بودن یا نبودن: تلاشی برای یافتن معیارهای عینی جهت ارزشیابی مهارت نگارش

مجید نعمتی
استادیار آموزش زبان انگلیسی دانشگاه تهران

مهارت نگارش از جمله مهارت‌هایی است که در محیط آکادمیک از اهمیت فوق‌العاده‌ای برخوردار است. با این وجود ارزشیابی این مهارت تا حد زیادی سلیقه‌ای می‌باشد و بنابراین یافتن راهی برای ارزشیابی دقیق و عینی نگارش ضروری به نظر می‌رسد. تحقیق حاضر تلاشی است برای یافتن ارتباط لغات مورد استفاده در نگارش یک متن و مهارت نگارش نویسندهٔ آن

متن. در این راستا رابطهٔ سه معیار فعل بودن، کلمات آسان پر استعمال و کلمات یک سیلابی با کیفیت نگارشی متنی که در آن استفاده شدند مورد بررسی قرار گرفتند. از این سه متغیر برای فعل بودن و کلمات یک سیلابی ضریب همبستگی منفی معناداری مشاهده شد ولی برای لیست کلمات آسان علیرغم وجود یک ضریب همبستگی منفی این ارتباط از نظر آماری معنادار نبود.

واژه‌های کلیدی: نگارش، ارزشیابی، انتخاب لغت، فعل بودن.

Pazhuhesh-e Zabanha-ye Khareji, No. 32, Special Issue, English, 2007, pp 175-186

'To be or not to be': A search for new objective criteria to evaluate EFL compositions

Majid Nemati
Assistant Professor, Department of English, Faculty of Foreign Languages, University of Tehran
e-mail:nematim@ut.ac.ir

Abstract

Writing, in the academic realm, is ubiquitous and thus of paramount importance. Nevertheless, assessing writing suffers from lack of objectivity which warrants further studies to find some more objective way for evaluating this skill. The present study is an attempt to investigate the probable links between the words used in a text and the mastery of its writer. Three criteria i.e. the verb *to be*, *monosyllabic words* and *easy words* were examined in this regard. Except for the latter for which just a trend was observed, the other two criteria showed statistically significant negative correlation coefficients with the quality of texts.

Keywords: writing, assessment, word selection, to be.

1. Introduction

Writing, compared to other skills and components of language, has always suffered a lot for its alleged lack of an established reliable and valid assessment method. This Achilles' heel has crippled both teachers and researchers working on writing ability and, as some scholars (e.g. Hirche, 1977) put it, it is one of the most important barriers which prevents practical progress in teaching and researching writing. After all, teachers and researchers have to evaluate the students' writing ability to realise any progress or weakness in a methodology, in teaching materials, or even in the test itself.

In the past three decades, evaluating writing skill has undergone various developments. In 1960s and 1970s the so-called objective tests which basically employed multiple-choice test items were considered as the only reliable way of evaluating writing (Dunlop, 1969). As well as theory, in practice major test batteries (e. g. TOEFL) preferred this type of test, too. The heyday of this quantifiable, and therefore reliable, method had been passed in the early 1980s (Oller, 1979; Jacobs *et al*, 1981; Carroll & Hall, 1985; Heaton, 1990). The introduction of the Test of Written English (TWE) by TOEFL in 1986 was a final seal to declare officially the end of the multiple-choice era.

The ever-lasting dichotomy of reliability versus validity drew attention towards holistically-scored compositions as the only valid approach towards evaluating writing skill. In the recent history of evaluating compositions, there has been a tendency to reconcile the two extremes by finding objective elements in the compositions to enable scorers to discriminate between well or poorly written essays. In this method, known as the frequency-count method, scorers 'tally or enumerate certain elements in the composition, such as: the number or type of words, clauses, T-units, cohesive devices, misspelled words, misplaced commas, or sentence errors' (Jacobs *et al* ,

1981, p.29).

Although Jacobs et al. (1981) admire the objectivity and high reliability of Frequency-Count method, they sharply criticise its dubious validity because in this method of assessment the communicative effect of the composition is replaced by the number or kind of elements. It should be pointed out that these methods, or different techniques of a method, are considered holistic method as they require the test takers to create a continuous written discourse(cf. objective multiple-choice tests), yet they are totally different from holistic scoring which rely on the subjective scores, mainly based on the rater's impression of the composition.

2. Objective measures of writing

Labrant(1933) can be considered the pioneer of this approach. She focused on some objective measures such as sentence length, clause length, and subordination ratio. Among these the subordination ratio proved to be a significant indicator of the student's maturity in writing. Similarly, Hunt(1965) developed five objective measures to evaluate writing skill. Among these five factors T-unit (terminable unit) proved to be the best index of writing maturity. Other researchers (e.g. Crowhurst, 1980; Robb et al., 1986, Wilkinson, 1989), however, refuted Hunt's measures, arguing that syntactic complexity depends on the writing task. To turn a bad situation to worse, Larsen-Freeman (1978) reported that neither the average number of words per T-unit, the average number of T-unit per sentence, nor the average number of words per error-free T-units were significant.

Many research studies have been carried out having the concept of error as their focal point. Perkins (1980) asserts that objective measures without taking the role of errors into account 'are of no use in discriminating among holistic evaluation at one advanced level of proficiency' (p. 64). In this connection Arthur (1979) found that there is a relationship between the

teacher's ranking of the essay and the frequency of spelling and grammar errors. Supporting Hunt, two other investigators, Flahive and Snow (1980) developed two new measures: index of complexity, and the second one error per T-unit ratio which revealed that Hunt's measures were quite accurate. Larsen-Freeman (1978) found that the error-free T-unit scores significantly differentiate between ESL writing levels. The element of committed errors is so appealing that a number of scholars (e.g. Humborg, 1984; Oller 1979) each tried to formulate the role of the error element in evaluating writing. Madsen(1983) points to an objective technique as 'points off' method. In his terms, 'students begin with 100 points or an A grade. Then, they lose points or fractions of a grade for errors' (p. 120).

The relationship of cohesive devices and the maturity of writing skill is another widely-used measure of students' writing ability. Some teachers believe that superior writers can handle cohesive devices more appropriately and, therefore, achieve a more coherent text. The most outstanding experiment done in this area is that of Evola, Mamer, and Lantz (1980). They investigated the relationship between cohesive devices used in a piece of writing and the objective score as well as the subjective rating given to it. It was found that there is a significant but weak correlation between the correct usage of conjunctions, pronouns, and articles and the level of writing proficiency. They admit that cohesive devices are not highly reliable measures because they provide only 'minimal indicators of overall language proficiency...(and) cannot be expected to reflect ... communicative ability...(Evola, Mamer, & Lantz, 1980, p. 191). Farzanehnezhad(1991) in a more recent study propounded two new measures called Measure of Cohesion (MC) and Maturity Index (MI). The MC can be computed by adding the number of cohesive devices and dividing the sum by the number of the T-units used in a composition. The MI is calculated through the following formula: MI = mean T-unit length + clause/ T-unit ratio + MC.

Although the stream of research in this area seems to be endless, and it is therefore impossible to cite all related research projects, there is one major branch of investigation which should not be overlooked and that is the question of syntactic differences viewed from a contrastive-rhetoric point of view. Reid (1990), possibly, can be named as the most outstanding endeavour of this type of investigation, however, the findings contribute rather indirectly to the field of writing assessment by providing textual information about essays of different types and written by speakers of different languages. The main shortcoming of this research and similar projects is that they do not examine the qualitative differences between the essays and the reader's attitude towards these differences, which is central to writing evaluation.

There is no doubt that writing is a sophisticated phenomenon in which many factors are interactively incorporated. Despite this complexity, one should not be unmindful that words, as the building blocks of any written text, are of paramount importance and, as assumed in this research, can manifest the writer's writing proficiency. Focusing on the words selected by writers, this research was designed to suggest new objective measures to discriminate between the beginners and more advanced writers. In other words it is an attempt to answer the question of whether or not beginner EFL writers use a different array of vocabulary in their writings compared to the vocabulary used by more advanced writers.

3. Research hypotheses

The above question, in terms of research studies, was too broad to answer. So, it was broken up into three manageable null hypotheses as follows:

H_0 #1 - There is no significant relationship between the ratio of easy words, listed in *"Dale's list of 769 easy words"*, used in a composition and

the writing ability of the writer.

Although more up-to-date frequency lists, such as the Cobuild list, would have been better choices for this experiment, as they were not available to the researcher at that time, this list, which is widely used for research purposes in Iran, was chosen.

H_0 #2 - The writing ability of those subjects who mainly use auxiliary verbs in general and *TO BE* in particular and those who replace them with main verbs are not significantly different.

The rationale behind this assumption is personal experience, encountering the Iranian students' common mistakes in their writings.

H_0 #3 - There is no statistically significant relationship between the number of monosyllabic words employed in an essay and the writer's writing maturity.

The third hypothesis is loosely based on the criteria used for calculating text readability in order to see whether or not the readability of texts written by learners with different writing proficiency differ.

4. METHOD

4.1. Participants and the procedure

To test these hypotheses, a group of 50 students, both male and female, were randomly selected from the population of Iranian graduate students who intended to continue their postgraduate studies abroad. They were all non-English majors selected from different fields of study.

In order to receive a scholarship they had to take a language proficiency exam, named after the Iranian Ministry of Culture and Higher Education (MCHE), which develops and administers it. As a part of this exam, the subjects were required to write an essay of about 200 words within 30 minutes. The limits of this writing test were in accord with guidelines given

by Jacobs et al. (1981) where they write :

"... we have found with a 30-minute composition test that students at all but most basic level of proficiency can generally write about a page or more. Directions for a 30-minute composition task might indicate, then, that students should write "about one page" or "200-300 words." (P.19)

Two different topics were given which subjects had to select and write on one of them. These two topics were only slightly different so that a choice of topics was offered without resulting in totally different compositions that could make the rating procedure difficult and even inaccurate. The topics were about advantages and disadvantages of learning English in Iran / studying and living abroad which were quite familiar and appealing issues to the subjects.

4.2 Scoring procedure

The collected compositions were separately scored by two raters, one Iranian male teacher of English and one female native speaker of English language with considerable teaching experience. The raters were asked to mark papers holistically on a 0 - 20 point scale which is the common scoring system in Iran. The two grades were then averaged and the result was regarded as the final mark for each writer. Following is the basic descriptive statistics for the final scores of compositions:

Table 1: Descriptive Statistics for the Final Composition Scores

N	Mean	SD	Min.	Max.
50	13.05	2.26	6	17

It was startling to find out that the inter-rater reliability was 0.68 which

may seem at first glance as a moderate estimate. However, considering the facts that there had been no co-ordinating meetings between the two raters and that the raters came from two completely different backgrounds, its significance becomes more evident. Undoubtedly, by having the raters in even a very short co-ordinating session the obtained inter-rater reliability would soar dramatically, but it was not held in this experiment to avoid any change in the existing criteria in the mind of the raters.

5. Results

The first null hypothesis ($H_0$1) was concerned with the role of easy words in compositions and their effects on the writing ability of the EFL learners. Although a weak correlation coefficient estimate (r = -.10) was found between the two variables of composition grades and the ratio of the easy words to the total number of words, it was not significant enough to reject the first null hypothesis at the level of $P < 0.05$.

It can, however, be inferred that the Dale list is not a proper choice for Iranian students simply because it is not based on their needs. As it was mentioned before this list is not suitable for Iranian learners and there is a probable possibility that a similar research using a more appropriate list of easy words would lead into a significant result.

The second research question and its corresponding hypothesis focused on the role of *TO BE* as an auxiliary verb. A negative correlation coefficient of r = -0.26 ($P < 0.05$) was observed between the number of *TO BE* in essays and the essay scores. Hence, it can be concluded that weaker students tend to (over)use auxiliary BE, correctly or incorrectly, in their writings. As they improve in writing they avoid doing so by replacing these auxiliaries with main verbs and/or omitting them.

Hypothesis III was proposed to investigate the significance of monosyllabic words in EFL essay writing. The idea was inspired by the Fog

index, a formula widely used for calculating the readability of English texts. This formula takes three-syllable words into consideration. Alternatively, to facilitate the procedure, one can use monosyllabic words as criterion which has a correlation of -0.91 with the previous one Bowen, et al. 1985). A significant correlation coefficient of $r = -0.23$, ($P < 0.05$), between the composition grades and the ratio of monosyllabic words was found. The significance of the estimated correlation coefficient rejects the third null hypothesis, indicating that there is a mutually exclusive relationship between the writing proficiency and the use of monosyllabic words in the written text.

6. Peripheral findings

The research process created an excellent opportunity to study some other pertinent variables, though these are not directly related to the research questions.

Firstly, the ratio of verbs to total words in each essay, as an indicator of sentence length, was correlated with final scores which showed a significant negative correlation coefficient of $r = -0.45$ ($P < 0.05$). In a further step, according to the scores assigned by the raters, the subjects were divided into three subgroups: *GOOD*, *FAIR*, and *WEAK*. A one-way ANOVA proved that this ratio can be used as a criterion to discriminate the learners with different levels of writing proficiency into the three pigeonholes. It means that the ratio of verbs to words can be viewed as a means of evaluating EFL compositions. This relationship was first established in 1933 by Labrant and confirmed later on by Hunt (1965) but they had only examined writing in a first language.

Secondly, the total number of words written down, signifying the length of the essay, was taken into account. The significant correlation coefficient of $r = 0.52$ ($P < 0.05$) would mean that more advanced learners are more fluent in writing, too. Similarly, a one-way ANOVA confirmed that the total

number of words written in a timed evaluation session can reliably discriminate better writers from less proficient ones by dividing them into three subgroups labelled as *GOOD, FAIR*, and *WEAK*.

Thirdly, and finally, the composition scores were validated against the MCHE test scores with a calculated correlation coefficient of 0.71 which endorses essay tests as a relatively valid test of general proficiency test. Considering the fact that the writing task discussed in this article, for practical reasons, was loosely assigned and administered, it can be logically concluded that in a more controlled condition with trained and co-ordinated raters the results would improve dramatically.

7. Conclusion

In this research study a few criteria were investigated in order to check the possibility of using them as objective measures to assess EFL writing skill. It has been concluded that words used by foreign language learners in their compositions can be, to some extent, reliable indicators of writing maturity. However, it is wise to use these findings as pedagogical guidelines to help learners improve their writing skill rather than judging them merely according to these factors or even worse incorporating them into major official testing schemes as it may cause disastrous washback effects on classroom teaching procedures.

Refrences

Arthur, B. (1979) Short-term changes in EFL composition skills, *in* Yorio, C.A., Perkins, K. and Schachter, J. (eds) ***On TESOL***. TESOL. pp.330-342.

Bowen, J.D., Madsen, H. & Hilferty, A. (1985) TESOL, techniques and procedures. Cambridge: Newbury House.

Carroll B., & Hall P. (1985). ***Make your own language tests***. Oxford: Pergamon Press.

Crowhurst, M. (1980) Syntactic complexity and teachers' quality ratings of narrations and arguments. ***Research in the teaching of English***, 14 (3), pp.223-231.

Dunlop, I. (1969) Tests of writing ability in English as a foreign language. ***ELT Journal***.

Evola, J., Mamer and Lantz (1980) Discrete point versus global scoring for cohesive devices, in Oller, J. W. Jr. and Perkins, K. (eds), ***Research in language testing***. Mass: Newbury House Publishers Incorporated, pp. 177-183.

Farzanehnezhad, M. (1992) Cohesive devices and writing proficiency. M.A. thesis, Tehran University.

Flahive, D., and Snow, B. G. (1980). Measures of syntactic complexity in evaluating ESL composition. In J. Oller and K. Perkins (Eds.), ***Research in language testing*** (pp. 171-176). Rowley, Mass.: Newbury House.

Heaton, J. B. (1990) ***Classroom testing***. London: Longman.

Hirsch, E. (1977) ***The philosophy of composition***. Chicago, IL.: The University of Chicago.

Homburg, T.J. (1989) Holistic evaluation of ESL compositions. ***TESOL Quarterly***, 18, pp. 87-107.

Hunt, K. (1965) ***Grammatical structures written at three grade levels***. Urbana, Ill.: national Council of teachers of English.

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). ***Testing ESL composition: A practical approach***. Rowley, MA: Newbury House.

La Brant, L. L. (1933) A study of certain language developments of children in grades four to twelve, inclusive. ***Genetic Psychology Monographs***, XIV, pp. 387-491.

Larsen-Freeman, D. (1978). An ESL index of development. TESOL Quarterly 12, (pp. 439-448)

Madsen, H.S. (1983) ***Techniques in testing***. Oxford: Oxford University Press.

Oller, J. W. Jr. (1979) ***Language Tests at School: A Pragmatic Approach***. New York: Longman Group, Ltd.

Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. ***TESOL Quarterly***, 14, 61-70.

Reid, J. (1990). Responding to different topic types: A quantitative analysis. In B. Kroll (Ed.), ***Second language writing: Issues and options***. New York: Cambridge University Press.

Robb, T., Ross, S. & Shortreed, I. (1986). Salience of feedback on error and its effect on ESL writing quality. ***TESOL Quarterly*** 20, (pp. 83-93)

Wilkinson, A. (1989) Our first great conversationalists. ***English in Education***, 23 (2), pp. 12-24.