The Effect of the Quantity of Writing Exams on Developing ESL/EL Writing Skills

Ali Osman ENGİN (*)

Turgay HAN (**)

Assiye BURGUCU (***)

Abstract: *Feedback in ESL/EFL writing has been inconclusive. In literature, several studies are available about the degree of the effect of the varying feedbacks (explicit vs. implicit; coded vs. uncoded, etc.) on the learners' motivation and success. The effect of optimum number of writing exams on the success of L2 learners may be important as much as the effectiveness of feedback types on learners' errors in their compositions. The aim of this study is to investigate the impact of number of writing exams on the linguistic errors in EFL/ESL student compositions. In this study there are two groups (e.g. experimental and control groups) consisting of 20 intermediate level students, each studying in the Department of Tourism Guidance at a state university in Turkey. Control group participants took three administrative examinations, which are two mid-term exams and one final exam and experimental group participants took three more exams. No feedback was provided for students compositions. The results showed that group students who took more three exams (e.g. experimental group) slightly outperformed those of control group.*

Keywords: *Feedback; EFL/ESL Writing, Error Treatment.*

İkinci Dilde Yazma Becerisinin Değerlendirilmesi Üzerine Nicel Bir Yaklaşım

Öz: *İkinci ve yabancı dil İngilizcede yazma becerileri geribildirimi sonuçsuz kalmaktadır. Alanyazında öğrencilerin motivasyonları ve başarıları üzerine farklı türlerde(açık, kapalı, kodlanmış veya kodlanmamış) geribildirimin etkilik derecesiyle ilgili birçok çalışmaya ulaşılabilir. Önemli sayıda yazılı sınavların öğrenenlerin ikinci dil başarıları üzerine etkisi,onların yazdıkları kompozisyonlarında karşılaşılan hatalarla ilgili olarak verilen çeşitli geribildirimlerin etkilikleri kadar önemlidir. Bu araştırma, birçok sayıda yazılı sınavların ikici/yabancı dil öğrencilerinin dilbilim hataları üzerine etkisini araştırmaktır. Bu çalışmada her biri bir Türkiye'de bir devlet üniversitesi Turizim Rehberliği bölümünde öğrenim gören 20 orta düzeye sahip öğrenciden oluşan iki grup (control grubu ve deney grubu) bulunmaktadır. Kontrol grubu katılımcıları tanımlayıcı iki ara ve bir final sınavına ve deney grubu katılımcıları da fazladan üç sınava daha alınmışlardır. Öğrencilerin kompozisyon çalışmalarına herhangi bir geribildirim verilmemiştir. Sonuç; üç adet daha fazla sınava giren öğrenciler (deney grubu)control grubundan daha fazla performans ortaya koyduklarını göstermiştir.*

Anahtar Kelimeler: *Geribildirim, İkinci/yabancı dilde yazma, Hata düzeltme*

*)   Assoc. Prof. Dr., Kafkas University, Faculty of Education, Department of Education Sciences. (e-posta: aosmanengin@gmail.com)

**)   Kafkas University, Faculty of Science and Letters, Department of English Language and Literature. (e-posta: turgayhan@yahoo.com.tr)

***)   Kafkas University, Faculty of Science and Letters, Department of English Language and Literature. (e-posta: sybrgc@gmail.com)

Introduction

Writing assessment has always been problematic. White (1985) asserts, "most testing of writing is poorly done, destructive to the goals of teaching, and improperly used" (p. 2). However, second language (L2) assessment studies indicate that certain forms of teacher feedback affect text quality more positively than others (Hedgcock & Lefkowitz, 1994). Through written feedback, L2 learners may overcome some challenges and understand the arbitrary nature of writing assessment and design; further, corrective feedback can be effective in improving students' accuracy (Van Beuningen, De Jong & Kuiken, 2008). Principally, grammar error correction in L2 writing classes has attracted most of the researchers to investigate the possible working ways of treating interlanguage errors of the learners. Error treatment is one of the key and important issues in L2 writing both teachers and students. Ashwell (2000) states that teachers believe that correcting the grammar of student writers' work will help them improve the accuracy of subsequent writing. Much of the related literature on the assessment of writing through error feedback helps L2 students to improve the accuracy and overall quality of their writing (Ferris, 1999; Kepner, 1991; Truscott, 1999). In this context, treating L2 errors in some way or another is thought to be a component of writing classes. Therefore, for majority of teachers responding to student writing by corrective feedback is an indispensable part of writing course and L2 writers aspire for teacher feedback on errors in their written productions (Ferris & Roberts, 2001). Even though there is a substantial number of a research suggest that error correction helps L2 writing skill (Ashwell, 2000; Chandler, 2003; Ferris, 1997), on the other hand, there are also some findings in literature that put forward the neutral effect of grammar correction; For example, Kepner (1991) and Truscott (1996) assert that grammar correction in L2 writing classes ought to be abandoned, even potentially harmful for students' writing ability. Truscott (1996) advocates, "(…) grammar correction has no place in writing courses… The reasons are: (a) Research data handed show that grammar correction is ineffective; (b) this lack of effectiveness is exactly what should be expected, given the nature of learning; (c) grammar correction has significant harmful effects; and (d) the various arguments offered for continuing it all lack merit" (p.328). Truscott (1999) also highlights that "By using constructive error correction, teachers encourage students to believe in it; because students believe in it, teachers must continue using it" (p.116).

However, as aforementioned, there is some opposed to these Truscott's claims that come from Ferris (1999), Ferris and Hedgcock (1998) and Ellis (2001).

> "Ferris (1999), Ferris and Hedgcock (1998) and Ellis (2001) criticized Truscott for being premature in his thesis and conclusions. First, Ferris claims that Truscott's use of the term 'error correction' lacks definition. Ferris states that there is sufficient research showing that effective error correction does improve students' writing. Secondly, Ferris criticizes Truscott's generalizations based on reviews of previous studies that

exhibited crucial variation with regard to groups of subjects, research and instructional paradigm and type of feedback. These variations are impediment in making generalizations" (cited in Delgado, Rocio; 2007: p.7).

The ongoing disagreement of the effects of direct versus indirect feedback types has led to other studies for grammar correction for the last decade. Even some radical disagreement on the positive effect of direct or indirect feedback types can be traced in some literature (e.g. Ferris & Hedgcock, 1998; Truscott, 1996; 1999; 2004; 2007; 2009; Ellis, 2001) Although little research argue about the effect of a certain type of error feedback, most state that working on grammar corrections or error correction in writing instruction is essential (Ferris, 1999; 2002; 2003). Additionally, there is some important emphasis on testing and assessing English. Madsen (1983) advocates that well made tests contribute to the positive attitudes and motivations of students and also efficient instruction. Besides, Madsen emphasizes that "(…) Where several tests are given, learning can also be enhanced by students' growing awareness of your objectives and the areas of emphasis in the course. Tests can foster learning too, by their diagnostic characteristics: They confirm what each person has mastered, and they point up those language items that need further attention.

If certain elements of a second language differed greatly from the student's native language, that student would likely encounter difficulties (Lado, 1957). It means sometimes your mother tongue interferes.

*Mary is interested to reading books.

* Mary is interested in reading books.

Naturally, a better awareness of course objectives and personal language needs can help your students adjust their personal goals…" (p. 4). Thus, we prefer to look for a different dimension of composition classes. An outlook for the effects of the numbers of exams and correlations between the exams and grammatical errors may be reasonable when Truscott's eye-catching emphasis is taken into consideration.

Therefore we raised the following research questions in this study:

1. Is there any negative or positive effect between the number of exams and errors/ mistakes in compositions?

2. If there is some effect of the number of exams, in what way or how does it deviate the number of errors whether positive, negative or neutral?

3. Number exams, the restriction of errors: Are they complementary because of domain effect?

Ali Osman ENGİN
Turgay HAN
Assiye BURGUCU

1. Methodology

1.1. Participants of the Study

The study is conducted on two groups of participants (control and experimental), consisting of 20 students each, attending preparatory day and evening classes in the Department of Tourism Guidance at Kafkas University. All participants were selected through convenience sampling strategy. The students enrolled in this program during the 2009-2010 academic years, took a proficiency test in the beginning of the academic year. Students who pass the exam were exempted from prep-class; so, they directly attended to the first-year classes. Besides, their English proficiency levels are thought to be similar when their University Placement Exam (YDS) results are taken into consideration. Hence, the participants of this study can be counted as homogenous.

The control group participants did not get any type of feedback (indirect, direct, conferencing, etc.). They only took three writing exams at four-week intervals during the first semester while experimental group participants took six writing exams at two-week intervals, they did not get any feedback as well.

Throughout the academic term, the lessons were conducted for fourteen weeks and the students were taught basic structure of writing skills such as how to write a paragraph, essay, decide a topic, create an outline, and write the body paragraphs etc., they learned writing content standards and also grammatical structures (*see Appendix B*).

2.2. Design of the study

The study is planned as an experimental study. First, each group was required to write a composition in English, and the papers were assessed by two independent and voluntary raters. Only grammatical errors were defined and classified, and each student's errors were recorded to an error chart.

To calculate inter-rater reliability ten exam papers were randomly marked based on five error categories by two raters. Using the SPSS Program (Version 11.0), it was found that there was no significant difference between raters.

2.3. Data Collection

All participants attended four-hour writing classes weekly during the term. During one of the two-hour session, it was aimed to help student develop their writing skills; therefore, the teacher introduced language structures, creative expressions, content and other composing elements through some reading activities. During the other two-hour session, students were required to compose on a given topic following a method.

The data was obtained weekly from the two raters' ratings for the writings by the both group participants. The writing teachers did not respond students' grammatical errors and even did not directly or indirectly give feedback to them. Nevertheless, the experimental group participants took three more exams in addition to the three administrative exams.

The two independent and volunteer raters only counted and classified the errors and then recorded to the chart (*see Appendix A).*

2.4. Procedure for Analyzing the Data

In the literature, analyzing errors with larger or smaller categories varies (Fratzen, 1995; Lane & Lange, 1999; Raimes, 1992; Ferris & Roberts, 2001). In this study, while identifying EFL students' grammatical errors, the raters used smaller categories of errors rather than larger categories (Ferris, 2002) because "students can focus on more a limited range of forms and rules when learning about a specific error type…" (Ferris, 2002; p. 68) further, "the use of 15 to 20 different terms or symbols to label errors may be overwhelming to teachers and students alike. Also the distinctions among error categories are not always as precise as we may think" (Ferris, 2002; p.68).

For each writing task by each participant, raters used an error-analysis form including 5 error categories (e.g. verb errors, noun ending errors, article errors, word choice errors, and sentence structure errors). In addition, total number of the errors, error ratio and total words were calculated and recorded to the form.

Three administrative writing exams were given to the control group participants during the twelve-week of writing courses; on the other hand 3 administrative writing exams and 3 more exams were given for the experimental group. Totally, raters marked 60 exam papers by the control group participants and 120 exam papers by experimental group participants.

All students wrote nearly 200-hundred word composition in each exam. They were not allowed to use dictionary, their text books and notes. However, it was difficult to assume that all the students would compose exactly 200-word writings; therefore, calculations were made to normalize the writings over 200 words.

Two calculations were made to normalize comparisons using the following formula:

1. [Number of all errors x 200] ÷ Number of Words

2. [Number of errors of a category marked by a rater x 200] ÷ Total Number of Words

The following table shows the non-parametric test results made through Wilcoxon Signed Ranks Test. There is no significant difference between the groups for overall markings (p= 0.589).

Table1: The comparison of pre-test scores between experimental and control groups

Groups	N	Mean	SD	Z	p
Experimental group	10	29.80	3.71	-0.541	0.589
Control group	10	30.00	3.20		

p>0.05

Table 2 shows the results obtained through the independent samples t-test analyses for each category. There is no significant difference between groups in terms of error category.

Table 2: The comparison of pre-test scores based on error categories between experimental and control groups

Error Category		N	Mean	Mean	Z	p
Noun Ending	Group 1	10	12.25	8.42	-1.501	0.133
	Group 2	10	8.75	6.74		
Verb	Group 1	10	8.55	1.38	-1.542	0.123
	Group 2	10	12.45	2.89		
Article	Group 1	10	10.85	1.38	-0.278	0.781
	Group 2	10	10.15	1.49		
Word Choice	Group 1	10	11.15	1.96	-0.511	0.610
	Group 2	10	9.85	0.85		
Sentence Structure	Group 1	10	8.85	0.81	-1.350	0.177
	Group 2	10	12.15	12.07		

Note: Control Group = *Group1* and Experimental Group = *Group2*.

The pre-test results of the Wilcoxon Nonparametric test shows that both raters marked errors of each category nearly same and there is no significant difference between ratings at p>0.05 level.

3. Results and Discussions

3.1. Analysis the data of Control Group Participants

Three administrative exam results were analyzed through independent samples t-tests. The number of the participants is below 30 and group variances are neglected. The tables below show the details of the analysis that were carried out in seven steps.

Comparing of mean scores between the first and second administrative exams, also second and third administrative exams of the control group students, Wilcoxon Signed Ranks Test results showed that there was a significant difference between exam1 and exam 2 (z=-3.375, p<0.001) and exam2 and exam3 (z=-3.616, p<0.000).

Table 3: The comparison between exam1-exam2 and exam2-exam3 of experimental
groups

Group1	EXAM2 - EXAM1	EXAM3 - EXAM2
Z	-3.375	-3.616
P<	0.001	0.000

Note: Control Group = *Group1* and Experimental Group = *Group2*.

Comparing the first and second administrative exams, also second and third
administrative exams of the experimental group students, Wilcoxon Signed Ranks Test
results showed that there was significant difference between exam1 and exam2 (z=-3.863,
p=0.000). However, there was no significant difference between exam2 and exam3 (z=-
2.621, p=0.009).

Table 4: The comparison between exam1-exam2 and exam2-exam3 of experimental
groups

Group2	EXAM2 - EXAM1	EXAM3 - EXAM2
Z	-3.863	-2.621
P<	0.000	0.009

Note: Control Group = *Group1* and Experimental Group = *Group2*.

Comparing three administrative exams of the control group and those of experimental
group, Wilcoxon Signed Ranks Test results showed that there was no significant difference
(p=0.226 at p>0.05 level)

Table 5: The comparison between mean scores of experimental and control groups

	The difference of experimental groups and control groups' average
Z	-1.211
P>	0.226

Note: Control Group = *Group1* and Experimental Group = *Group2*.

Three more exams were given to the experimental group students during the semester
and they were compared. Comparison between the fourth and fifth exams, also fifth and
sixth exams of the experimental group students were done with Wilcoxon Signed Ranks

Ali Osman ENGİN
Turgay HAN
Assiye BURGUCU

Test, and the results showed that there was a significant difference between exam4 and exam5, also between exam5 and exam6 (p=0.000 at p<0.05 level).

Table 6: The comparison between exam4-exam5 and exam5-exam6 of experimental groups

Group2	Exam5 - Exam4	Exam6 - Exam5
Z	-3.901	-4.013
P<	0.000	0.000

Note: Control Group = *Group1* and Experimental Group = *Group2*.

Descriptive results for the study show that there were some difference between the groups.

Table 7: Descriptive Statistics showing the distributions of mean scores of both groups.

Descriptive	N	Minimum	Maximum	Mean	Std. Deviation
Mean scores of 6 exams of group 2	20	19.17	31.67	25.0917	3.4727
Mean scores of 3 administrative exams of group 2	20	20.00	35.67	27.0500	4.4738
Mean scores of 3 administrative exams of group 1	20	19.00	37.00	27.4667	5.2864
4th 5th 6th exams of group 2	20	18.33	27.67	23.1333	2.5071

Note: Control Group = *Group1* and Experimental Group = *Group2*.

4. Conclusion

For the last two decades, debates have continued over the impact of different types of feedback (e.g. direct or indirect, coded or uncoded, teacher conferencing, etc.) on decreasing grammatical errors in ESL/EFL writing and speaking. Some concluded that giving feedback to the students' errors has a positive effect while some others argued that feedback must be abandoned as it has neutral effect, and even it is sometimes harmful in motivation context (e.g. Ferris, 1999; Truscott, 1996). However, to our best knowledge, there is not much research on the impact of the number of tests on ESL/EFL wiritng errors. This study was aimed to assess the effect of exams on the improvement. Students' errors in their compositions were not treated in terms of corrective feedback; no grammar correction feedback was given to the students' errors. Considering the first and second

research questions of our study, it was found that there is a negative correlation between the number of exam and number of grammatical errors. The number of errors on writing skill slightly decreased after each exam, this may constitute the positive effects of number of exams.

However, there are few further questions for the future research: if increasing number of exams works well in terms of decreasing number of exams, how many exams must be given to the students? In assessing process of writings of EFL students, what are the other factors that affect raters? And what should be the time span among the exams?

APPENDICES

Appendix A: Error-Analysis Form

Error Type	Number of Errors	Error Ratio
Verb Errors		
Noun Ending errors		
Article Errors		
Word Choice Errors		
Sentence Structure (Speech/Writing) Errors		
Total Errors Marked		
Total Words Marked		

Source: Adapted from Ferris, Dana R. (2002). Treatment of Error in Second Language Student Writing, USA: University of Michigan Press, p.134.

Appendix B: Content Standards

Content Standard 1—Students write clearly and effectively.
Content Standard 2—Students apply a range of skills and strategies in the writing process.
Content Standard 3—Students evaluate and reflect on their growth as writers.
Content Standard 4—Students write for a variety of purposes and audiences.
Content Standard 5—Students recognize the structures of various forms and apply these characteristics to their own writing.
Content Standard 6—Students use the inquiry process, problem-solving strategies, and resources to synthesize and communicate information.

Source: Adopted from Finn, C.E., Julian L., Petrilli, M.J., (2006). Montana Standards For Writing, October 1999. The State of State Standards. Thomas B. Fordham Foundation, p. 84.

Ali Osman ENGİN
Turgay HAN
Assiye BURGUCU

References

Ashwell, T. (2000). "Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method?". Journal of Second Language Writing, 9(3), 227-258.

Chandler, J. (2003). "The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2student writing". Journal of Second Language Writing, 12(3), 267-296.

Delgado, R. (2007). Effects of Different Error Feedback: Approaches in Students' Ability to Self-edit Their Writing. Divergencias. Revista de Estudios Lingüisticos y Literarios, 5 (2), 7.

Efe, H. (2008). "Punctuation Mistakes Committed by Lycée Graduate Turkish Students". Atatürk Üniversitesi Edebiyat Fakültesi Sosyal Bilimler Dergisi, Cilt : 8, Sayı: 41, Erzurum

Ellis, R. (2001). "Investigation form-focused instruction". In E.Rod (Ed.), Forma-Focused Instruction in Second Language Learning, (pp. 1-46). Malden, MA: Blackwell Publishers.

Ferris, D. R. (1997). The influence of commentary on student revision. TESOL Quarterly, 31(2), 315-339.

Ferris D. R. (1999). "The case for grammar correction in L2 writing classes: A response to Truscott (1996)". Journal of Second Language Writing, 8 (1), 1-11.

Ferris, D.R, Chaney, S.J., Komura, K., Roberst, B.J., & McKee, S. (2000). Perspectives, Problems, and Practices in Treating Written Error. Colloquium Presented at Intenational TESOL Convention, March 14-18, Vancouver, B.C.

Ferris, D. & Hedgcock, J.S. (1998). Teaching ESL Composition: Purpose, Process, and Practice. Mahwah, NJ: Lawrence Erlbaum Associates.

Ferris, D. R. & Roberts, B. (2001). Error feedback in L2 writing classes: Howexplicit does it need to be? Journal of Second Language Writing, 10(3),161-184

Ferris, D. R. (2002). Treatment of Error in Second Language Student Writing, USA: University of Michigan Press.

Ferris, D. R. (2003). Response to Student Writing: Implications for Second Language Students. Mahwah, NJ: Lawrence Erlbaum.

Fratzen, D. (1995). "The effect of grammar supplementation on written accuracy in an intermediate Spanish content course". Modern Language Journal, 79(3), 329-344.

Hedgcock, J. & Lefkowitz, N. (1994). "Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing". Journal of second language writing, 3, 141-163.

Kepner, C. G. (1991). "An Experiment in the relationship of types of written feedback to the development of second-language writing skills". The Modern Language Journal,75(3), 305-313.

James, C. (1980). Contrastive Analysis. London: Longman.

Lado, R. (1957). Linguistics across cultures: Applied linguistics for language teachers. University of Michigan Press: Ann Arbor.

Lane, J., & Lange, E. (1999). Writing Clearly: An Editing Guide, 2nd Edition. Boston: Heinle and Heinle.

Liu, Y. (2008). The effects of error feedback in second language writing. Arizona Working Papers in SLA & Teaching, 15, 65-79.

Madsen, H. S. (1983). Techniques in Testing. New York: Oxford University Press.

Raimes, A. (1992). Grammar Troublespots. New York: St. Martin's Press

Truscott, J. (1996). "The case against grammar correction in L2 writing classes". Language Learning, 46(2), 327-369.

Truscott, J. (1999). "The Case for the case against grammar correction in L2 writing Classes". A Response to Ferris. Journal of Second Language Writing, 8 (2), 111-122.

Truscott, J. (2004) Evidence and conjecture on the effects of correction: A response to Chandler. Journal of Second Language Writing, 13, 337 – 343.

Truscott, J. (2007). "The effect of error correction on learners' ability to write accurately". Journal of Second Language Writing 16(4), 255-272.

Truscott, J. (2009). "Arguments and appearances: A response to Chandler". Journal of Second Language Writing 19(1), 59-60.

Van Beuningen, C.G., De Jong, N.H., & Kuiken, F. (2008a). "The effect of direct and indirect corrective feedback on L2 learners' written accuracy". ITL International Journal of Applied Linguistics 156 , 279-296.

White, E. M. (1985). Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance. San Francisco: Josey-Bass.