Analytic Assessment of TEFL Undergraduate Students' Writings: Diagnosing Areas of Strength and Weakness

Ebrahim Mohammadkarimi[1]*

Abstract

Assessment of learners affects all processes of their learning. Generally, without assessment, it is very difficult to evaluate the effectiveness of the teaching and learning process. Understanding their weak points and difficulties in their writing would help teachers and administrators focus more on these areas. Therefore, this content analysis study was an attempt to investigate the relative contribution of each writing component to the variation in the overall writing performance of TEFL undergraduate students to determine areas of strength and weakness in their writing. To this end, the writing samples of 73 students who were willing to participate in this study at different universities in Tabriz, Iran, were assessed through an Analytical Scoring Scale. These participants were divided into two groups: group 1, first-year and second-year students, and group 2, third-year and fourth-year students. The results of the Stepwise Multiple Linear Regression analysis revealed that the greatest contributions to the variance of the total rating of TEFL undergraduate students' writings were related to the grammar and punctuation components in groups 1 and 2, while the least contributions were related to the content and spelling in group 1, and vocabulary and spelling in group 2. In addition, the results indicated that the weaknesses and issues of participants in both groups of this research were mostly related to the grammatical structure and using punctuation marks in paragraphs. Moreover, the positive effects of the analytical scoring scale as well as the implications of the findings were discussed.

Keywords: content analysis; TEFL undergraduate students; analytical scoring scale; writing components

1. Introduction

In language learning, writing is considered an important skill for students. Unlike speaking, writing is a more scholastic aspect of language, one that needs more instruction for its development. In addition, its improvement requires a lot of effort from students (Richards, 2002). In this regard, according to McMaster et al. (2020), substantial changes have been made in the assessment and instruction of writing over the last decades. In former instructions of writing, mainly grammar drills had an essential role in improving students` compositions. Nevertheless, there has been a relative shift from this approach to using portfolios, projects, reflections, and journals (Dinçman, 2002).

[1] Ph.D. in TESOL, Language and Development Center, University of Raparin, Kurdistan, Iraq.
Email: Ebrahim.karimi@uor.edu.krd

To check the quality and effectiveness of a piece of writing, it should be assessed, and it is usually the role of the teacher to assess their students' writing. However, it has been a problematic task not only for teachers but also for the whole education system. Considering English as a foreign language (EFL), learners with different linguistic backgrounds make this problematic task more and more of a striking issue. Moreover, the linguistic background of teachers as well as their previous training and experience in writing assessment fundamentally affect the assessment of learners` writing. Assessment of writing depends on different factors, including the size of the class (small-scale or large-scale) and the kind of writing (simple essay, portfolio, long-project, and thesis).

In Iran, English is taught as a foreign language from secondary schools and most of the methods of teaching and, consequently, assessment are still traditional ones, so relatively less attention has been paid to reform in this regard. In most countries, writing skills are less practiced by students, and mainly in their writings, students have only one topic and option to write about (Vaughn & Bos, 2012), and this is the case in our country too. Iranian students, even in the field of English Language Teaching (ELT), do not have enough practice in essay writing and the teachers do not provide enough feedback on all aspects of their writing. Their assigned essays are usually assessed impressionistically and without any scoring scales. Therefore, students would not be aware of the weak points in different aspects of their writing.

Reviewing the literature reveals that many studies investigated the implementation of various sorts of scoring scales in order to assess writing in large contexts rather than in classrooms. Moreover, these studies mainly talked about the impacts of various scales of scoring on raters and ratings. For example, Knock, Read, and Randow (2007) and Wiseman (2012) inquired about rater training impacts, and many other researchers surveyed the possibility of raters' bias, the process of rating, and the reliability of raters based on different scales of scoring (Bacha, 2001; Barkaoui, 2007; Ghanbari & Barati, 2014). Considering this, very few studies have used scoring scales to diagnose the students' problematic areas in writing to overcome their weaknesses in writing performance.

In the EFL context of Iran, where the majority of teachers in writing classes subjectively assess and give feedback on their students' writings, there is a need for different ways of measuring their writings to better understand their problematic areas. The main point is that at the tertiary level, less attention has been paid to the writings of EFL learners and assessing their writings in a way that gives more information regarding their weaknesses and strengths in writing. In this case, some changes are needed in teaching writing and evaluating learners' writing in academic settings. Therefore, this study aimed at diagnosing TEFL undergraduate students' strengths and weaknesses in different aspects of their writing performance through an analytical scoring scale. In addition, it sought to determine which aspect or component of writing significantly contributes to the variation in the overall writing performance of the students. The results of this study can help both teachers and students better understand their weaknesses and strengths in different aspects of writing and improve the quality of their future writing.

2. Review of Literature

Assessment of students is one of the essential components of education to evaluate the effectiveness of a program and decision-making (Mohammadkarimi & Amin, 2019). Based on Suskie (2018), there are two methods of assessing the writings of learners: direct assessment and indirect assessment. He added that assessment in a direct way is more authentic and valid than indirect assessment. In the direct assessment, writing is assessed subjectively and it is according to texts that are produced. Basic skills, as well as composition, could be evaluated using this method (Hyland, 2019). Lee (2017) claimed that although the direct approach possesses face validity, it needs to be assessed subjectively, leading to the disagreement of the raters. Mainly because of issues related to the reliability of the direct approach to assessment, we have witnessed the emergence of different approaches to indirect assessment. On the other hand, Suskie (2018) believes that the indirect method is more flexible, reliable, and standardized in its scoring and administering. Tests and examinations based on this approach are both short-scale and large-scale. Examples of short-scale are cloze tests and multiple-choice tests, in which learners reveal their skills in constructing successful aspects of writing such as grammar and sentences. A large-scale example of the indirect approach to assessment is TOFEL.

Teachers usually use rubrics and checklists for evaluating learners` language skills, including their writing. To evaluate learners` performance over time, they will use a checklist, while to evaluate the performance quality of both written and oral tasks, a rubric will be used. Rubrics are developed according to various kinds of scales and are divided into multi-trait, primary-trait, holistic, and analytic categories (Tedick & Klee, 1998).

Cohen (1994) noted that the multi-trait scale observes multiple aspects of learners` writing. Each trait indicates a particular aspect of the essay, and since the essay is evaluated according to expectations, its validity is increased. Although this scoring scale has been considered a beneficial approach, the development and validity of traits in different contexts could be a challenge. The primary-trait scale is a general evaluation of writing, and if we pay more attention to a particular part of writing, each time one aspect will be focused on. In this approach, it is not easy to concentrate on only one particular aspect of scoring, and considering one particular trait as "primary" may not be suitable (Cohen, 1994).

In a holistic approach, writing would be evaluated based on its general impression and as a whole. According to Wolcott and Legg (1998), this approach is usually applied for scoring a number of writings with the same procedure and topic, so the quality of writings could be compared. Scores sometimes start at 1 (for the worst essays) and go up to 6 (for the best essays) in holistic scoring, and to enhance the reliability of the writing, two raters assess each essay. Although in this approach, essays will be looked at and scored as a whole, if some errors, for example, grammatical or spelling errors, have been repeated more and made the reading difficult, then it would cause a loss of marks (Haswell & Elliot, 2019). Like other scales, the holistic scoring scale has some limitations. First of all, students` performance could not be judged solely on a single score. Second, this approach does not provide feedback for students to revise their writing. Third, there is no reasonable explanation for the assigned scores. Fourth, an assigned score does not show the real skills of a student. Lastly, scores may not be fair in all aspects of writing for all students (Cohen, 1994).

Analytical scoring scales are considered as rubrics that consist of clear action expectations for every criterion, and they are suitable for learning activities that include several criteria (Galti, Saidu, Yusuf, & Goni, 2018). This scale has some advantages. First, it decreases the influence of test scores on the development of various sub-skills among test takers. Second, it helps administrators recognize areas of strength that can often remain unnoticed. Third, greater scoring reliability is achieved by asking the scorer to analyze each of the specified categories individually. Furthermore, inexperienced raters might find the analytical method easier to conduct than the holistic method. On the other hand, it requires more effort and time for scoring, which makes it unsuitable for large-scale tests (Al-Zubeiry, 2020; Veloo, Abd Aziz, & Yaacob 2018). Weir (1990) proposed an analytic scoring scale for scoring different sections of a piece of writing, including relevance and adequacy, organization, cohesion, vocabulary, grammar, punctuation, and spelling.

Regardless of the method of scoring, either holistic or analytic, generally, teachers find writing assessment a time-consuming duty because it requires considering different aspects such as the obvious prompt of an essay, suitable tasks of writing, time, and feedback. Some researchers were concerned about rubric reliability (Ghanbari & Barati, 2020; Turley & Gallagher, 2008; Wilson, 2007) and focused more on the appropriate use of rubrics and criteria for evaluating learners` writings (Chowdhury, 2019; Gallardo, 2020; Pablo & Lasaten, 2018) since no rubric is the best.

Moreover, writing assessment of English as second language (ESL) learners is not as easy as assessing native speakers` writings (Song & August, 2002). In this regard, it is noted that learners with various backgrounds in education and culture have different techniques and levels in their writing (Hamp-Lyons & Condon, 2000; Odena & Burgess, 2017). In this case, we can imply that ESL and English as a foreign language (EFL) learners write differently based on their different contexts. This calls for identifying their strengths and weaknesses in writing, especially their weaknesses in different aspects of writing, mainly in academic settings. This also requires the selection of a good assessment scale that can identify the quality of their writing from different aspects. Some researchers indicated the appropriateness of analytic scales over holistic scales and expressed that since different aspects of writing ability develop in various ways and analytic scales provide different rates for different aspects of writing performance, they are ideal for measuring the learners' writing performance (Knoch, 2009; Weigle, 2002; Wiseman, 2012). The analytic scales are also useful for diagnosing students' weaknesses and strengths in writing (Ghalib & Al-Hattami, 2015; Knoch, 2011). That is why the researcher in this study aimed to use an analytical scale for assessing Iranian undergraduate students' writing performance.

One of the most common ways to diagnose learners' writing weaknesses is to assess and provide feedback on their work. However, the scoring scale that the teacher or rater applies for rating learner essays heavily affects the results and understanding of the components of writing that learners have problems with (Nodoushan, 2014). Unlike the holistic scale, the analytic scoring scale focuses on each aspect of the learner's writing, making it easier for both the teacher and the learner to understand the weaknesses of their writing.

In this regard, Ahur and Mukundan (2009) in their study analytically assessed the writings of 128 national and international students at one of the universities in Malaysia to find

out the weak points as well as the strong points of their writings. They applied Weir`s (1990) analytical scoring scale to analyze collected essays from students. Their results revealed that learners paid more attention to organization criteria, and this component had less variance, making it one of the strong points of their writings. On the other hand, grammar and cohesion were components with a high level of variance and were considered weaknesses in students` writings. Moreover, it was found that the analytical scoring scale was regarded as a beneficial tool for understanding the strengths and weaknesses of learners` writing.

Wiseman (2012) compared the effectiveness of holistic and analytic scoring rubrics for assessing second language writing. His study used Rasch many-faceted measurement to investigate the performance of both rubrics in assessing students` writings. The outcomes of the study, similar to Ahur and Mukundan`s (2009) study, revealed that analytic rubrics are better tools for the purposes of placement and diagnostics.

Through an analytic scoring scale, Al-Mudhi (2019) evaluated the essay writing of 88 Saudi students. Based on the results of this research, students` overall performance in their writing was poor to fair. Their areas of weakness were related to vocabulary and text mechanics, while organization and content components were their strong points in their writing. Surprisingly, Al-Zubeiry (2020), who similarly investigated the issues of Saudi students' writings, obtained the opposite results. His findings showed that the main weaknesses in students` writing were in the areas of content, organization, and punctuation.

In their study, Sukandi and Sani (2020) compared the differences between two rates in the analytical scoring of student writing. Their findings revealed that the main areas of strength in their participants' writings were related to content and organization, while their weaknesses were related to the mechanics component (spelling and punctuation).

To the best of the researcher`s knowledge, analytic assessment of students' writings is less investigated in the literature, particularly in the context of TEFL in Iran. Therefore, this study was an attempt to find out the areas of strength and weakness in the writing performance of TEFL undergraduate students as well as to determine which components of writing (i.e., content, organization, cohesion, vocabulary, grammar, punctuation, and spelling) can significantly explain the variance in their overall writing performance. In this regard, the following research questions were posed:
1. What are the areas of strength and weakness in the writings of TEFL undergraduate students?
2. Do all components of writing (i.e., content, organization, cohesion, vocabulary, grammar, punctuation, and spelling) significantly contribute to the variance of TEFL undergraduate students' overall writing performance?

3. Method

3.1. Population and Sampling

The population of this study was all TEFL undergraduate students at different universities in Tabriz. Availability sampling was used to choose the participants in the study. According to Farhady (2008), in availability sampling, participants are selected based on their availability and willingness to participate in the research. The researcher tried to contact the students that had taken writing courses at these universities.

For the purpose of this study, more than 300 students were contacted, either directly or through their instructors. However, only 73 students were ready to participate in this study. Writing samples were collected from 73 students majoring in teaching English as a foreign language at different universities in Tabriz, Iran. As Table 1 illustrates, there were 41 females and 32 males, ranging from freshman to senior students, with a mean age of 23 years old.

Table 1
Demographic Information of Participants

N	Freshman Student	Sophomore Student	Junior Student	Senior Student	Age (Mean)	Female	Male
73	19	12	14	28	23	41	32

Moreover, two independent raters assessed the sample essays of the students. As Table 2 illustrates, they were both university instructors with more than 10 years of teaching and assessment experience.

Table 2
Demographic Information of Raters

Raters	Teaching Experience	Field of Study	Age	Gender
1	13	TEFL	39	Female
2	17	TEFL	46	Male

3.2 Instruments and Materials

3.2.1 Analytic Scoring Scale. To evaluate the essay samples of the learners, Weir's (1990) four-point analytic scoring scale (0-3 points) was employed (see Appendix). This scale assesses different components of writing, including content, organization, cohesion, vocabulary, grammar, punctuation, and spelling. This scale's reliability has been tested in various situations (Weigle, 2002); nevertheless, in the current study, its reliability was checked through Cronbach's alpha, which was 0.78, and it is acceptable.

3.2.2 Writing Samples. To collect the required data, the researcher needed the writing samples of the participant students. In this regard, five prompts were given to the teachers of these participants, and they agreed on the following two writing prompts, considering different criteria, including familiarity with the topics, interest, and the level of the students. Prompts were selected in a similar style, and the participants were asked to choose one of them and write an essay about them in which they were required to express their agreement or disagreement with the provided issue. The two writing prompts were as follows:

1. Some people believe that universities should accept an equal number of men and women in every subject. To what extent do you agree with this statement?

2. In this period of COVID_19 (Corona Virus), one solution for the continuing process of education is online teaching and learning. Do you agree or disagree with online teaching and learning?

The reasons for choosing these topics were that the first topic was an ongoing issue among students, and the second topic was a current challenge for them; moreover, they were familiar with this topic. The difficulty level and validity of these prompts were checked by two instructors from each level (i.e., freshman, sophomore, junior, and senior) of these participants, and they agreed on the suitability of the prompts for the level of the students. Therefore, they were used in the study.

3.3. Procedure

The main objectives of the study were explained to the instructors, who asked their students to voluntarily participate in the study. Because of COVID_19, the researcher was unable to contact the students face-to-face. So, he asked some university teachers to help and show students how to participate in this study online. In addition, the researcher introduced the assessing criteria, which was an analytic scoring scale, to the students so they would know how their writing would be assessed, and also, they consider these criteria more while they are writing. The instruction paper that was sent to the students included two pages in word format. On the first page, the students were asked to write about their demographic information. Furthermore, a scoring rubric and instructions for choosing one option from two provided questions were included on this page. The second page contained two topics and enough space for their essays.

To reduce the risk of academic dishonesty, students were required to submit their writings at an assigned time, which was one hour (45 minutes for essay writing and 15 minutes for its submission). Students had to choose one of the two topics that were given to them and write their essays and submit them in no more than an hour.

The researcher collected students' online writing samples. After printing two copies of each collected essay, they were coded and the same numbers were allocated to each copy. Then, they were submitted to the two independent raters who were experienced in teaching and assessing writing. The analytic scoring scale was also given to these raters, and they were asked to assess the writings of the students based on the criteria on this scale.

The researcher also provided the scoring sheet to the raters in order to record their ratings for each participant regarding different aspects of their writing. After collecting their ratings, the inter-rater reliability was checked, and the average of their ratings was used in the data analysis through the software of SPSS 22.

3.4. Design

The nature of the study required the investigation of participants' writing samples, which made the use of content analysis an appropriate design. In addition, because the aim was to find the weaknesses and strengths in the writing samples of the participants, these two were considered variables in the study.

3.5. Data Analysis

Before giving the writing samples to the raters, the first sheet of the printed essays was detached from all the essays on which the demographic information of the students was written. Instead, to eliminate scoring bias, the scoring sheets were attached to all the essays by the researcher. Then, they were given to the raters, who were TEFL instructors with more than 10 years of experience (13 years and 17 years) in teaching writing and assessment. To enhance the reliability of the content analysis, the seven aspects of writing (i.e., relevance and adequacy of content, organization, cohesion, vocabulary, grammar, punctuation, and spelling) in the scoring scale were discussed between the researcher and the raters, and they were asked to strictly follow these aspects. Then, by Pearson correlation, the inter-rater reliability was checked. Later, to answer the first research question, descriptive statistics such as standard deviation and mean were obtained. A stepwise multiple linear regression was conducted to answer the second research question. In this case, the seven components of writing (i.e., relevance and adequacy of content, organization, cohesion, vocabulary, grammar, punctuation, and spelling) were considered independent or predictor variables, and the total score of writing was considered the dependent or predicted variable. The alpha level for significance testing was set at .05. In this regard, all the necessary assumptions were checked and they were all met, which legitimized the use of parametric tests in this study.

4. Results

4.1 Areas of Strength and Weaknesses in the Writings of TEFL Undergraduate Students (Research Question One)

In order to answer the first research question, the analysis of ratings of the essays by two independent raters provided the following data for the first and second-year students (Table 4.1), as well as for the third and fourth-year students (Table 4.2). These tables show the standard deviation and mean for all essays according to seven aspects of writing in Weir's (1990) analytical scoring scale. The results are explained in parts (a) and (b) below.

4.1.1 Results for the First and Second Year Students. Based on two ratings, the mean (M) and standard deviation (SD) for all parts of freshman and sophomore students' writings are shown in Table 3.

Table 3

Descriptive Statistics of First and Second Year Students' (Gr1111) EssaR Ratings ww Tw waters
N=31

Rater		Content	Organization	Cohesion	Vocabulary	Grammar	Punctuation	Spelling	Overall Writing	Skewness
1	M	2.23	2.03	2.19	2.05	1.45	1.64	2.84	2.06	.028
	SD	.38	.49	.51	.57	.64	.61	.26	.68	
2	M	2.41	1.98	2.07	1.94	1.65	1.78	2.69	2.07	.057
	SD	.34	.47	.44	.52	.62	.67	.22	.26	
	M	2.32	2.00	2.13	1.99	1.55	1.71	2.76	2.07	.064
	SD	.36	.48	.47	.54	.63	.64	.24	.26	

According to Table 3, considering the average mean scores of raters in the writings' of these students, grammar and punctuation had the lowest means (1.55 and 1.71, respectively) and the highest standard deviations (.63 and .64, respectively). Whereas, spelling and content had the highest means (2.76 and 2.32, respectively) and the lowest standard deviations (.24 and .36, respectively). Table 4.1 also shows the skewness values for the data, and because they are within plus/minus 1 standard deviation, the distribution of the data is considered normal for the ratings of both raters. The results of this table indicate that the first and second-year students had the lowest mean scores in the components of grammar and punctuation among the seven components of writing. Pearson product-moment correlation was applied for calculating the inter-rater reliability. The results are shown in Table 4.

Table 4
Inter-rater Reliability for the Scores of First and Second Year Students

		Rater 1	Rater 2
Rater 1	Pearson Correlation	1	.438**
	Sig. (2-tailed)		.001
	N	31	31
Rater 2	Pearson Correlation	.438**	1
	Sig. (2-tailed)	.001	
	N	31	31

**. Correlation is significant at the 0.01 level (2-tailed).

As Table 4 indicates, the overall correlation between the two raters' ratings was significant and moderate ($r=.438$, $p=.001 < α=.05$).

4.1.2 Results for the Third and Fourth Year Students. Table 5 illustrates the descriptive statistics for the essay scores of junior and senior students by two raters.

Table 5
Descriptive Statistics of Third and hhhrhh hear Sttdents' tGrppp pp Essay tatin ts yy Two Raters

Rater		Content	Organization	Cohesion	Vocabulary	Grammar	Punctuation	Spelling	Overall Writing	Skewness
1	M	2.78	2.33	2.12	2.84	2.06	2.01	2.89	2.43	.564
	SD	.39	.58	.53	.26	.54	.64	.39	.67	
2	M	2.63	2.43	2.51	2.71	2.14	2.13	2.84	2.48	.623
	SD	.38	.49	.53	.38	.62	.62	.34	.57	
	M	2.70	2.38	2.31	2.77	2.10	2.07	2.86	2.46	.574
	SD	.38	.53	.53	.32	.58	.63	.36	.65	

Based on Table 5, the highest means were related to spelling (M=2.86) and vocabulary (M=2.77), with standard deviations of SD= .36 and SD= .32 respectively, while punctuation and grammar had the lowest means (2.07 and 2.10, respectively), with standard deviations of SD= .63 and SD= .58. The highest means imply strength, and the lowest means indicate the

weakness of the learners in those components. In this regard, the weakness of third and fourth-year undergraduate students was in the components of punctuation and grammar. In addition, the skewness values less than 1 indicate the normality of the distribution of data for the scoring of raters 1 and 2.

For estimating the inter-rater reliability, the Pearson product-moment correlation was also used. Table 6 presents its results.

Table 6

Inter-rater Reliability for the Scores of Third and Fourth Year Students

		Rater 1	Rater 2
Rater 1	Pearson Correlation	1	.486[**]
	Sig. (2-tailed)		.001
	N	42	42
Rater 2	Pearson Correlation	.486[**]	1
	Sig. (2-tailed)	.001	
	N	42	42

**. Correlation is significant at the 0.01 level (2-tailed).

As Table 6 demonstrates, there was a positive and moderate correlation between the ratings of the two raters, and this result was also significant (r=.486, p=.001< $α$=.05).

Further investigation of the scores of raters in both groups showed that variation among different components was less than one point in both groups. However, for each component, the average score by two raters was calculated and used in the inferential statistical analysis.

4.2 Contribution of Components of Writing to the Overall Writing Performance of TEFL Undergraduate Students (Research Question Two)

In order to answer the second research question, Stepwise Multiple Linear Regression (MLR) was applied. This model of MLR included the following predictors: mechanical accuracy for spelling and punctuation, grammar, vocabulary sufficiency for purpose, cohesion, organization, relevance, and adequacy of content, as well as a dependent or predicted variable, which was the total ratings of essays. The results are presented in two separate parts for the first and second-year students' scores (part a) and the third and fourth-year students' scores (part b).

4.2.1 Results for the First and Second-Year Students. The results of the model summary in the stepwise multiple regression analysis of essay components of first and second-year students are shown in Table 7.

Table 7

Model Summary for Components of Writing in Group 1

Variables	R^2	R^2 change	F change	p-value
G	.379	.379	124.895	.0001
G+P	.592	.213	94.247	.0001

G+P+O	.774	.182	82.361	.0001
G+P+O+V	.871	.097	76.598	.0001
G+P+O+V+C	.936	.065	162.574	.0001
G+P+O+V+C+C2	.979	.043	235.658	.0001
G+P+O+V+C+C2+S	1.000	.021	475.234	.0001

G= Grammar; P=punctuation; O=organization; V=vocabulary; C=cohesion; C2=content; S=spelling

Based on the results in Table 7, as R^2 change shows, the component of grammar in group 1 had the greatest contribution to the variance of their overall writing performance (37.9%). The second greatest contribution was related to the punctuation variable (21.3%). Other variables which had a relative contribution to the overall writing performance were organization (18.2%), vocabulary (9.7%), cohesion (6.5%), and content (4.3%), respectively, as well as spelling (2.1%), which had a minimum contribution to the variance of essay quality.

4.2.2 Results for the Third and Fourth-Year Students. Multiple regression analysis for the third and fourth-year students' essay components is shown in Table 8.

Table 8

Multiple Regression Analysis for Components of Essay Writing (Group 2)

Variables	R^2	R^2 change	F change	p-value
P	.296	.296	124.895	.0001
P+G	.558	.262	94.247	.0001
P+G+C	.775	.217	82.361	.0001
P+G+C+O	.873	.098	76.598	.0001
P+G+C+O+ C2	.940	.067	162.574	.0001
P+G+C+O+ C2+V	.976	.036	235.658	.0001
P+G+C+O+ C2+V+S	1.000	.024	475.234	.0001

P=punctuation; G= Grammar; C=cohesion; O=organization; C2=content; V=vocabulary; S=spelling

According to Table 8, unlike in group 1, in group 2 students, punctuation (29.6%) had the greatest significant contribution to the overall writing variance, while grammar (26.2%) had the second greatest contribution. Other contributors to this group were cohesion (21.7%), organization (9.8%), content (6.7%), and vocabulary (3.6%). Similar to group 1, spelling (2.4%) was the least contributory factor in group 2.

Moreover, the beta regression coefficient enabled the researcher to better compare the contributions of the predictor variables and find the most significant predictors of the variation in the overall writing of the first and second-year undergraduate students (Group1) and the third and fourth-year undergraduate students (Group 2) (see Table 9).

Table 9

Beta Regression Coefficient for Essay Components

Group	Content	Organization	Cohesion	Vocabulary	Grammar	Punctuation	Spelling
Group1 β	.186	.273	.238	.245	.329	.298	.168

t	2.65	4.81	3.26	3.50	5.25	4.96	2.48
P	.181	.003	.095	.091	.000	.002	.194
Group 2 *β*	.237	.253	.264	.194	.287	.307	.154
t	2.98	3.24	4.76	2.72	4.91	5.15	2.24
p	.09	.092	.007	.184	.006	.000	.196

As Table 9 reflects, in group 1, grammar ($β$=.329) was the most significant contributor to the overall variation of essay quality followed by the components of punctuation (==.298, *p*= .002) and organization (==.273, p= .003). It proposes that changing one standard deviation of grammar leads to an increase of .329 standard deviations of the variable criterion. Spelling, on the other hand, had the least coefficient ($β$=.168, *p*= .194) among other components, indicating that changing one standard deviation in spelling results in an increase of .168 standard deviation in the overall essay. The contributions of other components in the essay quality variation were non-significant: organization (.273), vocabulary (.245), cohesion (.238), and content (.186).

However, in group 2, as illustrated in Table 4.7, the highest contributor to the overall variation was punctuation ($β$=.307, *p*= .000); that is, changing one standard deviation of punctuation leads to an increasing .307 standard deviation of the variable of criterion. This is followed by the significant contributions from the components of grammar (==.281, *p*= .006) and cohesion (==.264, *p*= .007). In contrast, spelling, with a coefficient of $β$= .154 was the smallest non-significant contributor. There was also a non-significant contribution from the components of vocabulary, content, and organization.

The obtained data from both groups revealed that tolerance in all predictor variables was greater than 0.10 and the data of the Variance Inflation Factor (VIF) was not more than 10.0, indicating that there was not an important issue of multicollinearity. In this regard, the obtained results would be trusted and interpreted confidently. Therefore, based on the results of Table 4.7, in the writings of the first and second-year undergraduate students, the variables of grammar, punctuation, and organization, respectively, were considered the greatest and most significant contributors to the variance of their overall writing scores. This is while, in the writings of third and fourth-year undergraduate students, the components of punctuation, grammar, and cohesion, respectively, were the significant contributors to the variance of their overall writing scores.

In conclusion, this section answers the research questions of the study.
1. What are the areas of strength and weakness in the writings of TEFL undergraduate students? As mentioned, writing strength in the students of group 1 was related to spelling and content components, and in group 2, their strong points were the spelling and vocabulary components of their writing. On the other hand, weaknesses in group 1 were mainly in grammar, punctuation, and organization components, and in group 2, punctuation, grammar, and cohesion components, respectively.
2. Do all components of writing (i.e., content, organization, cohesion, vocabulary, grammar, punctuation, and spelling) significantly contribute to the variance of TEFL undergraduate students' overall writing performance?

As revealed by the results, the degree of contribution of components to the variance of the total score was different. In group 1, the grammar component had the highest significant contribution, followed by the components of punctuation and organization. This is while the smallest non-significant contribution was related to the spelling component. In group 2, the most significant contributor was punctuation, followed by the components of grammar and cohesion, while the lowest contributor, similar to group 1, was the spelling component.

5. Discussion

In this research, 73 essay samples by TEFL students were investigated to find the respective contributions of all essay components to the variance in overall learners' essays. Since it was perceived that students in different years of study may vary in their writing performance, their results were analyzed into two groups: group 1, first-year and second-year students, and group 2, third-year and fourth-year students.

In multiple regression analysis, the stepwise technique was used, indicating that the great contribution in writing variance of group 1, is concerned with grammar, punctuation, and organization components. In addition, a lower contribution of spelling, cohesion, content, and vocabulary variables to the total variance is reported. Likewise, statistics from group 2 presented significant contributions from punctuation, grammar, and cohesion components, while spelling, vocabulary, content, and organization components, respectively, had smaller and non-significant contributions to the total variance.

The researcher compared the findings of this study with some previous studies in this regard. The results of the current research were not in complete agreement with some studies (Ahur and Mukundan, 2009; Al-Zubeiry, 2020; Astika, 1993; Sukandi & Sani, 2020). In his study, Astika (1993), through an analytic scoring scale, analyzed the essays of foreign students at Hawaii University. His study revealed that the greatest contributor to the overall variance of an essay was vocabulary (83.75%), whereas the contribution of mechanics was the least (.29%) to the variance. Other contributors were content (8.06%), language use (4.05%), and organization (2.48%), respectively. Moreover, in another study, Ahur and Mukundan (2009) analyzed the sample essays of TESL learners through the analytic scoring scale of Weir (1990) to find out the strong and weak points of their writing. Their findings showed that grammar (48.5%) was the greatest contributor, and the least contributor to the variance was organization (1.6%).

However, in the present study among the TEFL undergraduate students in group 1, the greatest contribution to the overall score variance was related to grammar (37.9%) and the least contribution was related to the spelling component (4.3%). In group 2, the punctuation component (29.6%) had the greatest contribution, and the spelling (2.4%) component had the lowest contribution.

In Astika's (1993) study, the number of components contributing to the variance of total ratings was presented as R^2 change, and it was compared with the findings of Ahur and Mukundan (2009) and the current study (See Table 10).

Table 10

Ctttarint t te FindinFs tt t tis St tdt witt tst ika's)))))) and)))r and)   )k undan's uuuuuu

R^2 change

Variables	Astika (1993)	Ahur and Mukundan (2009)	This study (2020) Group 1	Group 2
Vocabulary	.8375	.110	.097	.036
Content	.0806	.047	.043	.067
Language use	.0405	.695(Grammar+Cohesion)	.444(Grammar+Cohesion)	.479(Grammar+Cohesion)
Organization	.0248	.016	.182	.098
Mechanics	.0029	.132(Spelling+Punctuation)	.234(Spelling+Punctuation)	.320(Spelling+Punctuation)

As Table 10 illustrates, components of mechanics in Astika correspond with spelling and punctuation in the current study as well as in Ahur and Mukundan`s (2009) study, and language use in Astika corresponds with grammar and cohesion in these two studies. Therefore, in the current study, the variances of components in the total rating for group 1 (first and second-year students) were language use (44.4%), mechanics (23.4%), organization (18.2%), vocabulary (9.7%), and content (4.3%), respectively. For group 2 (third and fourth-year students), the variances of components were language use (47.9%), mechanics (32%), organization (9.8%), content (6.7%), and vocabulary (3.6%), respectively.

Generally, vocabulary in Astika's study was the greatest contributor to the variance, and mechanics was the smallest contributor. In the study by Ahur and Mukundan (2009), grammar and cohesion were the greatest contributors, and organization had the lowest contribution to the variance of the total rating.

On the other hand, comparing the results of this study with those of Astika (1993) and Ahur and Mukundan's (2009) studies, as indicated in Table 4.7, the greatest contributions to the variance of the total rating of TEFL undergraduate students' writings were related to the language use (grammar and cohesion) and mechanics (spelling and punctuation) components in groups 1 and 2, while the least contributions were related to the content and vocabulary in group 1, and vocabulary and content in group 2, respectively. It indicates that the weaknesses and issues of participants in both groups of this research were mostly related to grammatical structure (group 1) and using punctuation marks in paragraphs (group 2), whereas in Astika's (1993) study, foreign learners did not have these issues to that extent in their writings, but TESL students in the study of Ahur and Mukundan (2009) had a greater number of problems with grammatical structure and fewer problems with using punctuation marks.

In general, regardless of the classification of components in Astika's (1993) study, it is implied that learners in group 1 are better in spelling and content, and group 2 are better in spelling and using suitable vocabulary in writing essays. In group 1, knowledge of content could be related to giving two options for selecting the topic of an essay. When learners are free to select one option to write about, they normally select a topic that they are familiar with. In group 2, since they were third-year and fourth-year students (mostly in their fourth year), studying at the university helped them improve their lexical knowledge. Although better performance and strength in spelling could be related to their proficiency in this component of writing, using a computer and specifically Microsoft Word, might have helped them understand

and correct their spelling mistakes (if, of course, they noticed them), which was inevitable to prevent in this study.

There is a relative agreement between the results of this study and those of Sukandi and Sani`s (2020) study. Their findings revealed that the main areas of strength of their participants' writings were related to the content and organization components, while their weaknesses were related to mechanics (spelling and punctuation). In this study, spelling and content items were the students' strong points, while grammar and punctuation items were their weaknesses. As mentioned, higher scores of spelling in this study could be related to auto-correction in Microsoft Word. Similarly, the current study's findings are consistent with Al-Mudhi`s (2019) research, in which participants' strong points in writing were organization and content. However, in Al-Zubeiry`s (2020) study that investigated the problems of writing in Saudi students, the main weaknesses in students' writing were in content and organization components.

It is stated that the analytic scoring scale should be applied to the higher levels of writers (Galti et al. 2018) and that it is an appropriate tool for diagnostic purposes (Wiseman, 2012). The result of this research showed that the analytical scoring scale was suitable for assessing the writing of students and could diagnose the areas of weakness and strength in their writing at the tertiary level.

Overall, the issues and problems for the undergraduate students in both groups were producing accurate grammatical structures and appropriate use of punctuation marks in their writings. One reason for this could be the increased focus on producing abstract words and phrases rather than formulating correct paragraphs and essays in schools and even universities in Iran. Therefore, these students who are capable of producing high-quality essays require further awareness about structure use and its instruction.

6. Conclusion

Iranian students, even in the field of English Language Teaching (ELT), do not have enough practice in essay writing and the teachers do not provide enough feedback on all aspects of their writing. Their assigned essays are usually assessed impressionistically and without any scoring scales. Therefore, students would not be aware of the weak points in different aspects of their writing. In this case, the present study attempted to find the students' areas of difficulty in their writings based on an analytical scoring scale. This content analysis study was an attempt to investigate the relative contribution of each writing component to the variation in the overall writing performance of TEFL undergraduate students to determine areas of strength and weakness in their writing. Thus, the writing samples of 73 students who were willing to participate in this study at different universities in Tabriz, Iran, were assessed using an analytical scoring scale. These participants were divided into two groups: group 1, first-year and second-year students, and group 2, third-year and fourth-year students. Descriptive statistics, including mean and standard deviation, as well as inferential statistics such as stepwise multiple linear regression, were used to find the respective contributions of all components of writing to the total score of writing. Results revealed that the first and second-year undergraduate students in group 1 mainly concentrated on the spelling and content components that presented the lowest variance in overall rating, and in group 2, the third and

fourth-year students' main concerns were spelling and vocabulary, indicating the smallest amount of variance. These results imply that the familiarity of students in group 1 with the topic could help them to perform better in their writing and get higher scores and that years of study at university and paying more attention to the vocabulary helped students in group 2 to improve their lexical knowledge.

Based on Firoozi et al. (2019), despite the educational reform in Iran, teachers are required to modify their attitude toward language assessment. Moreover, training teachers in language assessment in both theory and practice is necessary. Particularly, it is important for them to attend training programs for designing rubrics for assessing students` writings.

Therefore, many related stockholders and authorities could benefit from the outcomes of this study. School teachers and university professors should pay more attention to all aspects of students' writing, particularly their areas of weakness. Rubrics could have benefits for both teachers and students, as they could improve instruction and increase learning. The rubric makes the criteria of instruction and assessment clear for teachers and students and tells them what is significant in assessment. As a result, it is suggested that teachers apply an analytical scoring scale for assessing the essays of students. An analytical scoring scale is beneficial for teachers to diagnose the students` writing weaknesses and accordingly tailor their instruction to alleviate these drawbacks and improve the quality of their essays. The results of the study can be useful for teachers who have writing courses to modify their teaching and evaluation methods of writing and consider alternative ways of assessing students' writing samples. In addition, it can help teachers identify the areas of difficulty in students' writings and put a greater focus on those aspects in their teaching. The findings can also be useful for students in writing classes in that they can be aware of their problematic areas and practice more to improve the quality of their writing.

Because of the incidence of COVID-19, universities were almost closed and it was very difficult to find more participants for this study. Further research could be conducted with a higher number of participants from various universities and contexts. In this study, the aim was to find out the areas of strength and weaknesses of TEFL undergraduate students' writing, other studies can reflect on these issues for postgraduate students. It is also recommended for further research to compare the ESL and/or EFL undergraduate and postgraduate students' weak and strong areas in their writing performance. Future studies can also compare the male and female undergraduate students' weaknesses and strengths in writing and find their problematic areas that require more consideration in teaching/learning situations. In addition, more studies about scoring scales, rubrics, and instructions for online and computer-assisted writing are required, because they may require different allocating of time, especially for typing, and may have issues with some components, such as auto-recognize and correction of spelling on the computer, as well as the possibility of academic dishonesty in this kind of writing. Finally, Hunter, Jones, and Randhawa (1996) concluded that coupling analytic with holistic ratings provided detailed diagnostic information with a more general profile of writing quality. Therefore, further studies seem to be necessary to investigate the quality of writing both holistically and analytically.

Declaration of Conflicting Interests

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Ahur, T., & Mukundan, J. (2009). Analytic assessment of writing: Diagnosing areas of strength and weakness in the writing of TESL undergraduate students. *Iranian Journal of Language Studies*, *3*(2), 195-208.

Al-Mudhi, M. A. (2019). Evaluating Saudi university students' English writing skills using an analytic rating scale. *Journal of Applied Linguistics and Language Research*, *6*(4), 95-109.

Al-Zubeiry, H. Y. A. (2020). Coherence problems in the English writings of Saudi university students. *Journal of Language Sciences and Literature*, *25*(1), 408-431.

Astika, G. G. (1993). Analytical assessments of foreign students' writing. *RELC Journal*, *24*(1), 61-70. https://doi.org/10.1177/003368829302400104

Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, *12*(2), 86-107. https://doi.org/10.1016/j.asw.2007.07.001

Chowdhury, F. (2019). Application of rubrics in the classroom: A vital tool for improvement in assessment, feedback, and learning. *International Education Studies*, *12*(1), 61-68. https://doi.org/10.5539/ies.v12n1p61

Cohen, A. D. (1994). *Assessing language ability in the classroom*. USA: Heinle & Heinle Publishers.

Gallardo, K. (2020). Competency-based assessment and the use of performance-based evaluation rubrics in higher education: Challenges towards the next decade. *Problems of Education in the 21st Century*, *78*(1), 61-79.

Galti, A. M., Saidu, S., Yusuf, H., & Goni, A. A. (2018). Rating scale in writing assessment: Holistic vs. analytical scales: A review. *International Journal of English Research*, *4*(6), 4-6.

Ghalib, T. K., & Al-Hattami, A. A. (2015). Holistic versus analytic evaluation of EFL writing: A case study. *English Language Teaching*, *8*(7), 225-236. http://dx.doi.org/10.5539/elt.v8n7p225

Ghanbari, N., & Barati, H. (2014). Iranian EFL writing assessment: The agency of rater or rating scale? *International Journal of Language Testing*, *4*(2), 204-228.

Ghanbari, N., & Barati, H. (2020). Development and validation of a rating scale for Iranian EFL academic writing assessment: A mixed-methods study. *Language Testing in Asia*, *10*(1), 1-21. https://doi.org/10.1186/s40468-020-00112-3

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio*. NJ: Hampton Press, INC.

Haswell, R., & Elliot, N. (2019). *Early holistic scoring of writing: A theory, a history, a reflection*. University Press of Colorado.

Hunter, D. M., Jones, R. M., & Randhawa, B. S. (1996). The use of holistic versus analytic scoring for large-scale assessment of writing. *The Canadian Journal of Program Evaluation*, *11*(2), 61-85.

Hyland, K. (2019). *Second language writing*. Cambridge university press.

Knoch, U. (2009). *Diagnostic writing assessment: The development and validation of a rating scale* (Vol. 17). Frankfurt: Peter Lang.

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from?. *Assessing Writing*, *16*(2), 81-96. https://doi.org/10.1016/j.asw.2011.02.003

Knoch, U., Read, J., & Randow, J. V. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, *12*(1), 26-43. https://doi.org/10.1016/j.asw.2007.04.001

Kroll, B. (Ed.) (1990). *Second language writing: research insights for the classroom*. Cambridge University Press.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer Singapore.

McMaster, K. L., Lembke, E. S., Shin, J., Poch, A. L., Smith, R. A., Jung, P.-G., Allen, A. A., & Wagner, K. (2020). Supporting teachers' use of data-based instruction to improve students' early writing skills. *Journal of Educational Psychology, 112*(1), 1–21. https://doi.org/10.1037/edu0000358

Mohammadkarimi, E., & Amin, M. Y. M. (2019). ELT students' attitudes toward the effectiveness the anti-plagiarism software, Turnitin. *Applied Linguistics Research Journal*, *3*(5), 63-75. https://dx.doi.org/10.14744/alrj.2019.66376

Nakamura, Y. (2004). *A comparison of holistic and analytic scoring methods in the assessment of writing*. Retrieved from https://jalt.org/pansig/2004/HTML/Nakamura.htm

Odena, O., & Burgess, H. (2017). How doctoral students and graduates describe facilitating experiences and strategies for their thesis writing learning process: A qualitative approach. *Studies in Higher Education*, *42*(3), 572-590. https://doi.org/10.1080/03075079.2015.1063598

Pablo, J. C. I., & Lasaten, R. C. S. (2018). Writing difficulties and quality of academic essays of senior high school students. *Asia Pacific Journal of Multidisciplinary Research*, *6*(4), 46-57.

Richards, J. C. (2002). *New interchange 1 lab guide: English for international communication* (Vol. 1). Cambridge University Press.

Schwarz, J. A., & Collins, M. L. (1995). *Improving the reliability of a direct writing skills assessment*. Paper presented at the annual meeting of the international personnel management association assessment council conference, New Orleans, LA. (ED 393880).

Song, B., & August, B. (2002). Using portfolios to assess the writing of ESL students: A powerful alternative? *Journal of Second Language Writing*, *11*, 49-72.

https://doi.org/10.1016/S1060-3743(02)00053-X

Sukandi, S. S., & Sani, R. D. (2020). Comparing two different instructor-raters in assessing EFL students' paragraph writings: An action research. *Journal of Language, Literature and Teaching*, *2*(1), 120-136.

Suskie, L. (2018). *Assessing student learning: A common sense guide*. San Francisco: John Wiley & Sons.

Tedick, D. J., & Klee, C. A. (1998). *Alternative Assessment in the Language Classroom.* Washington, DC: Center for applied linguistics. (ERIC Document Reproduction Service No. ED 433720).

Turley, E. D., & Gallagher, C. G. (2008). On the uses of rubrics: Reframing the great rubric debate. *English Journal*, 79(4), 87–92. https://doi.org/10.2307/30047253

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Weir, C. J. (1990). *Communicative language testing*. Englewood Cliffs, NJ: Prentice Hall Regents.

Wilson, M. (2007). Why I won't be using rubrics to respond to students' writing. *English Journal*, 96(4), 62–66. https://doi.org/10.2307/30047167

Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *International Journal of Language Testing*, *2*(1), 59-92.

Wolcott, W., & Legg, S. M. (1998). *An overview of writing assessment: theory, research and practice.* Urbana, IL: NCTE.

Appendix

Analytic Scoring Scale, Weir (1990)

A. Relevance and adequacy of the content

0. The answers bear almost no relation to the task set. Totally inadequate answer.

1. Answers of limited relevance to the task set. Possibly major gaps in the treatment of topic and/or pointless repetition.

2. For the most part answers the task set, though there may be some gaps or redundant information.

3. Relevant and adequate answer to the task set.

B. Compositional organization

0. No apparent organization of content.

1. Very little organization of content. The underlying structure is not sufficiently controlled.

2. Some organizational skills are evident, but not adequately controlled.

3. Overall shape and internal pattern clear. Organizational skills are adequately controlled.

C. Cohesion

0. Cohesion is almost totally absent. Writing is so fragmentary that comprehension of the intended communication is virtually impossible.

1. Unsatisfactory cohesion may cause difficulty in comprehension of most of the intended communication

2. For the most part satisfactory cohesion although occasional deficiencies may mean that certain parts of the communication are not always effective.

3. Satisfactory use of cohesion resulting in effective communication 73

D. Adequacy of vocabulary for purpose

0. Vocabulary inadequate even for the most basic parts of the intended communication

1. Frequent inadequacies in vocabulary for the task. Perhaps frequent lexical inappropriacies and/or repetition.

2. Some inadequacies in vocabulary for the task. Perhaps some lexical inappropriacies and/or circumlocution.

3. Almost no inadequacies in vocabulary for the task. Only rare inappropriacies and/or circumlocution.

E. Grammar

0. Almost all grammatical patterns are inaccurate.

1. Frequent grammatical inaccuracies.

2. Some grammatical inaccuracies.

3. Almost no grammatical inaccuracies.

F. Mechanical accuracy I (punctuation)

0. Ignorance of conventions of Punctuation.

1. Low standard of accuracy in Punctuation.

2. Some inaccuracies in punctuation.

3. Almost no inaccuracies in punctuation.

G. Mechanical accuracy II (spelling)

0. Almost all spelling is inaccurate.

1. Low standard of accuracy in spelling.

2. Some inaccuracies in spelling.

3. Almost no inaccuracies in spelling.