The Effect of Variations in Integrated Writing Tasks and Proficiency Level on Features of Written Discourse Generated by Iranian EFL Learners

H. Soleimani [*]
Assistant Professor, TEFL
Payame Noor University
email: arshia.soleimani@gmail.com

M. Mahdavipour
M.A., TEFL
Payame Noor University
email: maryam.mahdavipur.2009@gmail.com

Abstract

In recent years, a number of large-scale writing assessments (e.g., TOEFL iBT) have employed integrated writing tests to measure test takers' academic writing ability. Using a quantitative method, the current study examined how written textual features and use of source material(s) varied across two types of text-based integrated writing tasks (i.e., listening-to-write vs. reading-to-write) and two levels of language proficiency (i.e., high vs. low). Sixty Iranian English major students were selected through purposive sampling and divided into low and high proficiency groups based on an IELTS practice test. Then, they were required to write on a listening-to-write and a reading-to-write task. Results of two-way and one-way *ANOVAs* revealed that firstly, variations in integrated writing tasks together with level of proficiency had a significant effect on all the generated discourse features, secondly, the two types of integrated tasks produced features sharing the same features, and thirdly, some features could distinguish a certain level of proficiency. In addition, the results indicated that plagiarism (i.e., direct source use without quotation) is higher in response to the reading-to-write task than the listening-to-write task especially among the low proficiency writers. Implications of the study are presented.

Keywords: academic writing, integrated writing tasks, writing assessment, reading-to-write, listening-to-write

1. Introduction

Academic writing ability has been particularly regarded as one of the fundamental facets of language ability for successful academic achievement. Typical academic tasks at college level require students to work on several sources from outside rather than isolated information (Weigle, 2004). In this sense, integrated writing tasks characterize the type of writing that is expected in academic contexts of the tertiary level of education. Since the incorporation of direct writing tasks in L2 writing assessment, the independent tasks have been widely employed for evaluating writing skills of the university students (Gebril, 2009).

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) as they elicit actual writing performance rather than working on morphological and syntactic aspects of the target language (Camp, 1993) similar to what is expected in most of the indirect methods. Nevertheless, independent tasks have been criticized by many researchers (Gebril & Plakans, 2009; Plakans, 2007; Gebril, 2006; Weigle, 2002, 2004; Cho, 2003; Cumming, Kantor, Powers, Santos, & Taylor, 2000; Leki & Carson, 1997; Hamp-Lyons & Kroll, 1996). Given this criticism, integrated tasks have been regarded as an alternative component in writing tests.

Integrated writing tasks put forward an authentic measure for the writing skill (Cumming et al., 2000; Feak& Dobson, 1996; Guo, 2011; Read, 1990; Weigle, 2002, 2004; Yang, 2009) and as such have increasingly become a popular component in both large-scale writing and academic writing assessments (Gebril, 2009; Gebril & Plakans, 2009, Plakans & Gebril, 2012, Weigle & Parker, 2012). In recent years, the TOEFL iBT has included these tasks along with independent writing tasks in its writing section (Educational Testing Service, 2005). As Cumming, Kantor, Powers, Santos, and Taylor (2000) claim, authenticity is the most important justification for inclusion of integrated writing tasks in the new Test of English as a Foreign Language (TOEFL). These tasks measure the test takers' writing ability in academic settings and require test takers to exemplify "real-life English-language usage in university lectures, classes, and laboratories" (Educational Testing Service, 2007, p.6).

In typical academic contexts, language writing tasks often combine language skills including reading, listening, and writing. In activities such as summarizing, for instance, students work on reading material(s), interact

with ideas expressed by the author, and write the summary (Delaney, 2008). Thus, the motivation for inclusion of integrated writing tasks in the new generations of writing tests, according to Yang (2009), is that these tasks are "reflective of the real use of language that occurs in academic contexts" (p. 3). Lewkowicz (1997) believes that integrated tasks are intended to intimately resemble the language situations that students often experience in academic contexts (as cited in Gebril, 2009). Yang (2009) verifies that in academic contexts students in most of their writing tasks work with source material(s) to identify, synthesize, connect, and manipulate data in their writing. She adds that the integrated writing tasks are similar to real-life academic writing tasks in that both require test takers to combine multiple language skills.

In terms of potential consequences, it has been agreed that the authentic nature of integrated writing tasks leads to a positive washback effect (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Cumming et al., 2000; Esmaeili, 2002; Feak& Dobson, 1996; Weigle, 2004). That is, integrated writing tasks by asking test takers to produce authentic language similar to what they experience in real academic contexts encourage similarly authentic content in language curricula and teaching (Guo, 2011). Weigle (2004) believes that including such tasks in high stakes exams might make teachers and learners feel the need for skills that represent language usage in real academic writing contexts rather than relying solely on strategies for five paragraph writing.

Yang (2009) cites researchers (Lewkowicz, 1997; Wesche, 1987) who assume that authenticity of integrated writing tasks may improve predictive validity of writing assessment. Researchers have regarded fairness or accessibility as another advantage of integrated writing tasks (Yang, 2009). Source material(s) in integrated writing tasks are intended to support fairness of writing tests by minimizing topic effect (Read, 1990; Weir, 1993, as cited in Yang, 2009). In the writing or testing literature, topic effect is considered as one of the well-defined factors affecting writers' performance (Clapham, 1996). In independent writing tasks, the assigned topics may be unknown to some test takers and thus negatively affect their performance due to their lack of background knowledge (Guo, 2011). In contrast, integrated writing tasks by providing source material(s) can support test takers who may lack related knowledge or experience on the assigned topic (Reid, 1990; Wallace, 1997; Weigle, 2004; Weir, 1983, as cited in Guo, 2011). Read (1990) clarifies that:

> [Source material(s)] may help reduce the effects of differences in background knowledge among test-takers and, when the writing tasks are linked with earlier reading and listening tasks, may represent a better simulation of the process of academic study than simply giving a stand-alone writing test. (as cited in Lewkowics, 1994, p. 204)

Thus, according to Weigle (2002), providing students with source material(s) potentially diminish content bias in such tasks.

2. Background

It might be stated that the origin of integrated language testing in general dates back to the empirical studies by Oller (1979) to tackle the question of language proficiency. Oller (1979) examined whether language ability can be divided into separately testable constructs or not. He put forward three hypotheses regarding language proficiency construct. The divisibility hypothesis claims that language skills and components do not share common variances; on the contrary, the indivisibility hypothesis embraces that there is the same common variances are shared by all language skills and components. There still exists a partial divisibility hypothesis which takes the middle ground. The indivisibility hypothesis, also called unitary competence hypothesis, originally derives from cognitive science. Spearman (1904) believes that a general factor of intelligence dominates majority of the variance in human performance. From linguistics point of view, Spolsky (1968) argues that global proficiency tests are measures of linguistic competence rather than discrete skills.

In the recent decades again, lots of attention have been paid by researchers to integrated writing tasks to assess writing in academic contexts. Due to their increasing popularity and usefulness, we might be expecting of their replacing the independent writing tasks in assessing academic writing. Numerous writing researchers have referred to the merits of integrated writing tasks including their more authenticity (Weigle, 2004), lowering anxiety and creativity demands on writing (Plakans, 2008), and their positive washback effects in writing classrooms (Cumming, et al. 2004). Furthermore, Leki and Carson (1997) states that integrated writing tasks get the learners involved in reading-response writing and results in texts that are contend responsible. In a recently published article by Cumming (2013) on the perils and promises of integrated writing tasks for

academic purposes, several promises have been stated as emphasizing the positive effects and focusing more on complex activities such as discourse synthesis and textual borrowing as well as evaluating language abilities consistent with multiliteracies models of literacy.

Cumming, Kantor, Baba, Erdosy, Eouanzoui, and James (2005) in their study focused on differences in written discourse in independent and integrated writing tasks for next generation TOEFL. Text length, lexical sophistication, type-token ratio, syntactic complexity, grammatical accuracy, argument structure, and orientation to the source evidence were examined in detail across differing proficiency levels. For text level, for instance, they found a main effect for task type and a medium effect size for proficiency level. As far as lexical sophistication was concerned, a large effect size was found for task type and a small effect size was found for proficiency level. In their analysis of syntactic complexity, they found a medium effect size for task type and a large effect size for proficiency level. In argument structure, for the quality of propositions expressed, the results showed the independence essay tasks produced higher quality propositions for all three proficiency groups, than the two integrated tasks did.

Cho, Rijmen, and Novak (2013) examined the impact of prompt characteristics on the comparability of TOEFL iBT integrated writing tasks from 2005 to 2009. In the context of TOEFL iBT RLW tasks, the prompt includes a reading comprehension passage and a lecture. Evaluating 107 previously administered RLW prompts by subjects on nine measures of task difficulty through a questionnaire, they found that some of the variations in subjects RLW scores was because of differences in the English proficiency of testees that also varied across administrations. In addition, the results revealed that distinctness of ideas within the prompt and difficulty of ideas in the passage wee two variables attributing to the potential sources of variation in subjects' RLW scores.

In spite of advantages referred to in the literature concerning the superiority of integrated writing tasks over writing-only independent tasks for academic purposes, many scholars have been working on major concerns influencing the usefulness of the integrated writing tasks. Construct-related validity has been an issue of great concern to writing assessment researchers "to understand what scores from integrated tasks infer about English language writing ability" (Gebril & Plakans, 2009).

Some scholars state the risks related to the integration of writing with other skills in assessment, including mixing the measurement of writing with

comprehension of the source material, involving genres which are not well-defined and so are complex to score, misrepresenting the testees' language proficiency due to reliance on the language of the source text, and requiring threshold levels of abilities for competent performance (Cumming, 2013).

The existing literature puts forward a body of research which has quantitatively looked at the integrated writing tasks to support the validity arguments for them. "In the process of validation", as Yang (2009) states, "two types of data, writing products and process or strategies, are commonly collected" (p. 10). That is to say, the data derived from textual analysis on the essays written by test takers is required to support the inferences made from the test scores. This is in line with Chappell's (1999) assertion that the combination of validity evidence from different sources strengthens a validity argument. Also with reference to Bachman (2002), collection of information both "on test-takers' responses to individual assessment tasks," and "on processes or strategies that test-takers use in responding to assessment tasks," is required to shed light on the construct of integrated writing tasks (p. 470). In general, two lines of research can be detected in the literature which has validated integrated writing tasks. The first line has addressed the relationship between scores on integrated writing tasks and test takers' independent writing scores (Delaney, 2008; Esmaeili, 2002; Gebril, 2006; Lewkowicz, 1994; Messer, 1997; Watanabe, 2001), reading scores (Delaney, 2008; Enright, Bridgeman, & Cline, 2002; Trites&McGroarty, 2005; Watanabe, 2001), general language proficiency (Campbell, 1990; Connor &Krammer, 1995; Corbeil, 2000; Cumming, 1989; Delaney, 2008; Johns & Mayes, 1990; Kirkland & Saunders, 1990), as well as educational level (Conrad, 1996; Delaney, 2008; Guo, 2011; Mathison, 1996; Trites&McGroarty, 2005). The second line of research has aimed at understanding the relationship between generated textual features in integrated writing tasks and those in independent writing tasks (Campbell, 1990; Cumming et al., 2005, 2006; Guo, 2011; Lewkowicz, 1994; Watanabe, 2001). Furthermore, the interplay between textual features and integrated writing scores has been investigated with or without comparison to independent writing scores (Cumming et al., 2005, 2006; Gebril & Plakans, 2009; Guo, 2011; Johns & Mayes, 1990; Watanabe, 2001).

2.1 Objective of the study

Following the current writing research being conducted in the world, it seems a necessity for Iranian writing researchers to pursue writing-from-

sources tasks with a focus on reading and writing integration. In the same token, any research on integrated writing might find also try to find solutions to issues as the writers' use of sources in their performance, their proficiency level, their first language background, their gender, and so forth. The integrated view of writing leads to the holistic view to language in general and writing in particular, which results in the authenticity of writing tasks.

The present study was actually conducted with two major purposes: the theoretical and the pedagogical contribution. Theoretically speaking, we intended to examine whether some of the commonly referred to construct-related issues of integrated writing and its validation as claimed by authorities in writing are also confirmed employing data by Iranian EFL learners. From the pedagogy point of view, our research results might contribute to employing more emphasis on authentic writing tasks by practitioners in writing courses. Materials developers are also expected to use integrated writing tasks along with independent activities to meet the requirements of valid performances.

In comparison with the abundance of research on independent writing tasks, there are limited numbers of studies on integrated tasks in the literature of L2 writing assessment (Guo, 2011). Currently, little systematic evidence is available to describe the actual features of the written discourse that examinees produce in these new tasks. In what ways do the qualities of writing that examinees produce for reading-to-write tasks differ from those they generate for listening-to-write tasks? In fact, we could not find any study that exclusively examines how discourse features, proficiency level, and variations in integrated writing tasks interact for students in an EFL academic context. This work is an attempt to test the following hypotheses:

Null Hypothesis 1: Proficiency level and variations in integrated writing tasks do not have any significant effect on syntactic complexity of the written discourse in Iranian EFL learners' writing.

Null Hypothesis 2: Proficiency level and variations in integrated writing tasks do not have any significant effect on grammatical accuracy of the written discourse in Iranian EFL learners' writing.

Null Hypothesis 3: Proficiency level and variations in integrated writing tasks do not have any significant effect on text length of the written discourse in Iranian EFL learners' writing.

Null Hypothesis 4: Proficiency level and variations in integrated writing tasks do not have any significant effect on nature of source language use in Iranian EFL learners' writing.

3. Methodology

3.1 Participants

Using purposive sampling, 60 EFL Translation and Literature students from Tehran Payam-e-Noor University participated in the study. Several criteria were taken to sample the participants. First of all, they all had to be English majors. The reason was that, according to the pilot study, non-English majors with limited language proficiency, especially listening ability, tended to solely rely on the reading passages while ignoring the listening material(s). The second criterion was that to answer research hypotheses, the participants were to be divided into low proficiency and high proficiency levels. Therefore, considering the practicality issues, we decided to accept both junior (%45) and senior (%55) students. As for the third criterion, native language background and gender (based on the literature review) had to be controlled, so all the participants were selected from Iranian female students whose age ranged from 20 to 27. They were divided into two groups based on the dispersion of the IELTS practice test scores around the mean (4.99). Thirty two participants whose scores fell above the mean were assigned to high proficiency level group and 28 participants whose scores were below the mean were assigned to low proficiency level group.

3.2 Instruments

Test of English Proficiency: An academic practice version of the IELTS test was administered as a measure of English language proficiency. To make sure of its appropriateness, we found the claim that the test is "about the same level of difficulty as the real IELTS test" (University of Cambridge, 2000, p. 6) and evaluates all four language skills. Each of the four skills provides a band score, ranging from 0 (non-user) to 9 (expert user) and an average of the 4 scores yields an overall band score. The assumption was that both tests enjoy the same construct for writing competence.

    Writing Tests: The current study employed text-based integrated listening-to-write and reading-to-write tasks for both groups of test takers (The tests are accessible upon request). Text-based integrated writing tasks, of which TOEFL iBT integrated writing task is an example, involve construction of a text that summarizes or compares/contrasts information presented in source material(s) (Guo, 2011). In such tasks, since the writing is exclusively based on the information expressed in the source material(s), participants' performance is assumed not to be disadvantaged by a lack of

familiarity with the topic (Yang, 2009). That is to say, each participant has something to incorporate in writing apart from his/her individual, cultural, and educational backgrounds (Read, 1990). As the participants of this study are all university students majoring in English, authenticity of the English writing tasks entails a focus on academic contexts. Given that the writing tests of the new TOFEL focus primarily on academic contexts, we decided to adopt two reading-listening-writing tasks from the *BARRON'S TOEFL iBT Internet-Based Test* (Sharpe, 2010). For the purpose of the current research, the two selected tasks were modified into a reading-to-write and a listening-to-write task. To generate the listening-to-write task, the reading passages from the original task was recorded and presented orally. Also, some modifications were made to make sure that the texts were of equal level of difficulty. Hence, the administered tasks were not genuine intact TOEFL iBT tests. Table 1 shows the difficulty level of the texts.

Table 1. Difficulty level of the source texts

	Flesch-Kincaid Grade-Level	Flesch-Reading Ease	Sentences	words
Reading-to-write	11.4	51.189	21	490
Listening-to-write	11.9	52.981	21	488

As another type of writing task, the listening-to-write tasks consisted of two lectures on "disciplining children" which lasted about five minutes. The topic for the reading-to-write task was "system of school organization". The task included two short reading passages about 490 words long. The participants were given 6 minutes to read and comprehend the passages. Test takers were allowed to take notes during both tasks if they chose to. They were asked to write a response to a question asking them to summarize ideas and explain the relationship between ideas from the lectures for the listening-to-write task and from the reading passages for the reading-to-write task. An expected length of the essay was about 150 to 225 words.

3.3 Procedure

Four weeks before the main study, a pilot study was conducted with 30 participants to examine the instruments used and find out possible problems that would occur in the main study. The subjects were chosen randomly so that the results would be more reliable; however, they were not involved in any way in the main study to avoid replication effects. As for the main

study, the 60 participants received a full practice test of IELTS designed for academic candidates (University of Cambridge, 2000). Three days later, the participants were provided an information sheet describing the study and the procedures for the writing session. If participants agreed to continue, they were given the writing tasks. Verbal instructions were also given with the written prompt. It was told that they would have 35 and 30 minutes to complete the reading-to-write and listening-to-write tasks, respectively. To assure confidentiality, each writer was given an identification number; no names were included with their work. Then, the two integrative writing tasks were given to the test takers on a single session. We alternated the order of task presentation to minimize any effect that a certain order might have on score variability. To half of the participants, the reading-to-write task was given followed by the listening-to-write task. The rest of the students started with the listening-to-write task and then were given the reading-to-write task.

Two independent raters who hold M.A. degree in TEFL conducted scoring and text analyses. They were native speakers of Persian and both had more than a decade of experience teaching and assessing academic writing in English as a foreign language in both university and language institute contexts. Given their vast experience in this area, they did not need extensive training except for the training session in the piloting phase of the study. Although the IELTS program uses a single rating approach to assessing test takers' writing and speaking performance (Kim, 2010), in this study two raters scored the writing responses. The obtained coefficient displayed reasonable reliability indices of 0.73 and 0.75 for the integrated listening-to-write and integrated reading-to-write tests, respectively. 0.73 for the writing scores.

To address syntactic complexity, the essays were coded for the mean length of T-units. Grammatical accuracy was measured using a holistic rating of grammatical accuracy (from 1 to 3) adopted from Cumming, et al. (2005, 2006). Text length was operationalized as the total number of words per composition using Microsoft Word. In order to examine the effect of independent variables on direct use of source material(s), we adopted the indicators suggested by Gebril and Plakans (2009). That is, the raters coded direct source use as *without quotation marks* or *with quotation marks* to identify verbatim source use.

3.4 Data analysis

In order to test normality of the data, both for the IELTS and the integrated writing tests, the non-parametric test of Kolmogorov-Smirnov was run. In analyzing the collected data from the IELTS administration, descriptive statistics were calculated and an independent sample t-test was conducted. In order to address the results of textual analyses across the two integrated task types as well as the two proficiency groups, two-way and one-way Analysis of Variance (*ANOVA*) were run using the Statistical Package for Social Sciences (*SPSS*) 18.

4. Results

4.1 Normality of the data

The non-parametric test of Kolmogorov-Smirnov was run to see whether the participants' scores in the written features generated in response to the integrated listening-to-write and reading-to-write tests were normally distributed. As Tables 2 and 3 show, p values for words per T-unit, grammatical accuracy, text length, direct source use with quotation, and direct source use without quotation are greater than .05 for both groups in the listening-to-write test.

Table 2. One-sample Kolmogorov-Smirnov test on features generated by high-group in the listening-to-write task

		Words-per-T-unit	Grammatical-Accuracy	Text-Length	Source-Use-Quotation	Source-Use-No-Quotation
N		32	32	32	32	32
Normal Parameters a,b	*M*	15.0063	2.1641	194.9063	3.6875	5.1563
	SD	2.02595	.54480	9.90067	.96512	.95409
Most Extreme Differences	Absolute	.115	.120	.060	.231	.221
	Positive	.081	.120	.050	.231	.221
	Negative	-.115	-.109	-.060	-.163	-.155
Kolmogorov-Smirnov Z		.652	.680	.340	1.305	1.252
Asymp.Sig.(2-tailed)		.789	.745	1.000	.066	.087

Table 3. One-Sample Kolmogorov-Smirnov Test on Features
Generated by Low-Group in the Listening-to-Write Task

		Words-per-T-unit	Grammatical-Accuracy	Text-Length	Source-Use-Quotation	Source-Use-No-Quotation
N		28	28	28	28	28
Normal Parameters a, b	M	12.836	1.6964	172.8214	1.0714	2.1786
	SD	2.2316	.45824	11.91032	.76636	.77237
Most Extreme Differences	Absolute	.107	.130	.147	.216	.249
	Positive	.105	.130	.125	.216	.199
	Negative	-.107	-.103	-.147	-.213	-.249
Kolmogorov-Smirnov Z		.568	.689	.780	1.141	1.318
Asymp.Sig.(2-tailed)		.904	.730	.577	.148	.062

Similarly, as shown in Table 4 and Table 5, p values are greater than .05 for both groups in the reading-to-write test. Thus, distributions of scores for all the written features in the two tasks were normal.

Table 4. One-sample Kolmogorov-Smirnov test on features generated by high-group in the reading-to-write task

		Words-per-T-unit	Grammatical-Accuracy	Text-Length	Source-Use-Quotation	Source-Use-No-Quotation
N		32	32	32	32	32
Normal Parameters[a,b]	M	15.7156	2.2188	206.4688	4.5000	3.9375
	SD	2.14162	.49899	12.26320	.98374	.80071
Most Extreme Differences	Absolute	.100	.138	.100	.226	.223
	Positive	.087	.138	.079	.163	.223
	Negative	-.100	-.120	-.100	-.226	-.189
Kolmogorov-Smirnov Z		.565	.782	.568	1.276	1.261
Asymp. Sig. (2-tailed)		.907	.574	.903	.077	.083

Table 5. One-sample Kolmogorov-Smirnov test on features generated by low-group in the reading-to-write task

		Words-per-T-unit	Grammatic al-Accuracy	Text-Length	Source-Use-Quotation	Source-Use-No-Quotation
N		28	28	28	28	28
Normal Parameters[a],[b]	M	14.0000	1.6607	178.53 57	1.7500	6.2500
	SD	2.09178	.46255	12.509 20	1.14261	1.07583
Most Extreme Differences	Absolute	.112	.136	.171	.229	.235
	Positive	.112	.136	.171	.151	.235
	Negative	-.083	-.126	-.126	-.229	-.127
Kolmogorov-Smirnov Z		.594	.719	.905	1.214	1.242
Asymp. Sig. (2-tailed)		.872	.679	.386	.105	.091

4.2 Testing the hypotheses

The statistical procedure of two-way *ANOVA*s were run to explore possible effect of the independent variables, namely two types of integrated writing tasks and two levels of English proficiency, on the dependent variables including syntactic complexity, grammatical accuracy, text length, as well as direct use of source material(s). The results showed main effects with small effect sizes at the $p < .05$ level for task type in number of words per T-unit, F (1, 116) =5.825, p =.017 (Table 6). For task type, main effect with medium size was observed in text length, F (1, 116) =16.409, p =.000 (Table 8), and main effect with large effect size was reported in direct source use without quotation, F (1, 116) =83.169, p =.000 (Table 10). For proficiency level, the results of analyses yielded main effects with large effect sizes in the number of words per T-unit, F (1, 116) =25.057, p =.000 (Table 6), grammatical accuracy, F (1, 116) =32.122, p =.000 (Table 7), text length, F (1,116) =137.529, p =.000 (Table 8), and direct use of source material(s) with quotation, F (1, 116) =226.765, p =.000 (Table 9). The interaction effect between task type and proficiency level was statistically significant only for direct source use without quotations with a large size, F (1, 116) =289.032, p =.000 (Table 10). On the other hand, the analysis of grammatical accuracy, F (1,116) =.011, p =.917 (Table 7), yielded no significant difference. With respect to proficiency level, results revealed no statistically significant

difference in direct source use without quotation, $F (1,116) = 3.145$, $p = .079$ (Table 10). Accordingly, the

findings might be used to reject all the null research hypotheses which stated that proficiency level and variations in integrated writing tasks do not have any significant effect on syntactic complexity, grammatical accuracy, text length, and nature of using source material in the written discourse produced by Iranian EFL learners' writing.

Table 6. Tests of between-subjects effects dependent-variable: words-per-T-unit

Source	Type-III-Sum-of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	139.793a	3	46.59	10.35	.00	.21
Intercept	24736.14 1	1	24736.14	5496.6 5	.00	.97
Task-type	26.213	1	26.21	5.82	.01	.04
Proficiency-level	112.763	1	112.76	25.05	.00	.17
Task-type * proficiency-level	1.545	1	1.54	.34	.55	.00
Error	522.025	116	4.50			
Total	25732.57 0	120				
Corrected Total	661.818	119				

a. R Squared = .211(Adjusted-R-Squared = .191)

Table 7. Tests of between-subjects effects dependent-variable: grammatical-accuracy

Source	Type-III-Sum-of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.921[a]	3	2.640	10.797	.000	.218
Intercept	447.305	1	447.30 5	1829.187	.000	.940
Task-type	.003	1	.003	.011	.917	.000
Proficiency-level	7.855	1	7.855	32.122	.000	.217
Task-type * proficiency-level	.061	1	.061	.250	.618	.002
Error	28.366	116	.245			
Total	493.563	120				
Corrected Total	36.287	119				

R Squared = .218 (Adjusted R Squared = .198)

Table 8. Tests of between-subjects effects dependent-variable: text-length

Source	Type-III-Sum-of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	21276.208ᵃ	3	7092.06	52.21	.00	.57
Intercept	4230655.736	1	4230655.73	31147.72	.00	.99
Task-type	2228.705	1	2228.70	16.40	.00	.12
Proficiency-level	18680.002	1	18680.00	137.52	.00	.54
Task-type * proficiency-level	255.372	1	255.37	1.88	.17	.01
Error	15755.759	116	135.82			
Total	4324308.000	120				
Corrected Total	37031.967	119				

a.R Squared =.575(Adjusted-R-Squared =.564)

Table 9. Tests of between-subjects effects *dependent-variable: source-use-with-quotation*

Source	Type-III-Sum-of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	232.010ᵃ	3	77.337	81.568	.000	.678
Intercept	904.934	1	904.934	954.449	.000	.892
Task-type	16.601	1	16.601	17.509	.000	.131
Proficiency-level	215.001	1	215.001	226.765	.000	.662
Task-type * proficiency-level	.134	1	.134	.141	.708	.001
Error	109.982	116	.948			
Total	1311.000	120				
Corrected Total	341.992	119				

a.R Squared =.678(Adjusted-R-Squared =.670)

Table 10. Tests of between-subjects effects dependent-variable: source-use-no-quotation

Source	Type-III-Sum-of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	267.648ᵃ	3	89.21	118.74	.00	.75
Intercept	2303.029	1	2303.02	3065.35	.00	.96

Task-type	62.486	1	62.48	83.16	.00	.41
Proficiency-level	2.363	1	2.36	3.14	.07	.02
Task-type * proficiency-level	217.152	1	217.15	289.03	.00	.71
Error	87.152	116	.75			
Total	2678.000	120				
Corrected Total	354.800	119				

a.R Squared =.754(Adjusted-R-Squared =.748)

4.3 Task type and textual features

A one-way analysis of variance (*ANOVA*) was conducted to explore the impact of task type (i.e., reading-to-write and listening-to-write tasks) on syntactic complexity, grammatical accuracy, text length, and direct use of source material(s) with or without quotation marks. The findings indicated that in response to the reading-to-write task, the participants of both groups wrote longer essays, shorter clauses, with more instances of direct use of source material(s) with or without quotation marks (Table 14). However, only the observed differences in text length and direct use of source material(s) without quotation marks, respectively with medium and large effect sizes, were of practical significance.

As to the total number of words written, the difference was significant between the two task types, F (1,118) =7.962, p =.006 at the $p<.05$ level with a medium effect size, eta squared=.063 (Table 11). Table 12 confirms that mean score for the reading-to-write task (M = 193.433, SD=18.657) was higher than the listening-to-task (M =184.600, SD=15.487). On the measure of direct source use without quotation, there was a significant difference between task types, F (1,118) =17.975, p =.000 at the $p<$ .05 level. The actual difference between the means across the two tasks was also significant (M =3.766, SD=1.730 for L-to-R and M =5.016, SD=1.489 for R-to-W). The calculated effect size was .132 which represented a medium one.

Table 11. Descriptive statistics for the effect of task type on textual features

		N	*M*	SD	*Std.Error*
Word-per-T-unit	L-to-W	60	13.99	2.37	.30
	R-to-W	60	14.91	2.27	.29
	Total	120	14.45	2.35	.21

Grammatical-Accuracy	L-to-W	60	1.94	.55	.07
	R-to-W	60	1.95	.55	.07
	Total	120	1.95	.55	.05
Text-Length	L-to-W	60	184.60	15.48	1.99
	R-to-W	60	193.43	18.65	2.40
	Total	120	189.01	17.64	1.61
Source-Use-With-Quotation	L-to-W	60	2.46	1.57	.20
	R-to-W	60	3.21	1.73	.22
	Total	120	2.84	1.69	.15
Source-Use-No-Quotation	L-to-W	60	3.76	1.73	.22
	R-to-W	60	5.01	1.48	.19
	Total	120	4.39	1.72	.15

It was mentioned that the difference should be interpreted with respect to proficiency level since the results yielded an interaction effect for this variable. As displayed in Figure 1, there tended to be fewer verbatim phrases as participants' level of proficiency increased for the reading-to-write tasks, but more verbatim phrases as participants' proficiency increased for the listening-to-write tasks.
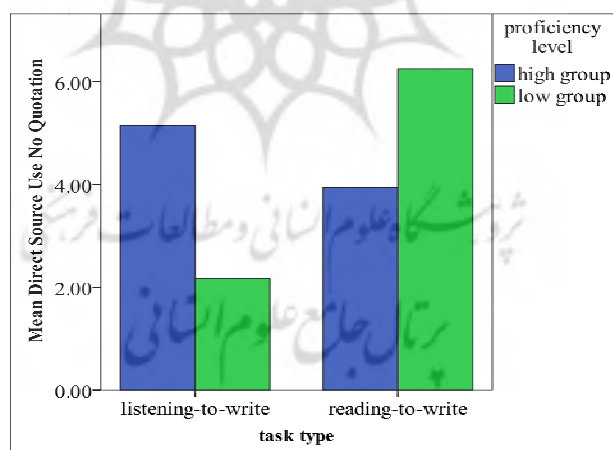

Figure 1. Interplay between task-type and proficiency-level for source-use-with-no-quotation

Explicitly, in the listening-to-write test, the high group borrowed more instances of the source material(s) while in response to the reading-to-write test, the low proficient group employed more strings of words directly from the source material(s).

4.4 Proficiency level and textual features

The results of one-way analysis of variance (*ANOVA*) indicated that across the two task types, the high proficiency group tended to compose longer clauses, longer essays, and more grammatically accurate texts, with more instances of direct use of material(s) with quotation marks. Text length and source use with quotation marks produced the largest effect sizes among all the textual features analyzed at this phase. On the measure of text length, there was a significant difference between proficiency levels, *F* (1,118) =120.109, *p* =.000 at the *p*<.05 level with a large effect size, eta squared= 0.504 (Table 13). As Table 12 displays, the actual difference between the means across the two levels was significant (*M* =175.67, *SD* =12.44 for Low-group and *M* =200.68, *SD* =12.49 for High-group). As to the source use with quotation marks, the difference between the two levels was significant, *F* (1,118) =199.778, *p* =.000 at the *p*<.05 level with a large effect size, eta squared=.628 (Table 13). As shown in Table 13, the difference between the means across the two levels was also significant (*M* =1.41, *SD* =1.02 for Low-group and *M* = 4.09, *SD* =1.04 for High-group).

Table 12. Descriptive statistics for the effect of proficiency level on textual features

		N	*M*	*SD*	*Std.Error*
Words-Per-T-unit	Low-group	56	13.41	2.22	.29
	High-group	64	15.36	2.09	.26
	Total	120	14.45	2.35	.21
Text-length	Low-group	56	175.67	12.44	1.66
	High-group	64	200.68	12.49	1.56
	Total	120	189.01	17.64	1.61
Grammatical-Accuracy	Low-group	56	1.67	.45	.06
	High-group	64	2.19	.51	.06
	Total	120	1.95	.55	.05
Source-Use-With-Quotation	Low-group	56	1.41	1.02	.13
	High-group	64	4.09	1.04	.13
	Total	120	2.84	1.69	.15
Source-Use-No-Quotation	Low-group	56	4.21	2.25	0.30
	High-group	64	4.54	1.06	0.13
	Total	120	4.39	1.72	0.15

Table 13: One-way-ANOVA-results-for-textual-features-across-proficiency levels

		Sum of Squares	df	Mean Square	F	Sig.	Eta squared
Words-Per-T-unit	Between-Groups	112.763	1	112.763	24.235	.000	.170
	Within-Groups	549.054	118	4.653			
	Total	661.818	119				
Text-length	Between-Groups	18680.002	1	18680.002	120.109	.000	0.504
	Within-Groups	18351.964	118	155.525			
	Total	37031.967	119				
Grammatical-Accuracy	Between-Groups	7.855	1	7.855	32.600	.000	.216
	Within-Groups	28.432	118	.241			
	Total	36.287	119				
Source-Use-With-Quotation	Between-Groups	215.001	1	215.001	199.778	.000	.628
	Within-Groups	126.991	118	1.076			
	Total	341.992	119				
Source-Use-No-Quotation	Between-Groups	3.304	1	3.304	1.110	.294	.009
	Within-Groups	351.288	118	2.977			
	Total	354.592	119				

5. Discussion

The results yielded 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. This is in agreement with Cumming et al. (2005, 2006) who found that the mean length of compositions was significantly higher in the reading-to-write tasks than the listening-to-write tasks. One possible justification is that test takers had visual access to the source texts in the reading-to-write test and could have employed more strings of words either with or without mentioning its sources. As a result, they produced longer essays in response to this task in comparison to the essays they generated for the listening-to-write task. However, considering that different time duration was assigned for the tasks, the evidence does not

directly lead to a definite explanation of the correlation between task type and the word count of the resultant essays. To be exact, text length in the listening-to-write test may have been limited by the time restriction both in the present research and that of Cumming et al. (2005, 2006). The findings indicated that the influence of task type on direct source use without quotation (i.e., plagiarism) depended on proficiency level. Albeit this type of plagiarism is not directly taken into account in the formal tests, it appears as an unavoidable habit applied by test takers to improve their test scores. The high proficiency group tended to employ more instances of the source material(s) in the listening-to-write test in comparison to the low proficiency group which borrowed more strings of words directly from the source material(s) in the reading-to-write task.

Similarly, Cumming et al. (2005, 2006) reported that test takers' uses of verbatim strings of words from the source material(s) were significantly higher in the listening-to-write tasks for proficiency level 5 (the most proficient). In opposition, this tendency was significantly higher in the reading-to-write task for proficiency level 3 (the least proficient groups). They attributed the differences to the interaction of a number of factors including test takers' proficiency levels, the medium of comprehension of source material(s), memory factors, as well as task characteristics and conditions. The possible explanations for the observed patterns in the present study, in line with Cumming et al. (2005, 2006), are medium of comprehension and memory factors. Degrees of comprehension may have been lower in the listening-to-write task than in the reading-to-write test notably for the low group. Namely, the less proficient writers may not have understood the listening tasks or the vocabulary sufficiently to have been able to use verbatim strings of words from those source material(s). In addition, for the listening-to-write task, test takers had to remind the source material(s) that they had heard or resort to their brief notes which they had taken during the listening phase. But for the reading-to-write task, they could read the passages as many times as they needed while responding to the task. It might be the case, for instance, that a less proficient participant who lacked enough linguistic resources to fulfill the requirements of the task might use parts of the source text directly in this task. Another probable reason may be related to testing conditions. Test takers were allocated less time, about 30 minutes, for the listening-to-write task whereas the time assigned to the reading-to-write task was 35 minutes.

The observed results point toward a significant rate of plagiarism: copying phrases directly from the material(s) without referring to its source by using quotations. It indicates that the participants of our research, especially the less proficient group, were not familiar enough with the conventions of academic writing although they were all English major students at university level. A number of researchers (e.g., Pennycook, 1996; Matalene, 1985) have noted that the inappropriate use of sources or plagiarism is related to the cultural aspects of the educational context in which L2 writers study (Gebril & Plakans, 2009). They believe that some students in non-native contexts do not consider plagiarism inappropriate. The same can be true for Iranian EFL students as they are provided with very few activities and assessments in which they are asked to write essays in response to integrated tasks.

In this research, the more proficient writers, in comparison to low proficient ones, generated longer clauses, longer essays, and more grammatically accurate texts, with more instances of direct use of material(s) with quotation marks. Text length and source use with quotation marks produced the largest effect sizes among all the textual features analyzed at this phase.

This is in agreement with Gebril and Plakans' (2009) findings in that text length increased as L2 writers' level of proficiency raised. Unlike this research, it was the only variable in their study which demonstrated significant differences across the three levels of proficiency. Cumming et al. (2005, 2006) also found that length of essays increased between proficiency levels 3 and 4 as well as between levels 3 and 5. Their findings revealed that the examinees wrote longer clauses and grammatically more accurate essays if they were more proficient in English. However, their findings did not yield a direct relationship between proficiency and the amount of using quotations.

One possible reason for the observed differences between our findings and those of previous researchers (e.g., Cumming et al., 2005, 2006; Gebril & Plakans, 2009) is that while they measured test takers' proficiency by writing scores, we distinguished low and high groups by scores derived from an English proficiency test. In addition, unlike mentioned studies which made use of thematically-related integrated tasks, this study measured integrated writing ability via text-based integrated tasks. The relationship between general language proficiency and integrated writing performance on both task types has been explored by Delaney (2008). She reported a

significant impact of language proficiency for thematically-related integrated tasks but not for text-based integrated tasks. Her findings yielded that these tasks represented different dimensions of reading-to-write ability since they were affected differently by language proficiency.

The observed interaction in our research points toward a distinction between thematically-related integrated task and text-based integrated task at least in terms of using source material(s) with quotations. While the more proficient writers, in response to text-based integrated writing task, tended to compose more instances of direct use of material(s) with quotation marks, the high proficiency groups in the earlier studies (e.g., Cumming et al., 2005, 2006; Gebril&Plakans, 2009) did not show such a tendency. Text-based integrated writing tasks, according to Guo (2011), require test takers to construct a text exclusively based on the information presented in the source material(s) whereas in response to thematically-related writing tasks, test takers are allowed to use their own ideas on the topic together with those expressed in the source material(s).

In view of that, the participants in our study heavily relied on the source material(s) in their writings rather than their personal opinions due to the nature of the text-based integrated writing tests. In consequence, the more proficient writers used more instances of the source material(s) with quotation marks rather than merely copying the texts. It can be due to the reason that the participants with higher levels of proficiency were more familiar with the conventions of academic writing, namely, using quotation marks with strings of words they directly borrowed from the source material(s). It verifies the findings of previous studies (e.g., Johns & Mayes, 1990) in which the less proficient L2 learners were reported to do more copying from the source texts. Yang (2009) also declared that effective test takers tended to avoid plagiarism by citing the sources in their essays.

Interpreting these results, alternatively, may suggest that some features could distinguish a certain level of proficiency across the two tasks. Weak performance on such variables as length of clauses and essays, grammatical accuracy, as well as direct use of material(s) with quotation marks was more characteristics of the participants with lower level of proficiency. That is, the participants' level of proficiency could differentiate linguistic performance in these tasks. The present study did not use scoring rubrics for the integrated writing performance, but it yielded valuable information which can be used for developing scoring schemes for assessment purposes.

According to Cumming et al. (2006), knowing this information is helpful in developing and refining current scoring rubrics for these tasks:

> … consideration also needs to be given to how the written discourse of examinees varies in particular tasks with their English proficiency. This information is needed to verify, or refine, the scoring schemes being developed to evaluate examinees' performance on these writing tasks. (p. 2)

6. Conclusion

As Guo (2011) asserts, a great deal of the discussion on the integrated tasks is based on theory rather than empirical research. Due to the small number of studies in this area, limited information is at hand about how linguistic features vary across text-based integrated listening-to-write and reading-to-write tasks. Our findings revealed that textual features generated by Iranian English major students in response to the reading-to-write and listening-to-write tests were similar to a large extent. This is in agreement with Cumming et al. (2005, 2006) who made a similar comparison between thematically-related integrated listening-to-write and reading-to-write tasks. By and large, these tasks can be considered as alternative types of integrated writing task that can be employed jointly to provide adequate evidence of writing ability. Integrated writing tasks, compared to independent ones, are regarded to offer more authenticity, improve fairness, and provide positive washback effect on learning and teaching of English (Yang, 2009). However, for assessment purposes, our concern is directed to the problem of high rate of plagiarism in the reading-to-write test among Iranian EFL students. This can be a threat to the reliability of the ratings and validity in interpreting test scores. Therefore, more validation studies are needed to shed light on the nature of integrated writing construct as well as to address issues of writing assessment concerning the use of source material(s).

Our findings could support using integrated tasks for writing assessment in Iranian universities. To the researchers' knowledge, these tasks in contrast to the independent ones have not normally been used for assessment purposes in the tertiary level of education in Iran. The tasks are believed to improve predictive validity as they resemble writing tasks similar to the ones students are involved in real-life academic contexts (Cumming et al., 2000).

The scores from integrated writing tests can offer useful information for stakeholders in recognizing how well students may perform when dealing

with real academic writing tasks throughout their education. Actually, for assessing L2 learners' performance on academic tasks, inference should be derived from integrated writing tests and not from independent reading and writing tasks (Delaney, 2008).

The use of integrated writing tests lead to positive washback effect (Weigle, 2004) as they require teachers to incorporate corresponding skills in their teaching programs. Further, by providing source material(s), the hindering effect that a lack of background information on a certain topic may have on students would be controlled. Presenting a common source for all students helps to achieve equity or fairness in testing (Plakans, 2007; Weir, 1993).

Concurrent use of independent and integrated writing tests has also been supported (Cumming et al., 2005, 2006) as it can yield a more comprehensive picture of the test takers' writing ability. Guo (2011) explains that essays written in response to the integrated tests represent characteristics of the general academic writing such as infrequent use of personal pronouns.

The results suggest the need for instructional materials to help Iranian EFL students properly recognize the requirements of the integrated tasks and involve them in practicing this ability in their classes. Teachers should not suppose that independent teaching of reading and writing to students would prepare them to cope with the requirements of integrated tasks (Delaney, 2008). The first challenge we may face is incorporating instructions on integrated writing into the current teaching materials and courses in our universities in general and Payam-e-Noor University in particular. Another challenge is devoting some instructional time to teaching integrated writing skill especially to the English major students.

References

Bachman, L.F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L.F. (2002). Some reflections on task-based language performance assessment. *Language Testing, 19*(4), 453−476.

Camp, R. (1993). Changing the model for the direct writing assessment. In M. M. Williamson & B. A. Huot (Eds.).*Validating holistic scoring for writing assessment: theoretical and empirical foundations* (pp. 45-78). Cresskill, NJ: Hampton Press, Inc.

Chapelle, C.A. (1999).Validity in language assessment. *Annual Review of Applied Linguistics, 19*, 254−272.

Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, *8*, 165–191.

Cho, Y., Rijmen, F., & Novak, J. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT integrated writing tasks, *Language Testing*, *30,* 4, 513-534.

Clapham, C. (1996).*The development of IELTS: A study in the effect of background knowledge on reading comprehension.* Studies in Language Testing 4. Cambridge: UCLES/Cambridge University Press.

Conrad, S. (1996). *Academic discourse in two disciplines: professional writing and student development in Biology and History*. Unpublished doctoral dissertation, Northern Arizona University.

Connor, U.M., & Kramer, M. G.(1995).Writing-from-sources: Case-studies of-graduate-students-in-business-management. In Diane-Belcher, George Braine-&-Diane-Dewhurst-Belcher (Eds.). *Academic-writing-in-a second-language* (pp. 155−182). Norwood, NJ: Ablex-Publishing Corporation.

Corbeil, G.(2000).Exploring-the-effects-of-first-and-second-language proficiency-on-summarizing-in-French-as-a-second-language. *Canadian Journal-of-Applied-Linguistics, 3*, 35-62.

Cumming, A.(1989). Writing expertise and second language proficiency. *Language Learning, 39*, 81−141.

Cumming, A. Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Langugae Testing, 21 2*, 107-145.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005).Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5−43.

Cumming, A., Kantor, R. Baba, K., Erdoosy, U., Eouanzoui, K., & James, M. (2006).*Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL* (TOEFL Monograph No. MS-30). Princeton, NJ: ETS.

Cumming, A., Kantor, R., Powers, D. Santos, T., & Taylor, C. (2000).*TOEFL 2000 writing framework: A working paper.* TOEFL

Monograph Series, Report No. 18. Princeton, NJ: Educational Testing Service.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004).A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 159–197.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly, 10, 1*, 1-8.

Delaney, Y.A. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes, 7*, 140-150.

Educational Testing Services.(2005). *Test of English as a foreign language*. Retrieved April 4, 2012, from www.toefl.org.

Educational Testing Service. (2007). TOEFL iBT Tips: How to prepare for the TOEFL iBT. Princeton, NJ. Retrieved June 7, 2012 from http://www.ets.org/toefl

Enright, M., Bridgeman, B., & Cline, F. (2002, April).Prototyping-a-test-design-for-a-new-TOEFL. In-D. Eignor-(chair),-*Research-in-support-of-the-development-of-new-TOEFL.   Symposium-conducted-at-the-meeting-of-the-National-Council-on-Measurement-in-Education.*New-Orleans.Retrieved-March-13,2010-from-http://ets.org/research/conferences/aera2002.html.

Esmaeili,   H.   (2002).   Integrated-reading-and-writing-tasks-and-ESL-students'   reading-and-writing-performance-in-an-English-language-test. *The Canadian-Modern-Language-Review, 58*(4), 599-622.

Feak, C. &Dobson, B. (1996).Building on the impromptu: a source-based academic writing assessment. *College ESL, 6*(1), 73-84.

Gebril, A. (2006). *Indpendent and Integrated academic writing tasks: A study in generalizability and test method.* Unpublished doctoral dissertation, The University of Iowa, Iowa City.

Gebril, A. (2009).Score generalizability of academic writing tasks: Does one test method fit it all? *Journal of Language Testing, 26(4),* 507–531.

Gebril, A. &Plakans, L. (2009).Investigating source use, discourse features, and process in integrated writing tests. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 7*, 47-84.

Guo, L. (2011). *Product and process in TOEFL iBT independent and integrated writing tasks: An investigation of construct validity.* Unpublished doctoral dissertation, Georgia State University.

Hamp-Lyons, L. & Kroll, B. (1996).Issues in ESL writing assessment: an overview. *College ESL, 6*(1), 52-72.

Johns, A., & Mayes, P. (1990).An analysis of summary protocols of university ESL students. *Applied Linguistics, 11*, 253-271.

Kirkland, M., & Saunders, M. (1990).Maximizing student performance in summary writing: managing cognitive load. *TESOL Quarterly, 25*, 105-121.

Kim, Y. (2010). *An Argument-Based Validity Inquiry into the Empirically-Derived Descriptor-Based Diagnostic (EDD) Assessment in ESL Academic Writing.* Unpublished doctoral dissertation. University of Toronto.

Leki, I. & Carson, J. (1997).Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31*, 36-69.

Lewkowicz, J. (1994).*Writing from sources: does source material help or hinder students' performance?* Paper presented at the Annual International Language in Education Conference, Hong Kong. [ERIC Document Reproduction Service No. ED386050].

Lewkowicz, J. (1997).The integrated testing of a second language. In C. Calpham (Ed.), *Encyclopedia of language and education*, Vol. 7: *Language testing and assessment* (pp. 121–130). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Matalene, C. (1985).Contrastive rhetoric: An American writing teacher in China. *College English*, *47*, 789–807.

Mathison, M. A. (1996).Writing the critique, a text about a text. *Written Communication, 13*, 314-354.

Messer, S. D. (1997).*Evaluating-ESL-written-summaries:-an-investigation-of-the-ESL-Integrated-Summary-Profile-(ISP)-as- a-measure-of-the-summary-writing-ability-of-ESL-students.* Unpublished-doctoral-dissertation, Florida-State-University. Tallahassee, FL.

Oller, J. W. Jr. (1979). *Language tests at school: A pragmatic approach.* London: Longman.

Pennycook, A. (1996).Borrowing others' words: Text, ownership, memory and plagiarism. *TESOL Quarterly*, 30, 201–230.

Plakans, L. (2007). *Second language writing and reading-to-write assessment tasks: a process study.* Unpublished doctoral dissertation.The University of Iowa.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13,* 111-129.

Plakans, L., &Gebril, A. (2012).A Close Investigation into Source Use in Integrated Second Language Writing Tasks. *Assessing Writing*, *17*(1), 18-34.

Read, J. (1990). Providing relevant content in an EAP writing test.*English for Specific Purposes, 9*, 109−121.

Reid, J. (1990). Responding-to-different-topic-types:-a-quantitative-analysis-from-a-contrastive-rhetoric-perspective. In-B. Kroll (Ed.).*Second Language-Writing:-research-insights-for-the- classroom* (pp. 191-210). Cambridge: Cambridge-University-Press.

Sharpe, P. J. (2010).*Barron's TOEFL iBT: Internet-based test*. Hauppauge, N.Y: Barrons Educational Series, Inc.

Spearman, C. (1904). General Intelligence objectively determined and measured. *The American Journal of Psychology, 15 (5)*, 201-292.

Spolsky, B. (1968). Language testing: The problem of validation. *TESOL Quarterly, 2 (2),* 88-94.

Trites, L. & McGroarty, M. (2005). Reading to learn and reading to integrate: new tasks for reading comprehension test? *Language Testing, 22*, 174-210.

University of Cambridge. Local Examinations Syndicate. (2000).*Cambridge IELTS 2: examination papers from the University of Cambridge Local Examinations Syndicate.* With answers ed. Cambridge: Cambridge University Press.

Watanabe, Y. (2001).*Read-to-write tasks for the assessment of second language academic writing skills: investigating text features and rater reactions.* Unpublished doctoral dissertation. University of Hawaii.

Wallace, C. (1997). IELTS: global implications of curriculum and materials design. *ELT Journal, 51*, 370-373.

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

Weigle, S.C. (2004).Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing, 9*, 27-55.

Weigle, S.C., & Parker, K. (2012).Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing, 21(2),* 118-133.

Weir, C.J. (1983).*Identifying the language needs of overseas students in tertiary education in the United Kingdom.* Unpublished PhD Thesis, University of London Institute of Education.

Weir, C.J. (1993).*Understanding and developing language tests*. Hemel, Hamstead, England: Prentice Hall.

Wesche, B. (1987).Second language performance testing: The Ontario test of ESL as an example. *Language Testing, 4*(1), 28−47.

Yang, H.C. (2009).*Exploring the complexity of second language writers' strategy use and performance on an integrated writing test through structural equation modeling and qualitative approaches.* Unpublished doctoral dissertation.The University of Texas at Austin.