. Another study (Davis & Stohrer, 1989) surveyed Department of Defense middle managers working for the U.S. Air Force; respondents identified grammar, syntax, and mechanics as among some of the most important writing skills, in addition to purpose of the message, audience analysis, and organization of the writing tasks. Also, West (1983) interviewed and surveyed Middle Tennessee Bar Association lawyers and legal secretaries/paralegals, discovering from them that necessary employee communication skills included vocabulary development and proofreading skills as well as grammar, organization of the content, punctuation, sentence structure, and spelling.

Over the last 20 years, research and pedagogy in technical/business communication have strongly suggested the need for skills in grammar and writing mechanics. Waltman and Smeltzer’s study (1988) explored the relationship between grammatical proficiency and a number of variables in an introductory business communication course, suggesting that there is a small but significant correlation between grammatical proficiency and overall performance in the course, and that grammatical competency is a predictor of successful course completion. In an earlier study, Waltman (1983) sought to determine the relationship between grammar competency and success in a business communication course by comparing 236 students’ English competency test scores with grades on a formal written report. The results indicated students’ grades for formal reports and for the course could be raised by improving their grammatical competency; the grammar test in the study proved to be a good predictor of success. Other studies in technical/business writing have investigated
such matters as active and passive voice verbs (Riggle, 1998; Rodman, 1994), choice of words and their strict meanings (Nadzieika, 1993), subject-verb or pronoun-antecedent agreement (Allison, 1993), syntactical and rhetorical characteristics (Myers, 1999; Myers, 1996), as well as other grammatical criteria in business letters (Goodin & Swerdlow, 1987). Other sources show the need for skills in grammar and mechanics and how to improve those skills in technical/business writing (Darvin, 2001; Robbins, 2001; Fatt, 2000; Gerson & Gerson, 2000; Carlisle et al., 2000; Tichenor, 1999; Gray, Ingram, Bodson, 1998; Spears, 1998; Burt, 1995; Samson, 1993; Killingsworth & Walter, 1990; Allen & Southard, 1987; Hall, 1986; Vaughn, 1985; Allison, 1983. This related research literature strongly supports this current study on assessing the effects of teaching grammar in writing not only for college but also for business and technology.

**Design and Procedures**

This experimental study used statistical analysis to assess the effects of testing skills in grammar and mechanics on the overall quality of college freshmen’s essays. The purpose of this study was to discover any statistically significant differences in gains and outcomes on the writing quality and grammatical correctness of 83 first-year college students between their pretest and posttest essays: 42 in the experimental group and 41 in the control group.

After both groups had initially written a pretest essay within 60 minutes before any instruction began, the professor then isolated the teaching treatment for the 42 students in the experimental group by first reviewing chapters in the grammar handbook and then assigning the following exercises to review skills in grammar and writing
mechanics: (1) writing complete, complex sentences by creating subordinate clauses attached to independent clauses from a list of subordinate conjunctions and relative pronouns; (2) identifying and changing fragments into complete sentences; (3) using the correct punctuation to eliminate identified comma splices and fused/run-on sentences; (4) choosing the correct verbs and pronouns to agree with their respective subjects and antecedents; (5) taking two major tests by proofreading two essays of five paragraphs each to detect and correct errors in grammar, mechanics, and punctuation. The 41 students in the control group, however, did not do these same assignments. In fact, they never did the two major tests on proofreading the two essays for errors in grammar, mechanics, and punctuation.

The pretest and posttest essays were the data collected, which revealed the effects of each class’s instructional mode on the gains and outcomes in overall writing quality of these 83 students; these variables were measured in their writing samples before and after instruction (Sanders & Littlefield, 1975), because most theorists believe that a direct sample is the best way to measure writing ability (Dieterich, 1974; Cooper, 1975, 1977). Forty-two students in the experimental group and 41 students in the control group each wrote an impromptu essay, the best way each knew how, during the first 55-minute class on a choice of four unannounced topics. This writing sample served as the pretest in class at the beginning of the course before any instruction began. Then each student in both groups chose one of four different unannounced topics and wrote another impromptu essay of around 500 words as a posttest sample after instruction at the end of the course, again using 55 minutes of class time to plan, compose, revise, edit and correct the essay. The only difference between the two groups was that the experimental group
had also completed several exercises and had taken two major exams by proofreading and detecting errors in two essays to test their skills in grammar, word choice, spelling, writing mechanics, and punctuation before undertaking this posttest essay; however, the students in the control group did not receive these major grammar tests of proofreading two essays for errors before they wrote the posttest essays. Studies show that the pre-test/posttest design is one effective way to assess the effects of an educational treatment (Bloom, Hastings, & Madaus, 1971; Campbell & Stanley, 1963; Cronbach, 1963).

In evaluating writing quality, a general-impression holistic rating session using expert readers—generally experienced college English teachers trained to agree on certain global characteristics of a piece of writing—can produce acceptable reliability (Diederich, 1974; Cooper, 1975, 1977). For this study two college English faculty members served as raters, who did have several years of expertise in teaching and grading college freshmen’s essays, in addition to several years of training, experience and reliable agreement on holistic scoring for the Board of Regents’ undergraduate, state essay exam in the University System of Georgia, a requirement for graduation.

The two faculty raters read and scored a total of 166 randomly sorted essays, without knowing which were the 83 pretests or 83 posttests. The raters scored these essays based on the following scale: (1) Lowest failing score; (2) Minimally passing score; (3) good passing score; (4) highest passing score. The raters followed the same scale and grading criteria used to score Regents’ Testing Program essays in the Georgia University System. For any essay on which the raters disagreed by more than one point, a third experienced rater scored the piece in question; then the other raters used the two closest scores.
The study’s researchers then separately totaled the raw scores for each group’s pretest and posttest set of essays to get the mean score for each comparison group. For rate of agreement on the test scores, calculating the “product-moment correlation” determined the reliability between the two raters. For both groups’ pretest scores, the correlation coefficient was .96, indicating good reliability. For both groups’ posttest scores, the correlation was .81, implying adequate reliability between the two raters. In addition, the researchers categorized and counted the number of traditionally serious errors (e.g., fragments, subject-verb disagreements, comma splices and run-on sentences) in both of the 83 students’ pretest and posttest essays, as well as counting less serious “minor” errors (e.g., misspellings, pronoun/antecedent disagreements, and punctuation errors) in a separate category.

In the statistical analysis of these scores, “paired T-tests” determined whether there were significant differences between the pretest and posttest scores on the essays and differences between the experimental and control groups’ posttest scores and their reductions in grammatical errors. The confidence levels of statistical significance for 95% of the time on the “paired T tests” are based on the following “p” values: (1) * p < .05, significant; (2) ** p < .01 highly significant; (3) *** p < .001, very highly significant. In addition to these “p” values showing the “Significance of T” in the analysis of variance, the statistics also showed “degrees of freedom” (df).

**Analysis of the Results**

In overall writing quality, each of the 83 students wrote two essays, one pretest essay before instruction at the beginning of the course and one posttest essay after instruction at the end of the course. The two faculty raters scored each essay holistically
on a scale of “one” (1—poor/failing) to “four” (4—superior/passing). This procedure resulted in two scores on each pretest writing sample and two scores on each posttest sample for each student in the study. For statistical analysis (paired T-tests), the raters combined their two pretest essay scores on each student, and then they combined their two posttest essay scores, resulting in the very lowest possible score of “two” (2) for a failing essay and the very highest possible score of “eight” (8) for a superior, passing essay.

In overall writing quality, the results suggested that the 42 students in the experimental group seemed to benefit from taking the two tests of proofreading essays for errors in grammar and writing mechanics. This group started with a mean score of 2 on the pretest writing sample but ended with a mean score of 4.53 on the posttest essay, an improved change of 2.53 (See Table 1). The T-test in statistical analysis indicated that the difference between the experimental groups pretest and posttest scores was very highly significant ($T = -10.663; df = 41; p = .00001; *** p< .001$) in this main effect for this method of teaching college freshman composition: the two major tests reviewing skills in grammar and writing mechanics by having students proofread and detect errors in two essays (See Table 2). Statistically, the students made very highly significant gains in overall writing quality.
Table 1
Mean Scores of Writing Quality Change from Pretest to Posttest
between the Experimental Group and the Control Group

<table>
<thead>
<tr>
<th>Method/Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>2.00</td>
<td>4.53</td>
<td>2.53</td>
</tr>
<tr>
<td>Control</td>
<td>2.66</td>
<td>4.49</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Table 2
Significance Tests for Overall Writing Quality
between the Experimental Group and the Control Group

<table>
<thead>
<tr>
<th>Method/Group</th>
<th>df</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>41</td>
<td>-10.663 ***</td>
<td>.000011</td>
</tr>
<tr>
<td>Control</td>
<td>40</td>
<td>-11.849 ** *</td>
<td>.000058</td>
</tr>
<tr>
<td>Both Groups’ Posttests</td>
<td>69</td>
<td>2.65 **</td>
<td>.00099</td>
</tr>
</tbody>
</table>

* p < .05  ** p < .01  *** p < .001

The 41 students in the control group did not receive the teaching treatment of testing their skills in grammar and mechanics by doing the major exams of proofreading...
two essays for errors. This group started with a mean score of 2.66 on the pretest but
ended the course with a mean score of 4.49 on the posttest, an improvement of 1.83 (See
Table 1). The T-test in statistical analysis still indicated that the difference between these
students’ pretest and posttest scores was also very highly significant \( T = -11.84; \text{df} = 40; \)
\( p = .000058; *** p < .001 \) in overall writing quality, even without any extra tests in
proofreading essays for errors in grammar and mechanics (See Table 2).

Statistical analysis also indicated that the difference between the experimental and
control groups’ posttest scores was highly significant \( T = 2.65; \text{df} = 69; p = .0099; ** p
< .01 \). The experimental group’s mean posttest score was significantly higher than the
control group’s mean score on writing quality in their final essays (See Table 2). These
results strongly suggested that the students in the experimental group made greater gains
in overall writing quality and significantly higher outcomes on the posttests than the
students made in the control group.

Between pretest and posttest essays, statistics also measured the reduction of
traditionally serious errors: (1) subject-verb disagreements, (2) fragments, (3) comma
splices, (4) fused/run-on sentences. On the pretest essays, the experimental group had a
mean number of 1.5 serious errors and a mean number of 0.93 error on the posttest, a
reduction of 0.57 (See Table 3). The experimental group’s reduction in these errors was
statistically significant \( T = 1.75; \text{df} = 41; p = .04; * p < .05 \). The control group’s mean
number on the pretest essay was 1.23 serious errors and a mean posttest number of 0.64
error, a reduction of 0.59 (See Table 3). The control group’s reduction of these serious
errors was also significant \( T = 2.04; \text{df} = 40; p = .02; * p < .05 \). Between both groups’
posttest essays, however, statistics showed no significant difference in reduction of these serious errors \( (T = -.02; \ df = 79; \ p = .98; \ p > .05) \). The statistics are presented in Table 4.

Statistics further measured the reduction of less serious “minor” grammatical errors (e.g., misspellings, pronoun/antecedent errors, or punctuation errors) between pretest and posttest essays. The experimental group had a mean number of 7.4 minor errors on the pretest essay and a mean number of 5.12 errors on the posttest essay, a reduction of 2.28 errors (See Table 3). Statistically, the experimental group’s reduction of the minor errors was highly significant \( (T = 3.58; \ df = 41; \ p = .004; ** p < .01) \). The control group had a mean number of 12.5 minor errors on the pretest and a mean posttest number 7.42 errors, a reduction of 5.08 errors, shown in Table 3. Statistically, the control group’s reduction of the minor errors was very highly significant \( (T = 5.44; \ df = 40; \ p = .00014; *** p < .001) \). Moreover, statistics showed a significant difference between both groups’ posttest essays in the reduction of minor errors, with the control group making more significant reductions in these errors than the experimental group \( (T = 2.515; \ df = 70; \ p = .014; * p < .05) \). The statistical findings for both groups are presented in Table 4.

These results further implied that the students in the experimental group made highly significant statistical gains and outcomes in writing quality between the pretest and posttest essays because of the additional exercises and tests of proofreading essays for errors in grammar and mechanics. Even though the control group did make statistically significant gains between the pretests and posttests, the results suggested that the experimental group’s additional tests in skills on grammar and writing mechanics gave these students a statistically significant edge in the gains and outcomes of their writing quality over the students in the control group, thus supporting the researchers’
first hypothesis at the beginning of the study, as well as supporting the related research on teaching grammar and mechanics in college composition. However, the findings also suggested that the two major tests on proofreading essays for correctness had little effect on the outcomes in the posttest essays for reducing the number of grammatical errors between the experimental group and the control group. These findings appear to reject the second hypothesis that the reduction of these errors alone affected writing quality in both groups. Thus, correct grammar does not necessarily correlate with overall writing quality.

Table 3

Mean Scores for the Reduction in Errors from Pretest to Posttest Between the Experimental Group and the Control Group

<table>
<thead>
<tr>
<th>Method/Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental’s Serious Errors</td>
<td>1.5</td>
<td>0.93</td>
<td>0.57</td>
</tr>
<tr>
<td>Control’s Serious Errors</td>
<td>1.23</td>
<td>0.64</td>
<td>0.59</td>
</tr>
<tr>
<td>Experimental’s Minor Errors</td>
<td>7.4</td>
<td>5.12</td>
<td>2.28</td>
</tr>
<tr>
<td>Control’s Minor Errors</td>
<td>12.5</td>
<td>7.42</td>
<td>5.08</td>
</tr>
</tbody>
</table>
### Table 4

Significance Tests for the Reduction of Errors from Pretest to Posttest Between the Experimental Group and the Control Group

<table>
<thead>
<tr>
<th>Method/Group</th>
<th>df</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental’s Serious Errors</td>
<td>41</td>
<td>1.756 *</td>
<td>.04</td>
</tr>
<tr>
<td>Control’s Serious Errors</td>
<td>40</td>
<td>2.039 *</td>
<td>.02</td>
</tr>
<tr>
<td>Both Groups’ Serious Posttest Errors</td>
<td>79</td>
<td>-.024</td>
<td>.98</td>
</tr>
<tr>
<td>Experimental’s Minor Errors</td>
<td>41</td>
<td>3.589***</td>
<td>.00004</td>
</tr>
<tr>
<td>Control’s Minor Errors</td>
<td>40</td>
<td>5.445***</td>
<td>.000014</td>
</tr>
<tr>
<td>Both Groups’ Minor Posttest Errors</td>
<td>70</td>
<td>2.515 *</td>
<td>.014</td>
</tr>
</tbody>
</table>

* p < .05  ** p < .01  *** p < .001

### Conclusion

The researchers of this study concluded that the two major tests for detecting and correcting grammatical errors in essays may have had some effect on the experimental group’s significant gains in overall writing quality, at least for correctness. However, the findings strongly suggest that these grammar tests had very little if any effect on the students in this group for reducing the number of errors significantly in their own essays. Therefore, more studies on teaching useful skills in grammar and writing mechanics are needed to help college educators realize how they may benefit their students the most to make greater gains and learning outcomes in overall writing quality for freshman composition. This study, however, strongly suggests that having students take tests by proofreading essays to detect and correct grammatical errors will not necessarily carry
over into proofreading their own essays to reduce errors significantly, as demonstrated by the control group’s students who reduced errors significantly without being exposed to this type of testing in grammar and writing mechanics. According to this study, teaching students to detect and correct errors through exercises and tests has little effect on their ability to eliminate these errors in their own essays. In fact, teaching errors can be counterproductive in teaching students to write. This study calls for more research on what variables in composing and what teaching methods really affect overall writing quality, especially in improving first-year college students’ essays. Instead of exercises and tests to help reduce the number of errors, could the significant difference actually lie in each instructor’s grading policies of grammatical errors? Only more studies may give the evidence and the truth.
References


