

Investigating Source-Text Use in Academic Writing Tasks of EFL Students through Integrated Writing Assessment

¹Amin Karimnia*

²Mahsa Safnyyat

IJEAP- 2105-1725

Received: 2021-07-08

Accepted: 2021-09-26

Published: 2021-09-30

Abstract

This study investigated the writing quality of EFL students through the factors of source text use (source idea, linguistic modification, and content accuracy) in integrated writing assessment. The participants were 30 MA-level EFL students from three different universities in Fars province, Iran. The participants were selected through purposive sampling. The test used in the study was a standardized three-part TOEFL package with reading, listening and writing sections. The study used both qualitative and quantitative data to report the findings. In the qualitative part, some instances of source idea, substantial linguistic modification, and content accuracy were listed in tables to show how the participants generally responded to these factors. In the quantitative part, multiple regression analysis was used and all of the assumptions were verified. The results revealed that the predictors could significantly explain 25.8% variance in the writing quality scores. Furthermore, out of the three factors of source text use, “substantial linguistic modification” was the only significant component, which was seemingly influenced by the participants’ advanced level and the nature of the writing task. The related implications were also elaborated.

Keywords: Source Text Use, EFL Writing, Integrated Writing, Writing Assessment, MA-Level EFL

1. Introduction

In education systems, especially those focusing on research, learners have to be able to write systematically and academically (Dangal et al., 2017). An academic piece of writing usually relies on and responds to other texts. Today, using writing tasks requiring other language skills (listening and reading) are becoming more popular in standardized academic examinations (e.g. TOEFL) in which learners have to use their “creativity”, rather than verbatim memorization, to respond to writing questions (Derntl, 2014). As such, the test-taker must be capable of incorporating reading-based and listening-based data while trying to produce the writing task (Cumming, 2013).

In line with the new expectations both in EFL tests and academic writing concerns (e.g. style, paraphrasing, and avoiding plagiarism) (Keck, 2006; Uludaga et al., 2019), integrated writing assessment in which learners are presented with reading or listening material (source text) and are required to compose a text based on the given information can be a highly effective method of teaching MA-level students about how to prepare their final research projects and theses. Considering the *focused* nature of integrated writing assessment, its accurate measurement of the criteria involved, and the new trends of critical writing (both in EFL teaching and academic research), this method could very effectively serve the teaching-related aspects of research in preparatory courses.

Integrating skills are significant for communicative competence for logical structuring of sentences to express desires and ideas both in oral and written communication. Yet, one of the serious difficulties in integrated writing assessment is how to measure test-takers’ written productions in response to the information in listening and reading sections (source texts). How should examiners make sure that students produce original ideas, diversified linguistic patterns, and accurate content? Another problem is that testing multi-literacies seems to be a challenging task for many students who are not adequately prepared (Cumming, 2013). This specification of integrated writing assessment

¹Associate Professor in TEFL, aminkarimnia@yahoo.com; Department of English Language, Fasa Branch, Islamic Azad University, Fasa, Iran.

²MA in TEFL, safnyyatt@yahoo.com; Department of English Language, Fasa Branch, Islamic Azad University, Fasa, Iran.

makes it particularly relevant to the level of advanced learners such as MA-level EFL students who academic study English. Source text use in such a context could be very effective as it involves tasks that provide a simplified version of academic writing. However, exploring integrated writing assessment in academic EFL contexts remains an underdeveloped topic that deserves attention.

Using integrated writing assessment at an MA level in a non-English speaking context seems to be an innovative contribution of the present study. Such a context avoids the problems associated with integrated writing assessment, since novice writers find it difficult to respond to a reading and listening startup and then transform their understanding into a written medium. As a result, this study could help us to figure out how integrated writing assessment could help strengthen the educational methods in research writing at a higher education level.

Employing a mixed research method, this study investigates the writing productions and writing quality of Iranian MA-level EFL students by exploring the three criteria of source text use in integrated writing assessment. This study seeks to answer two basic questions: (a) what are some instances of source idea, linguistic modification, and content accuracy in the writing of Iranian MA EFL learners? And (b) can source idea, linguistic modification, and content accuracy significantly predict writing quality in the sample under investigation? In doing so, the study first finds the instances of each of the three criteria, as a qualitative contribution. Following that, the study uses multiple regression test to examine the predictability of variance in the dependent variable (writing quality) through the independent variables of source text use (source idea, linguistic modification, and content accuracy).

2. Literature Review

2.1. Theoretical Studies

2.1.1. EFL and integrated writing assessment

One of the academic majors in which writing is highly significant is English as foreign language (EFL), which is offered at many universities worldwide. Preparing, writing and publishing research productions is one of the major purposes of postgraduate EFL studies in which learners have to be able to write systematically and academically (Dangal et al., 2017). Academic writing, however, represents a challenging process, because students over the course of their studies may need to incorporate data from various oral (e.g., lectures, conferences, seminars) and visual (e.g., electronic books or papers, printed materials) resources. As such, students must prepare for both listening and reading activities to be able to write essays (Gholami & Alinasab, 2017; Plakansa & Gebrilb, 2016).

Furthermore, texts in an academic context are frequently produced to respond to other writings (or publications). In fact, one of the major reasons an academic article is written is to respond to one or more texts (Dangal, Hamal & Giri, 2017; Derntl, 2014; Grabe & Zhang, 2013; Whitaker, 2010). An academic writing task must deal with the ideas, arguments, and topics in a source text. Considering these issues, standardized academic examinations require learners to use their “creativity”, rather than verbatim memorization, to respond to writing-based questions (Soleimani & Mahdavi-pour, 2014; Wette, 2018). In line with the new expectations both in EFL tests and academic writing concerns (e.g., style, paraphrasing, and avoiding plagiarism) (Keck, 2006; Uludaga et al., 2019), integrated writing assessment can be an effective method of teaching MA-level students about how to prepare their final research projects and theses.

Integrated writing tasks are normally used to assess academic writing competence in standard proficiency exams (Cumming, 2013; Yang & Plakans, 2012). According to Gholami and Alinasab’s observation (2017, p. 128), many valid language proficiency examinations, including the “Internet-based Test of English as a Foreign Language (TOEFL iBT), Canadian Academic English Language (CAEL), Ontario Test of English as a Second Language (OTESL), and Certificate of Proficiency in English (COPE)”, have merged reading- and/or listening-based tasks within integrated writing designs. Under such circumstances, examinees construct an integrated essay by considering the information communicated in source texts. TOEFL iBT seems to be a highly valid and widely used test currently employed (Yang & Plakans, 2012). Such tests are usually distinguished from what is

traditionally called “independent tests” (Gholami & Alinasab, 2017). Many researchers have also tried to compare integrated and independent writing tests. For instance, as Soleimani and Mahdavi-pour (2014) state:

Independent writing tasks are believed to offer a more valid demonstration of underlying writing ability in comparison to indirect writing assessment (e.g., multiple choice items) [...] Nevertheless, independent tasks have been criticized by many researchers [...] Given this criticism, integrated tasks have been regarded as an alternative component in writing tests (Soleimani & Mahdavi-pour, 2014, p. 132).

Considering this issue, some researchers have come to believe that integrated tasks may be a stronger and more valid representative of writing competence among test-takers.

2.1.2. Source text use in integrated writing assessment

One of the major issues in integrated writing assessment is how to distinguish and score students' answers to the writing task. In what could be a meta-study of integrated writing assessment, Cumming (2013) has provided a list of “promises” and “perils” of this mode of assessment. Realistic literacy activities, specific content, and multi-literacies are the promises of such mode of writing assessment (Cumming, 2013, p. 2). Despite these specific advantages, integrated writing assessment may exhibit some difficulties as well, such as complex measurement, the involvement of ill-defined genres, and threshold levels of abilities for competent performance (Cumming, 2013, p. 2).

As such, integrated writing assessment seems to be relatively difficult to be implemented, particularly as far as the final scoring is concerned. In the past few years, however, new frameworks have been proposed to address the problem of *source text use* in integrated writing assessment. The question is how to assess a student's writing in response to the listening and reading sections preceding the writing test. Uludaga et al. (2019) have proposed a measurement tool that could be employed in the case of source text use. The framework could help to solve the problem of source text use. The elements Uludaga et al. (2019, p.3) implemented in their source text use assessment are: source idea, substantial linguistic modification, content accuracy.

Source idea: This variable clarifies how the source text idea was reflected in the writing essay; for this purpose, three sub-measures are taken into account: (a) exemplification: in this case, the examinee's mention of the examples in the source text are considered; such instances are normally accompanied with a trigger such as *for example* or *as an instance*; (b) facts: facts are generally accepted propositions usually reporting a scientific reality or any mode of common knowledge; and (c) ideas: the assumptions, hypotheses, and possibilities mentioned in the source text; these are usually known as implications stated by the source text author.

Substantial linguistic modification: this category seeks to find how much variation is implemented on a source text piece when represented in a participant's writing. A substantial modification occurs when the participant implements an acceptable mode of paraphrasing (lexically and syntactically) in his/her representation of the source text in the writing essay.

Content accuracy: This measure tries to examine how precise the information in the source text is reflected in the writing. For instance, a proposition, even if paraphrased, must basically report the fact/idea in the source text; for instance, in the sentence “There is a correlation between air pollution and industries”, it would be a misrepresentation if the participant wrote “Industries cause air pollution” by changing the epistemic content in the writing.

2.2. Empirical Studies

Gholami and Alinasab (2017) investigated source-based tasks in writing independent and integrated essays, trying to explore the effects of source-based writing instruction and the juxtaposition of independent and integrated essay tasks on EFL learners' writing competence. In doing so, 20 advanced female learners who had attended a three-month TOEFL iBT preparation course at a private language institute in Iran, were selected as participants of their study. To gather data, they modified Longman Preparation Course for the TOEFL Test and Barron's writing for the TOEFL iBT. It

included five independent essay writing tasks for both groups of learners that were implemented in five sessions. The researchers found that frequent practice and incorporation of independent plus integrated essay tasks improved EFL writing ability on the part of the learners. The experimental group exhibited variety in using appropriate vocabulary, adequate ideas, and proper structures in integrated writing tasks. In addition, different aspects of integrated writing such as summarizing, note-taking, and paraphrasing, as addressed in the treatment, seemed to have been implemented in the integrated tasks in post-test compared to pre-test.

Sheibani and Ahmadi (2018) compared the writing performance of EFL learners in independent and integrated writing tasks. They also tried to find out if writing performance varies with task types. Thirty Iranian male and female advanced EFL learners participated in their study. The students were asked to write on two writing tasks of IELTS Academic module. Three experienced raters scored the writing samples. The result of the two one-way ANOVAs revealed the difference between the raters in rating integrated writing task while they were not performed differently in rating independent task. Paired sample t-tests also indicated that the learners' performance did not differ significantly across the two related tasks. In another study, Soleimani and Mahdavi-pour (2014) scrutinized the impacts of variations in integrated writing tasks and proficiency level on features of written discourse. Through the purposive sampling method, 60 EFL students were selected from translation and literature majors through some criteria. The participants were then divided into low proficiency and high proficiency levels. The instruments in the study were: test of English proficiency and writing tests. Two independent raters who hold M.A. degree in TEFL conducted scoring and text analyses. The findings revealed differences in the qualities of writing that emerged across the two task types with respect to text length and direct use of source material(s) without quotation marks.

In a recent study, Darong (2021) investigated the impact of the integrated task on the quality of students' writing at college level. Twenty - two learners participated in his study. Twelve students were asked to write an essay following an integrated task while the other 12 wrote based on an independent task. Next, the task division was switched in the second writing session. Three raters scored the writing quality. The quality was also descriptively and inferentially analyzed using Wilcoxon signed rank test. The findings revealed that the two groups were not significantly different. The author believes that there are other aspects contributing to writing quality and merely task types do not guarantee the students' writing quality.

3. Research Method

3.1. Research Design

This study was a descriptive, statistical investigation which relied on empirically collected data and tried to conduct a regression analysis. The analysis tried to find out how a number of independent variables could predict a dependent variable. Following the framework suggested by Uludag et al. (2019), this study focused on three measures as the independent (predictor) variables: source idea, linguistic modification, content accuracy. These variables sought to find out how much variation could be accounted for in the writing scores. The three predictor variables were measured as interval variables and no categorical item (e.g., gender) was examined.

3.2. Materials and Instruments

The integrated test for the present study involved three interconnected sections: reading, listening and writing. The test materials were channeled directly from TOEFL Preparation, Practice and Online Courses website on which there are TOEFL-based tests for the purpose of integrated writing.

3.3. Participants

The data were gathered from the statistical population including MA-level EFL students from Islamic Azad University Shiraz Branch, Islamic Azad University Fasa Branch, and Payame Noor University Shiraz Branch based on accessibility. The sample (of participants) in this study were 30 MA-level EFL students who were selected based on the purposive sampling method. Participants ranged in age from 22-37 years. The students were selected according to some inclusion criteria to make sure

they would not experience the problems associated with integrated writing tasks (see Uludag et al., 2019): (a) they should have obtained their BA in a language related discipline (e.g. translation, literature); (b) they should have been second-year students (who were starting to focus on research and thesis projects).

3.4. Data Collection

The data were collected in individual sessions during which a participant was given 70 minutes to read and listen to the materials and finish the essay part. Each session was held in the library of each university during which each participant read and listened to the materials via a laptop computer. However, they had the option of writing the essay either on the computer (in a notepad file without any smart edition options) or in handwriting. In each case, the listening track was played on the laptop and the participant used a standard headset for this section. The test sections were channeled on the computer from TOEFL Preparation, Practice and Online Courses website on which there are TOEFL-based tests for the purpose of integrated writing. The test had a "Read a short passage" section and a "Listen to part of a lecture on the same topic." Following those parts, the participants needed to write the essay. Each essay was expected to be around 250-300 words. The participants were advised to take notes during the listening sections so that they could more easily conduct the writing section.

3.5. Data Analysis Procedure

The writing task the students had to deal with after the reading and listening sections was:

Writing Task: Summarize the point made in the lecture and explain how the speaker cast doubt on specific points made in the reading passage.

After the data were collected, the instances were analyzed according to the three categories of source text use in integrated writing assessment. Primarily a qualitative sample of the cases was collected from the writings. Next, statistical tests were conducted to find the impact of the predictor variables that might account for the variance in the writing skill scores. To see the texts used for integrated tests, see Appendix A (section 1 and section 2).

Each of these variables was measured (interval scale) as follows: (a) source idea: this variable included any instances in which a source text idea was used in the writings; such instances mostly appeared as examples, fact, and ideas; (b) substantial linguistic modification: this category sought to find how much variation was implemented on a source text piece by focusing on acceptable lexical and syntactic modes of paraphrasing; and (c) content accuracy: this measure tried to examine how precisely the information in the source text was reflected in the writing.

The scores observed were submitted to three TEFL professors whose scores were subject to inter-rater reliability through averaging and rounding each score. The scores were based on the scale used by Uludaga et al. (2019), which involved seven levels, including "Low Beginner" (1-2), "High Beginner" (3), "Intermediate" (4), "High-intermediate" (5), "Advanced" (6), "Adept" (7), and "Expert-Fluent" (8-9). To find how much the independent variables explained the variance in the dependent one, this study relied on multiple regression test (see Ross & Willson, 2017, p. 47). In a multiple regression test, categorical (group numbers), interval (frequencies), and binary (e.g. 1 or 2) variables could be used. Because this study did not address any group-based tests, no categorical set or binary set was under consideration. Before conducting the final calculations, all of the assumptions of multiple regression were tested and confirmed.

In this multiple regression test, an ANOVA table was created which showed the degrees of freedom, f-statistic, and p-value. There was also a coefficients table which clarified any significant impact of the independent variables on the dependent one. Levels of significance were decided with reference to the p-value ($p < .05$). The data collected were processed in SPSS.

4. Findings and Discussion

4.1. Findings

4.1.1. Demographic information

The first piece of information reported in this section is the demographic specifications of the participants (see Table 1).

Table 1: The Demographic Information of the Participants in the Present Study

Age		Gender	
24-26	16	53.3%	Male
27-29	9	30.0%	Female
30-32	5	16.7%	9
			30%
			21
			70%

As Table 1 shows, the total number of the participants was 30, although an equal gender distribution was not a question in this study because gender was not seen as a predictor in this study. Furthermore, the majority of the students who took part in the study were aged 24-26. Only 5 participants were aged 30 or beyond.

4.1.2. Answering the first research question

Research Question One: What are some instances of source idea, linguistic modification, and content accuracy as statistically measured in the writing sections produced by the sample of the Iranian MA EFL students in the integrated tests?

This research followed both qualitative and quantitative purposes. This section primarily reports a sample of the responses as linguistic instances produced by the participants. These instances represent all of the three independent (predictor) variables, source text idea, linguistic modification, and content accuracy, as explored in this study. See Appendix B (Table 2, 3 and 4).

4.1.3. Answering the second research question

Research Question Two: Can source idea, linguistic modification, and content accuracy significantly predict writing quality in the sample of the Iranian MA EFL students in the integrated tests?

In order to answer the above question, it was primarily restated as a hypothesis to be tested through the statistical procedure employed in this study (multiple regression test):

H₀: Source idea, linguistic modification, and content accuracy cannot significantly predict writing quality in the sample of the Iranian MA EFL students in the integrated tests.

The question in the TOEFL writing section, following the reading and listening sections, asked the test-taker to summarize the points and explain what the problem was in the two passages. The two passages contained two opposing views regarding the pollution decline of otters. To summarize the views, the students would have to make many modifications to the two passages and put them in a summary-report writing frame. In many cases, nonetheless, they tended to use the materials in the passages directly, especially the reading passage. Many details were omitted in the writings and some new sentences were used to combine the passages. To score the writings, the scale used by Uludaga et al. (2019) was used (see section 3.5 above).

Table 5: The Scores of Students on Source Text Idea Use, Linguistic Modification, and Content Accuracy

Student No.	Source text idea use	Linguistic modification (substantial)	Content accuracy	Writing score
1	7	9	9	7
2	8	8	8	6
3	9	6	6	5
4	8	7	7	6
5	10	6	8	5
6	7	6	8	5
7	6	8	6	6
8	9	8	7	5
9	8	7	8	4
10	10	6	9	5
11	6	8	7	5
12	9	9	7	7
13	5	9	6	7
14	10	10	6	6
15	8	9	7	6
16	7	9	5	4
17	6	6	7	5
18	8	8	8	6
19	5	9	4	6
20	7	8	8	6
21	7	8	6	5
22	8	10	7	6
23	6	7	5	6
24	9	11	8	5
25	7	6	6	5
26	8	6	5	5
27	7	4	7	4
28	9	9	8	6
29	7	9	4	6
30	8	6	6	6
Total	229	185	203	

As Table 5 reveals, the students linguistically modified the sentences they used in 185 cases. "Source idea use" was found in 229 cases, whereas correct and acceptable propositions or ideas were observed in 203 cases. In order to conduct the multiple regression test, several other assumptions should be considered. The data used in this study were all interval values, as a result of which the type data was appropriate for the test. Meanwhile, the relationship between the dependent variable and the three other variables had to be linear. Figure 1 illustrates the linearity test of all of the variables.

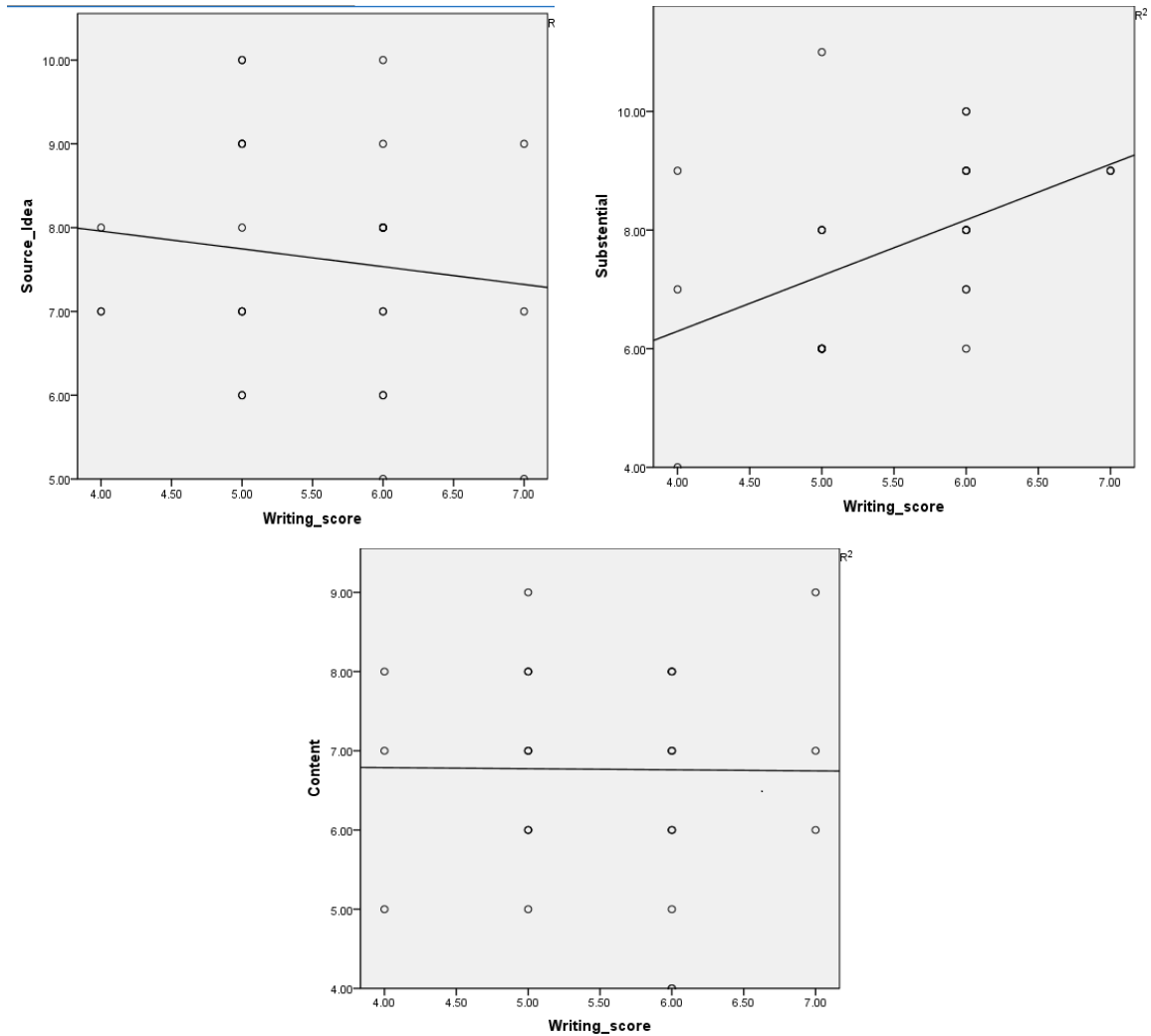


Figure 1: The Test of Linearity of the Variables Under Investigation

As Figure 1 most clearly shows, in all of the three cases individually, the relationship between the independent variables and the dependent one was linear. Next, to test multicollinearity, collinearity statistics was run, as shown in Table 6.

Table 6: The Collinearity Statistics of the Variables

Model	Collinearity statistics	
	Tolerance	VIF
Source idea	.804	1.245
Substantial revision	.994	1.006
Content accuracy	.801	1.249

As can be seen, the VIF values are less than 10 and the tolerance values are greater than 0.2, which demonstrates the acceptable collinearity measures in this study. The next assumption was to verify the independence of the residuals; to make sure of this issue, Durbin-Watson test was conducted, which was 1.767. The value of this test should be closer to 2, while not less than 1 or greater than 3. Considering these conditions, the independence of the residuals was verified. The next issue was homoscedasticity, or the constant variance of the residuals. To test this item in this study, the plot structure of the values was considered, which, as Figure 2 illustrates, shows randomly scattering dots without any particular shape pattern.

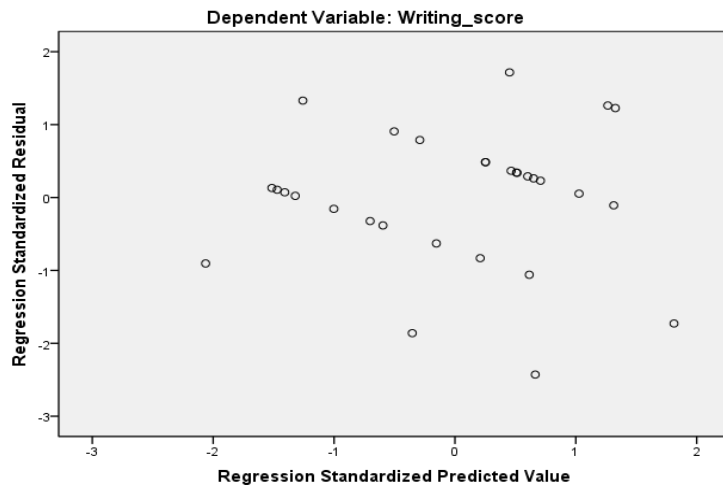


Figure 2: The Homoscedasticity Test

The next issue was to make sure of the normal distribution of the residuals. One of the methods of verifying this assumption in regression analysis is to use the P-P plot, in which the dots must be closer to the diagonal line. As Figure 3 depicts, in many cases the dots are close to the line, although there are some instances that show a relative distance from it.

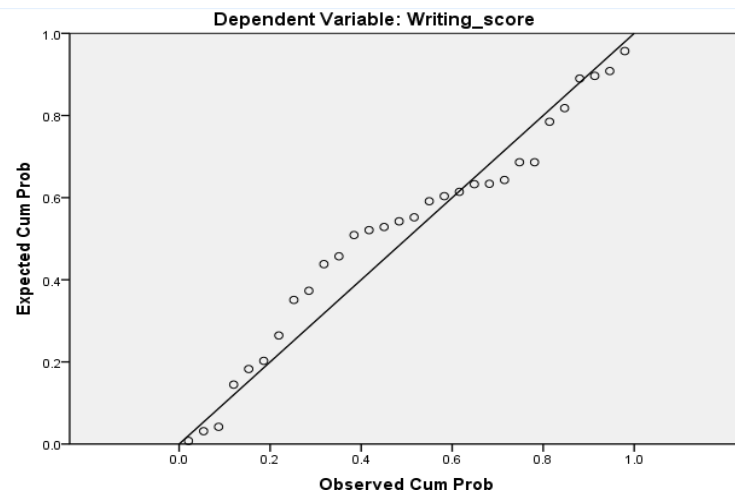


Figure 3: Normal Distribution of the Residuals

The last assumption was concerned with finding any data that could bias the results by showing values that were unduly effective in the results. In SPSS, the Cook’s distance has been installed that adds a new column to the results. In this column, values that exceed 1 could be eliminated as they could leave an unwanted impact on the final findings. In this study, the list of the residuals through Cook’s distance showed no case greater than 1. As a result, this final assumption for multiple regression was verified. Following all of the above assumption verifications, the multiple regression test was finally conducted (see Tables 7-9).

Table 7: The Model Summary in Multiple Regression

R	R square	Adjusted R Square	Std. Error of the Estimate
.508	.258	.173	.74517

Table 7 shows the model summary of the regression test conducted; what is mostly important is the R square value (coefficient of multiple determination), which shows the percent of the variance in the dependent variable explained solely or jointly by the independent variables. As such, the combination of the variables under examination (source idea, substantial modification, and content accuracy)

managed to account for 25.8% of the variance in source text use in integrated testing. Of course, this measure only shows the percent, and the ANOVA table must be referred to see whether the R square value was statistically significant (see Table 8).

Table 8: The ANOVA Table

Model	Sum of Squares	df	Mean square	F	Sig.
Regression	5.030	3	1.677	3.019	.048
Residual	14.437	26	.555		
Total	19.467	29			

As the Sig. value shows in Table 8, at the 95% level of confidence, the multiple regression was statistically significant ($p < 0.05$). This observation clearly revealed that the combination of “source idea”, “substantial modification”, and “content accuracy” could significantly predict writing quality (the score assigned to each writing task by the professors). This finding showed that the null hypothesis (H_0) was rejected and the relationship of the factors to the dependent variable was significant. Of course, although such a collective impact may be confirmed, there is still another piece of information to be considered regarding the degree and possible function of each of the variables in explaining the variable in the dependent variable (writing quality).

Table 9: The Coefficients in the Regression Test

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Source idea	-.107	.112	-.179	-.952	.350
Substantial revision	.252	.087	.491	2.898	.008
Content accuracy	.062	.116	.101	.538	.595

Table 9 provides statistical information regarding each of the independent variables. As can be seen, the significance values of “source idea”, “substantial revision”, and “content accuracy” are 0.350, 0.008, and 0.595, respectively, at the 95% level of confidence. Considering these measures, out of the three independent variables, only “substantial revision” was shown to be significantly effective in writing quality ($p < 0.05$). The other two, however, were not significantly impactful. The Beta measures, too, provide some information regarding the strength of each of the effect of each of the variables. More specifically, such values explain the impact of the independent variable for one standard deviation unit change one the dependent variable. As such, one standard deviation unit change in “substantial revision” could result in 0.49 units of change in writing quality.

4.2. Discussion

The purpose of this study was to investigate the writing productions and quality of 30 Iranian MA-level EFL students by exploring instance of source text use through integrated writing assessment. In doing so, the study first focused on the instances of each of the criteria of source text use as the predictors: “source idea”, “linguistic modification”, and “content accuracy.” The integrated test involved three interconnected sections: reading, listening and writing; the test material was channeled directly from TOEFL Preparation, Practice and Online Courses website on which there are TOEFL-based tests for the purpose of integrated writing.

The qualitative tables (Tables 2, 3 and 4) listed some of the choices made by the participants according to the three predictors under investigation. Although such choices may include some degree of imperfection as non-native writing, they at least show the patterns of MA students’ source text use, as potentially advanced and academic EFL learners. These instances of writing, of course, show one of the important contributions of this study, because Uludaga et al. (2019) did not provide any qualitative sample of their participants’ sample writings. The quantitative investigation used multiple regression analysis by considering all of the possible assumptions for this test before it was conducted (e.g., multicollinearity, collinearity, homoscedasticity).

The test revealed that the combination of the variables accounted for twenty five percent of the variance in writing quality through integrated testing. This suggested that, at least from the perspective of the professors (raters), source text use only explained *one-fourth* of the ultimate quality of the writing product. As a result, students taking integrated tests must develop the ability to flexibly handle several other factors in relation to source text use. Table 8 showed the ANOVA results, which clarified at the 95% level of confidence, the multiple regression was statistically significant ($p < 0.05$). As a result, the frame suggested by Uludaga et al. (2019) could help to explain source text use significantly in integrated testing.

Of course, this combinational effect would not necessarily mean that all of the factors were equally significant; in fact, in many multiple regression tests, several of the predictors are found insignificant in terms of quantitative data. Table 9 revealed that only “substantial revision” was statistically significant at the 95% level of confidence. This observation stood in exact opposition to what Uludaga et al. (2019) found. In their conclusion, Uludaga et al. (2019) stated that here was a positive association between band scores and source ideas, despite the low frequency of source ideas. They also claimed that the two aspects of source-text use, the proportion of accurate source ideas and the number of source ideas, predicted EAP students’ performance on a CAEL integrated writing task positively.

In this study, however, “substantial linguistic revision/modification” was the most considerable factor. Some reasons could account for this marked difference in the results found in this study and that of Uludaga et al. (2019). They report about their participants that in their first language (L1) backgrounds, there was substantial variation with the most frequent L1s including Arabic (20), French (16), and Spanish (12), Mandarin (43). They were registered in degree programs in Arts and Science (37), Business (50), Engineering and Computer Science (20) and Fine Arts (4).

As a result, in that study the participants were majored in such disciplines as Business, Art and Engineering, while in the present study MA students of EFL participated. This sample would perform differently from non-EFL students, particularly in the case of using linguistic patterns. Because “substantial revision/modification” is inherently a linguistic ability, the EFL participants seemed to have very adeptly implemented this factor. In fact, the students used a considerable number of the two other factors, “source ideas” ($n = 229$) and “content accuracy” ($n = 203$), but in multiple regression a case-by-case association of each score with the dependent variable scores was analyzed in software. In Uludaga and colleagues’ (2019) study, “substantial revision” was found in 101 cases, approximately half of the “source idea” variable measure ($n = 210$). Considering the advanced level of the Iranian MA-level EFL students in this study, their linguistic abilities outweighed their source text idea use and content accuracy.

Another issue that must be emphasized is the “task” to be addressed in an integrated test. The one used in this study was a summary-report type, which is different from compare-contrast or stance-taking tasks. In a summary, the discourse of the writing is considerably changed, as a result of which numerous linguistic modifications may be needed. In a compare-contrast frame, however, source ideas may be directly cited with a few original moves to be added by the test-taker.

Furthermore, creativity, as opposed to verbatim reproduction, is highly significant for today’s academia (Soleimani & Mahdavi-pour, 2014; Wette, 2018). The students in this research seemed to be sufficiently creative in re-designing the textual formulation, while using considerable “source ideas” and “accurate content.” This suggests that their reading abilities were good enough, even though no significant relationship was found regarding these two variables and writing quality prediction. This study did not explicitly measure writing quality dimensions such as clear purpose, audience engagement, clear point of view, single focus, logical organization, strong support, and so on (Whitaker, 2010, pp. 1-4). It was the responsibility of the professors to reflect their expectations in the scores they reported for each participant.

5. Conclusion and Implication

This investigation explored the writing quality of Iranian MA-level EFL students through the notion of source text use in integrated writing assessment. The study followed qualitative and quantitative purposes; the first question dealt with some instances of source text use including “source idea”, “linguistic modification”, and “content accuracy” in the writing samples. The second research question, however, addressed a quantitative issue, which tried to find out whether the three components could significantly predict writing quality in the sample of the Iranian students. A null hypothesis was formulated, which was tested via multiple regression.

The study concluded that the predicators could significantly explain one-fourth of the variance in the writing quality scores. Furthermore, out of the three factors of source text use, “substantial linguistic modification” was the only significant component, which was seemingly influenced by the participants’ advanced level and the nature of the writing task.

The qualitative findings of the study could help researchers to know how source ideas, linguistic modifications, and content accuracy may be used by participations in integrated writing-based tasks. Following assumption verifications for multiple regression analysis, the last three tables of the model summary, ANOVA, and the coefficients were constructed. Another important implication of the present study is that the writing tasks might be more effective than what it may seem on surface; in fact, different writing demands could lead to various responses. For instance, a summary test would require many changes and alterations that involve many linguistic modifications; other tasks may display different results. The statistical data also suggested that source text use in integrated testing could explain one-fourth of the score of TOEFL test. In fact, such writing tests require abilities in many other issues, such as style, syntax, coherence and cohesion.

Data collection, however, was very challenging in this integrated assessment; first, due to coronavirus restrictions, the universities were not fully active, and because of that few students would attend. Additionally, some potential participants said they were not ready for a listening or writing task. Furthermore, the library sections were only open on some days. All these factors made it difficult to gather data under standard test conditions. The original framework, too, could be ambiguous in some cases where source text idea use and linguistic modification may coexist. In this study, such cases were counted separately.

As for the suggestion for further research, the original framework could be analyzed in writings of other majors, such as the Humanities, Engineering, and Medical Sciences, in various Iranian universities. Also, this framework could be used in larger samples to verify the results here in other studies in Iran or abroad.

References

- Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, 10(1), 1–8. DOI: <https://doi.org/10.1080/15434303.2011.622016>.
- Dangal, G., Hamal, P. K., & Giri, M. (2017). Understanding research and scientific publication. *J Nepal Health Res Counc*, 15(35), 1-2.
- Darong, H.C. (2021). Integrated task on students’ writing quality: Is it more effective? *English Language Teaching Educational Journal*, 4(1), 25-33. doi.org/10.12928/eltej.v4i1.3336
- Derntl, M. (2014). Basics of research paper writing and publishing. *Int. J. Technology Enhanced Learning*, 6(2), 105-123.
- Gholami, J., & Alinasab, M. (2017). Source-based tasks in writing independent and integrated essays. *International Journal of Instruction*, 10(3), 127-142.
- Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4(1), 9–24. DOI: <https://doi.org/10.1002/tesj.65>.

- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15(4), 261–278. DOI: <https://doi.org/10.1016/j.jslw.2006.09.006>.
- Plakansa, L., & Gebrilb, A. (2016). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, 1-2. DOI: <http://dx.doi.org/10.1016/j.asw.2016.08.005>
- Ross, A., & Willson, V. L. (2017). *Basic and advanced statistical tests: Writing results sections and creating tables and figures*. Rotterdam & Boston: Sense Publishers.
- Sheibani, R., & Ahmadi, A. (2018). A comparison of writing performance on independent and integrated writing tasks. *International Journal of Foreign Language Teaching & Research*, 6, 18, 53–65
- Soleimani, H., & Mahdavi-pour, M. (2014). The effect of variations in integrated writing tasks and proficiency level on features of written discourse generated by Iranian EFL learners. *The Journal of Teaching Language Skills*, 75(4), 131-159.
- TOEFL Preparation, Practice and Online Courses. Retrieved January, 02, 2020, from www.bestmytest.com
- Uludaga, P., Lindberga, R., McDonougha, K., Payantb, C. (2019). Exploring L2 writers' source-text use in an integrated writing assessment. *Journal of Second Language Writing*, 46, 1-7. DOI: <https://doi.org/10.1016/j.jslw.2019.100670>
- Wette, R. (2018). Source-based writing in a health sciences essay: Year 1 students' perceptions, abilities and strategies. *Journal of English for Academic Purposes*, 36, 61–75. DOI: <https://doi.org/10.1016/j.jeap.2018.09.006>.
- Whitaker, A. (2010). *Academic writing guide: A step-by-step guide to writing academic papers*. Seattle: Seattle City University.
- Yang, S.C., & Plakans, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21, 118–133. DOI: doi:10.1016/j.jslw.2012.03.004

Appendix A: Section 1: Reading Section

Sea otters are a small mammal that lives in the waters along North America's west coast from California to Alaska. A few years ago some of the sea otter populations off of the Alaskan coast started to decline rapidly and raised several concerns because of their important role in the coastal ecosystem. Experts began investigating and came up with two possible explanations. One explanation was environmental pollution and the second was attacks by predators.

At first it seemed as if the pollution was the most likely cause for the population decline. One reason pollution was more likely was because of the known pollution sources along the Alaskan coast such as oil rigs. Also water samples taken in the area showed increased levels of chemicals that could decrease the otters' immune systems and indirectly result in their deaths.

Another thing that pointed to pollution as the culprit was the decline of other sea mammals such as seals in the same areas. This indicated that whatever was affecting the otters was also affecting the other sea mammals. Environmental pollution usually affects an entire ecosystem instead of just one species. Only predators that occurred in a large area, such as orcas (a large predatory whale), could cause the same effect, but they usually hunt larger prey.

Finally, scientists believed the pollution hypothesis would also explain the uneven pattern of otter decline. In some Alaskan locations the otter population declined greatly while other populations remained stable. Some experts suggested this could be explained by ocean currents, or other environmental factor, might have created uneven concentrations of pollutants along the coast.

Section 2: Listening Section

Professor:

Ongoing investigations have shown that predation is the most likely cause of the sea otter decline after all.

First, there is a lack of dead sea otters washing up on Alaskan beaches which is not what you would expect from infections caused by pollution killing them off. However, the fact that dead otters are seldom found supports the predator hypothesis. If a predator kills an otter, then it is eaten right away and can't wash up on shore.

Second, orcas prefer to hunt whales, but whales have essentially disappeared in that area from human hunters. Since the whales are not as available, orcas have had to change their diet to what is available. Since it is only smaller sea animals available, the orcas have probably started hunting those more and created the decline in all of the species mentioned in the passage.

Finally, predation is a more likely reason for the uneven pattern of otter decline. Otters in locations that are more accessible to orcas are more likely to be hunted. This is supported by the stable populations of otters in shallow rocky locations where orcas can't access.

Appendix B: Table 2: Samples of Source Text Idea as Used in the Writings of the MA-Level Students

Source text idea	Student source text idea use
A few years ago some of the sea otter populations off of the Alaskan coast started to decline rapidly and raised several concerns because of their important role in the coastal ecosystem.	Then, scientists started to feel worried about their population decline.
	..., as such animal had significant functions in the environment.
	Otters had some considerable impact on the ecosystem.
One explanation was environmental pollution and the second was attacks by predators.	Two possibilities were thought to be related to the otters' population decrease.
	The scientists arrived at a conclusion which said what had caused the problem.
	..., and after some time they observed that pollution and predators had leading to the population decline of the otters.
	The situation was caused by two factors apparently: environmental pollution and other hunting animals.
One reason pollution was more likely was because of the known pollution sources along the Alaskan coast such as oil rigs. Also water samples taken in the area showed increased levels of chemicals that could decrease the otters' immune systems and indirectly result in their deaths	It was assumed that the oil rigs at first led to the pollution in the area.
	There were, like oil, other substances in the water that could have been used by the otters.
	Afterwards, the scientists found that the animal's might have had their immune system weakened by the chemicals in the waster as well as the leakage from oil rigs.

	Thus in many case the sea creatures suffered from chemicals and they died.
Finally, scientists believed the pollution hypothesis would also explain the uneven pattern of otter decline. In some Alaskan locations the otter population declined greatly while other populations remained stable	They used some geographical evidences that showed pollution was the main factor.
	There was also another possibility that demonstrated that they perished through pollution.
	..., causing them to die as a result of pollution although this was not the case in other areas.... yet the otters were not dying as much in other regions which made a stronger probability that they been killed by substances in the water they had been consuming.
Ongoing investigations have shown that predation is the most likely cause of the sea otter decline after all. First, there is a lack of dead sea otters washing up on Alaskan beaches which is not what you would expect from infections caused by pollution killing them off.	The professor says other cause could lead to the death of the otters, which is predation.
	..., and based on what was found later they came to believe hunting was the major cause of their deaths.
	If the bodies of the otters were not found on the beach, so how did they die by infection, as asked by the professor.
Second, orcas prefer to hunt whales, but whales have essentially disappeared in that area from human hunters. Since the whales are not as available, orcas have had to change their diet to what is available.	There were living in the region and they did not have other sources of food.
	To feed, orcas decided to rely on the available otters, which is why other scientists think predation was the cause of the deaths.
	Secondly, orcas seemed to hunt otters instead of whales, because the latter was not living in the areas anymore

Table 3: Samples of Linguistic Modification in the Writings of the MA-Level Students

Source text idea	Student linguistic modifications
... the pollution was the most likely cause for the population decline	The decrease in population was accounted for by pollution
	... pollution caused he population reduction
	... because apparently pollution led to population decline
	... which was a condition causing the observed decrease in the population of the otters.

Finally, scientists believed the pollution hypothesis would also explain the uneven pattern of otter decline	The third observation of scientists was that pollution too had an impact on the uneven pattern of the animals' population reduction.
Environmental pollution usually affects an entire ecosystem instead of just one species.	In an environment even other creatures are hurt and not just one type.
	... because a chemical may reach various animals in a whole ecosystem.
	Some argue when pollution occurs more than just one species is at risk and in the case of otters
Some experts suggested this could be explained by ocean currents, or other environmental factor, might have created uneven concentrations of pollutants along the coast	Again experts insisted that environmental elements, including ocean currents, could account for the issue at hand.
	Some other scientists that a factor like ocean currents could have disturbed the distribution of polluting materials across the otters population.
	..., as a result of which an environmental cause could have happened that unequally influenced the population at various levels.
Ongoing investigations have shown that predation is the most likely cause of the sea otter decline after all.	In contrast to the above claims, there were other analyses that demonstrated the impact of predation and not pollution.
	They found through some continuous surveys that hunting by orcas was the major cause of the population decline of otters.
	However, as the professors explains some investigations revealed the outcome as caused by predatory animals.
... orcas prefer to hunt whales, but whales have essentially disappeared in that area from human hunters.	Because there were not many whales inhabiting the area anymore due to human hunting of them.
	..., and this investigation clarified that whales were a short supply as opposed to the past.

Table 4 Samples of Content Accuracy in the Writings of the MA-Level Students

Source content	Student content accuracy
A few years ago some of the sea otter populations off of the Alaskan coast started to decline rapidly and raised several concerns because of their important role in the coastal ecosystem.	The falling rate of otters' population has been a source of concern.
	The reason such creature are important is their effect on coastal ecosystem

<p>One reason pollution was more likely was because of the known pollution sources along the Alaskan coast such as oil rigs</p>	<p>As the professor states, however, predation caused the population decrease not the pollution sources identified in the Alaskan coast.</p>
<p><i>[The whole reading passage]</i></p>	<p>There were some pieces of evidence that pollution was responsible for the issue; first,</p>
<p>Only predators that occurred in a large area, such as orcas (a large predatory whale), could cause the same effect, but they usually hunt larger prey</p>	<p>..., although according to the professor's opinion these pollution sources cannot be attributed.</p>
<p>This is supported by the stable populations of otters in shallow rocky locations where orcas can't access.</p>	<p>but it was later learned than in this case orcas hunted smaller animals, otters.</p>
<p>However, the fact that dead otters are seldom found supports the predator hypothesis. If a predator kills an otter, then it is eaten right away and can't wash up on shore.</p>	<p>Contrary to the listening section information, orcas used otters because whales were not available anymore even though otters are not as large as whales.</p>
<p></p>	<p>For this reason, it was concluded that the location depth was effective</p>
<p></p>	<p>The pollution idea was rejected because dead bodies of otters were not seen much on the shore.</p>
<p></p>	<p>One of the reasons mentioned by the professor is that orcas ate the otters immediately and no carcass was found up on the shore.</p>