

# A Study on the Effects of Feedback Given to Students by Computer to Increase Their Performance in English Writing

Xiaoxia Lu

School of Foreign Studies, Xi'an University, Xi'an, China

1260728600@qq.com

**Keywords:** Computer; Feedback; English writing; Performance; Effect

**Abstract.** Computer evaluation has widely been used to automate grading of students papers, especially in English writing in college and universities in China, and computer can also provide feedback to the students. This study is to examine whether or not feedback given by computer to the students has a positive impact on the following essay in terms of fluency, accuracy and complexity. Two students of English majors were chosen as the subjects, and they submitted their essays on the line once every two weeks. Their four essays were selected and studied and analyzed. After the essay analysis, the author finds that both before feedback and after feedback, reveal that these two students improved greatly in accuracy, though not much progress either in fluency or complexity. So it gets the conclusion of the study that computer evaluation really has positive effects on students' English writing and the students could make progress if they take computer feedback seriously.

## Introduction

Whether or not feedback, from both the instructor and from the computer, has a beneficial effect on EFL students, is controversial and has a long history of discussion and speculation. Computer evaluation has widely been used to automate grading of student papers, and computers can also provide feedback to the student. At the same time this computer automation liberates instructors from the laborious work of dealing with large numbers of student papers. Assessment of writing skills is time consuming, and nowadays some universities in China resort to the use of computers to assist in most of the grading. However, as to the validity of this grading, there is controversy.

## Literature Review

Selected below are representative samples of the kind of feedback and perceived results obtained by the various researchers. Weigle [1] presented critical considerations in which part of the discussion centred around the e-rater scoring engine used in the TOEFL IBT, contrasted with human raters. McCurry's trial experiment [2] reveals that automated essay scoring was not as reliable as human scoring in general written assignments. The comparative study of Wang and Wang [3], found the continuous and instant feedback from computers did assist students in being aware of where they need to make improvement. It is suggested by Parr and Timperley [4], that feedback does have a positive effect on student essays where the feedback compared good quality writing with that of the student's, and showed the student what good quality writing consists of and what is needed to bridge the gap between from the current quality of work to the desired performance.

Dikli and Bleyle[5], compared computer-feedback with instructor-feedback that college ESL students received on three written assignments and found that there was a large gap between the two types, and the results indicated that the feedback from the instructor was better. Stern and Solomon [6], mentioned that students were often given feedback and technical corrections such as spelling, word choice, missing words and grammar, while feedback on content and organization was absent which might prevent the students from improving essay qualities. Ware [7], stated that computers can only be used to assist in writing instead of producing quality assessment since feedback from a computer is very narrow. Hyland[8], found that the students did respond to teacher-feedback on a standard form.

However, some researchers hold negative attitudes and doubt that such form-focused feedback is able to develop students writing capability. Wingate's [9], findings suggest that students who didn't use computer feedback didn't improve in areas where they were weak while students who had paid careful attention to the feedback did improve.

Stevenson and Phakiti [10], critically reviewed the effects of computer-generated feedback and their evidence suggested a modest positive effect on student text, and as a result, there was minimal indication of feedback increasing general writing proficiency. The investigations done by Storch [11], Knoch, Rouhshad, & Storch [12], Knoch, Rouhshad, Oon & Storch [13], on the writing development of some undergraduate ESL students in English-medium University for one semester, over one year and three years periods, each found consistent results: that students only achieved fluency and had no gain in accuracy and complexity. The follow-up qualitative research revealed that feedback hadn't done much to improve writing quality related to their majors, because the feedback they could get was related to content only. Stevenson and Larsen-Freeman [14], mentioned that any linkage between complexity and feedback in language development, though theoretically independent beneficial, was actually dependent on circumstance. Wingate [15], concluded that error-correction had no benefit in student writing accuracy. Chandler [16], found that direct feedback contributes to accuracy.

## Methodology

**Participants.** The two subjects whose essays were chosen for this study are English majors in their second year at Xi'an University, China. Both are females, supervised by the same professor in their English writing class, though they are not in the same class. And these two students are average performers among the students enrolled in the class. However, their ability to understand is above average.

**Methods.** The study attempted to ascertain whether or not the feedback and comments given to the students in the first assignment, essay one, facilitated an improvement in the second essay. These improvements could be detected by observing the students' performance in the follow-up assignment, essay two, with regard to fluency, accuracy and complexity etc. Complexity-Accuracy- and- Fluency (CAF) methods are used to investigate a few areas such as the length-of-clauses, length of T-units, word count, clauses, dependent clauses and errors, just to name a few. After investigating the difference between the two results, the author attempted to understand why each student performed better in one area, and less so in some other areas. The author also looked at whether or not the feedback from the computer was consistent with the non-computer analysis. In addition, the author did a holistic analysis of the difference between the results of each student's first and second essays, then she also made some suggestions for future studies.

**Data Collection.** The data for this article was obtained from a computer system which automatically analyzed essays submitted from two separate students, each of whom wrote an essay on the same two subject matters; "We should pursue long term goals" and "Will electronic dictionaries kill printed dictionaries?" This work was undertaken at Xi'an University, China. It consists of four short essays, each approximately 200 words in length. The students submitted their finished essays via the university's online computer system. Remarkably, modern day computers are able to analyze these essays and automatically generate helpful feedback to the students. Feedback to students was also given by the teacher both jointly, in class, and separately during private teacher-student discussions. The students were given two weeks to complete each essay.

The essay assignments were given to each student so that they may practice the writing skills in English, hopefully this will better prepare them for taking TEM-4, a test for English majors. The written portion of the test requires each student to complete an essay of between 200 and 250 words, note that the two assignments were chosen to be of the same length. Scoring for the written portion of Band Four is a number between 1 and 15 with 15 being the best score.

The students submit their essays online with a two week deadline. After submission, the students still have an opportunity to make changes to their essays as many times as they wish before the due date. With each new submission, the computer analyses the content and provides a new score, that new score

could be either higher or lower. The feedback is generated by the computer and has three parts. Part 1 comments on some key aspects of sentence structure: verb-subject agreement, appropriate use of words and grammar etc. Part 2 quantifies the student's vocabulary, paragraph count, sentences count, subordinate clause usage, vocabulary in general, percentage of academic words, percentage of advanced words, average length of words, vocabulary beyond the syllabus, and total number of the words, just to name a few. This feedback can be by number or by percentage. And finally Part 3 consists of a mark and short general comment pointing out both the strengths and weaknesses that the student should consider in future writing. The marks were awarded based on the quality of content, organization, grammar and appropriateness on a 1-15 point scale.

**Results.** From Table 1 and Table 2, it can be seen that as far as fluency is concerned, the second essay didn't show any big difference; roughly both stayed at the same level. Compared with the first essay, obvious the second essay doesn't make any difference as far as fluency goes.

Table 1 on fluency of script 1 and script 2 (student A)

Items	Script 1	Script2
T-units	14	14
Clauses	22	23
Measures:		
W/T	217/14=15.5 (each T-unit is approx. 16 words in length)	205/14=14.6 (each T-unit is approx. 15 words in length)
W/C	217/22=9.9(each clause is approx. 10 words in length)	205/23=8.9(each clause is approx. 9 words in length)
C/T	22/14=1.57 ( each T-unit is made up of roughly 2 clauses)	23/14=1.64 ( each T-units is made up of roughly 2 clauses)

Table 2 On fluency of script 1 and script 2 (student B)

Items	Script 1	Script2
T-unit	16	14
clauses	24	23
Measures:		
W/T	216/16=13.5 (each T-unit is approx. 14 words in length)	208/14=14.85 (each T-unit is approx. 15 words in length)
W/C	216/24=9 (each clause is approx. 9 words in length)	208/23=9.04 (each clause is approx. 9 words in length)
C/T	24/16=1.5 (each T-unit is made up of roughly 2 clauses)	23/14=1.64 (each T-unit is made up of roughly 2 clause)

From Table 3 we can see the big difference that student A made in her second essay, the errors dropped by 4%, error free units increased by 36% and error free clauses went up by 24 %.

Comparing the two scripts in Table 4, we can see that errors dropped by 4% in the second script. There is a 16 percent of increase of error free T-units in the second script, at the same time error free clauses increased by 16 % as well.

Table 5 tells that in the second essay among the 4 areas measured the only improvement lies in C/T, but only on very tiny amount.

Table 6 shows that the second essay of student B has a tiny increase in each of the 4 areas measured, though not much yet there is a sign of improvement anyway.

Table 3 On accuracy of script 1 and script 2 (student A)

Items	Script 1	Script2
No. Errors	18	8
No. Error free T-units(EFT)	7	12
No. Error free clauses (EFC)	15	21
Measures:		
E/W	19/217=0.08	8/205=0.04
EFT/T	7/14=0.5 (50% of T-units are error free)	12/14=0.86 ( 86% of T-units are error free)

Table 4 On accuracy of script 1 and script 2 (student B)

Items	Script 1	Script2
No. Errors	14	4
No. Errors free T-units (EFT)	10	11
No. Errors free clauses (EFC)	17	20
Measures:		
E/W	14/216=0.06	4/208=0.02
EFT/T	10/16=0.63 (63% of T-units are error free)	11/14=0.79 ( 79% of T-units are error free)

Table 5 on complexity of script 1 and script 2 (student A)

Items	Script 1	Script2
Dependent clauses (DC)	10	9
Clauses ( C )	22	23
Measures:		
C/T	22/14=1.57 (each T-unit is made up of roughly 2 clauses)	23/14=1.64 (each T-units is made up of roughly 2 clauses)
DC/C	10/22=0.45	9/23=0.39
W/C	217/22=9.86 (each clause is approx. 10 words in length)	205/23=8.9 (each clause is approx. 9 words in length)

Table 6 On complexity of script 1 and script 2 (student B)

Items	Script 1	Script2
Dependent Clauses (DC)	8	8
Clauses (c )	24	23
Measures:		
C/T	24/16=1.5 (each T-unit is made up of roughly 2 clauses)	23/14=1.64 (each T-unit is made up of roughly 2 clauses)
DC/C	8/24=0.33	8/23=0.35
W/C	216/24=9 (each clause is approx.. 9 words in length)	208/23=9.04 ( each clause is approx. 9 words in length)

## Analysis

The purpose of the study is to examine whether or not feedback given to the students after the first essay had a positive impact on the second essay in terms of fluency, accuracy and complexity. To accomplish this, CAF methods are employed and some rather interesting results are found. A careful review of those aspects (see Table. 1 -2) shows no obvious improvements in fluency. Fluency in language is the quality of being able to speak, and/or read-and-write a given language without the need to consciously think much about it - in other words the language comes as naturally as most people find their native tongue. In student A's case, the approximate word count in both T-units and clauses dropped by one word each in the second essay. In fact, the second essay was 12 words shorter. In student B's second essay, the approximate word count in each T-unit decreased by one word. But, the word count in each clause was seen to be more or less the same, although the total word count of the second essay was seen to be 6 words less than the first essay. Here, the author is just reporting the results, and she is perfectly aware that, statistically speaking, one should not draw conclusions from such a small sampling. It seems that fluency is hard to achieve in a short period of time. Suppose both students had increased their word counts from the first-to-second essay. Should we use this as an indication that the student has acquired increased fluency? In a large sampling, the answer is "yes". However, this needs to be further investigated. Regardless of how slight, it seems obvious that this simple experiment did not show any deterioration in student performance.

By contrast, the results of the accuracy-measurements (see Table.3-4) are strikingly different between the two assignments - accuracy in the second essay was seen to display a remarkable increase. And this demonstrates that feedback could increase accuracy. Fewer errors and overall writing skills were apparent in the second essays. Feedback helps to increase the accuracy in both clauses and T-units. Student A and student B both made great progress in keeping most T-units and clause error free. Specifically, a 36% improvement in T-units was achieved by student A, and a 16% improvement was made by student B. Accordingly, student A increased error free clause by 24 % and student B showed a 16% increase. In addition, both students had a 4% reduction in errors in the second essay, with Student A having decreased the errors from 18 in the first essay to 8 in the second essay and student B having reduced the errors from 14 in the first essay to 4 in the second essay. Presumably, the reason for this increase is because feedback was given and focused on the appropriate uses of words, collocations, clauses and subordinates. Though the algorithm used by the computer in my own feedback analysis differs in method, the results are found to be consistent and these results lend themselves nicely to demonstrate that feedback does, in fact, contribute to the improved accuracy.

Finally, a close examination of the complexity measurements (see Table 5-6) in their second essays demonstrated that the second two essays were, in fact, no better than the first two, which might imply that complexity is harder to achieve in a short period of time. Complexity involves more than accuracy. It requires an advanced knowledge of grammar, and understanding of sentence structure, subordinates, compounds, complex and extensive reading in the language of interest. What I've seen is that a student's confidence in using complicated structures can be gradually cultivated. Too much concern over accuracy may restrain a student from making an effort to develop complicated sentences. The more complicated a sentence is, the more time it takes to write it. And, more time spent writing means more time for errors to be made. Surely, a high proficiency in writing is characterized by fluidity and complexity. And to achieve this high level of proficiency a student should receive more, and better, guidance and encouragement as well as undertake extensive practice.

Overall, both students were good at spelling, flexible in use of vocabulary, strong at syntax. Their sentence structures, relatively speaking, were well organized; they both used transitional words well to make paragraphs smooth and natural.

## Conclusion

The findings from the essay analysis, both before feedback and after feedback, reveal that these two students improved greatly in accuracy, which is notable, considering only two weeks was between the



first assignment and the second one. After completing the first assignment, the feedback pointed out difficulties in grammar, and the students seemed to comprehend the feedback well, and apply their newly acquired knowledge to the second essay. After receiving feedback they spent time focusing on word usage, tenses, subject-verb agreement, and so forth. They also chose better connective words to make the paragraphs read more smoothly and seem more natural. These findings do agree with Wingate [9], and Chandler[16], on the positive effects of feedback in leading to the accuracy of student essays. My findings are also consistent with the studies of Storch (2009), and Knoch et al (2014, 2015).

However, the two students in the study didn't make much progress either in fluency or complexity. The cause of this is unclear, and should require further investigation. At this stage, it can only be surmised that fluency could possibly be connected with higher word counts. Students that use more words are often the ones producing more fluent writings. However, students may be concerned that the more they write the more mistakes they make. Students didn't want to take the risk of receiving a poor grade in a failed attempt at writing more complex sentences. Or they might value accuracy more than fluency. Likewise, in the studies of Storch [11], Knoch et al [12] [13], they found fluency in ESL student essays, which indicates that when students are free from worries about grammar, their fluency would increase. Time seemed not to be a factor for improvement in fluency. In the case of my study, two weeks seemed plenty long enough for producing a quality short essay. In this study, students who fail to achieve fluency and complexity can be understood as those not willing to expend the necessary time and also as those who do not have a full and good familiarity of grammar and lexical resource; knowledge which must be accumulated via extensive reading and reflections. Feedback can help achieve better complexity and fluency as Pallotti[17], postulated. It should be demonstrated, and not assumed that complexity is achievable with proper feedback over time. Further studies should be undertaken to determine the best way to structure feedback and how to arrange for quantitative measure to justify the results. It is also important that controlled, impromptu essays be assigned for this research, since impromptu timed essays force the students to think quickly and to articulate their thoughts spontaneously. Future research could reveal yet to be discovered methods for formulating feedback that could be of real help in training future students and allowing them to progress from their academic life into young professionals contributing meaningful work for the betterment of themselves and their future employer.

## **Acknowledgements**

The thesis is a phased achievement of the reform and the practice of the course English Writing based on the environment of net work study in Xi'an University Fund Project for curriculum reform (Program No.: KC2015B07)

## **References**

- [1] Weigle, S. C. English language learners and automated scoring of essays: critical considerations *Assessing Writing*, (2013) No.18, p. 85–99.
- [2] McCurry, D. Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, (2010) No.15, p. 118–129.
- [3] Wang, F. W., & Wang, Sh. W. A Comparative Study on the Influence of Automated Evaluation System and Teacher Grading on Students' English Writing *Procedia Engineering*, (2012) No.29, p. 993 – 997.
- [4] Parr, J. M., & Timperley, H.S. Feedback to writing, assessment for teaching and learning and student progress . *Assessing Writing*, (2010) No.15, p. 68–85.
- [5] Dikli, S., & Bleyle, S. Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, (2014) No.22, P. 1–17.
- [6] Stern, L. A. & Solomon, A. Effective faculty feedback: The road less traveled. *Assessing Writing*, (2006) No.11, p. 22–41.

- [7] Ware, P. Computer generated feedback on students writing. *TESOL Quarterly*, (2011) Vol. 45 No.4, p. 769-774 doi:10.5054/tq.210.12.7252
- [8] Hyland, F. Focusing on form: student engagement with teacher feedback. *System*, (2003) No.31, p. 217–230.
- [9] Wingate, U. The impact of formative feedback on the development of academic. *Assessment & Evaluation in Higher Education*, Vol.35 (2010) No.5, p. 519–533.
- [10] Stevenson, M., & Phakit, K. The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, (2014) No.19, p.51–65.
- [11] Storch, N. The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, (2009) No.18, p. 103–118.
- [12] Knoch, U., Rouhshad, A., & Storch, N. Does the writing of undergraduate ESL students develop after one year of study in an English-medium university? *Assessing Writing*, (2014) No.21, p.1–17
- [13] Knoch, U., Rouhshad, A., Oon, S. P., & Storch N. What happens to ESL students' writing after three years of study at an English- medium university? *Journal of Second Language Writing*, (2015) No.28, p. 39–52
- [14] Larsen-Freeman, D. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, Vol.27 (2006) No.4, p.590-619 doi:10.1093/applnl/aml029
- [15] Wingate, J. The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, ( 2007) No.16 p. 255–272.
- [16] Chandler, J. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, (2003), No.12, p. 267–296.
- [17] Pallotti, G. CAF: Defining, Refining and Differentiating Constructs *Applied Linguistics*, 2009, (30/4): 590–601 doi:10.1093/applin/aml045 Weigle, S. C. English language learners and automated scoring of essays: critical considerations. *Assessing Writing*, (2013) No.18, p. 85–99.