

Investigating the Scope of Source Use and Discourse Features in Integrated Academic Writing among Iranian Intermediate EFL Learners

Mahdiyeh Mirshahvalad^{1*}, Keivan Mahmoodi², Faramarz Azizmalayeri²

1. Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran

2. Assistant Professor, English Department, Malayer Branch, Islamic Azad University, Malayer, Iran

* Corresponding Author: m.mirshahvalad@yahoo.com

Abstract – In this study the researchers tried to investigate the effect of source use and discourse features in the integrated academic writing among intermediate EFL learners of one of the English institutes. First the researchers chose 15 students of IC6 level according to their final scores of their last semester, and then they were instructed how to write essays in ten sessions and they were asked to read the passages in their student books and their work books and then write an essay accordingly. The researchers then collected them and chose some of them with the same topic. After that they were asked to answer a questionnaire that was based on the usage of these criteria in their essays. Finally, two raters rated their writings and reported their scores which were then analyzed both qualitatively and quantitatively. It follows from the results that the usage of these factors are different in the distinctive proficiency levels.

Keywords: source use, discourse features, integrated writing

1. INTRODUCTION

As integrated writing tasks appear in more assessments of writing for academic purposes, research on their validity and usefulness is needed to assist users in interpreting scores. This study investigated features of writing from integrated reading-writing tasks. Integrated reading-writing tasks are increasing in popularity and either replacing or complementing writing-only independent tasks used in assessing academic writing.

The integrated tasks are seen to have more authenticity (Feak & Dobson, 1996; Weigle 2002), and may provide test takers with content, lowering anxiety and creativity demands on writing (Plakans, 2008; Read, 1990). These benefits extend to potentially positive washback effects in academic writing classrooms. However, problems exist for these tasks, such as development, plagiarism, and construct related validity. In terms of validity, more is needed to understand what scores from integrated tasks imply about English language writing ability. One component of a validity argument is “explanation,” which requires evidence such as the connection between an expected score and academic language proficiency (Chapelle, Enright, & Jamieson, 2008).

Some research has looked at discourse features and process in writing assessment with relation to proficiency (Engber, 1995; Manchon et al., 2000; Sasaki, 2000); however, most studies addressed the relationship based on impromptu writing tasks. Although holistic scores on independent and integrated tasks have been found to correlate (Brown, Hilgers, &

Marsella, 1991; Gebril, 2006; Lewkowicz, 1994), the written products have been shown to have significantly different discourse features (Cumming, Kantor, Baba, Erdosy, Eouanzoui, & James, 2005, 2006). Thus, it may be inconclusive to base score interpretations for integrated tasks on research from independent writing tasks. To fill this gap, our research explored the connections between writers' scores on integrated test tasks and features of the written products in completing the tasks with a group of writers from Malayer. Not use the content from source texts in their essays.

Another line of research has focused on writers' verbatim source or plagiarism (Campbell, 1990; Currie, 1998; Deckert, 1993; Johns & Mayes, 1990), suggesting that there is a relation between proficiency and verbatim source use. Johns and Mayes (1990) studied the writing of nonnative writers at two proficiency levels on a summary task. The findings showed no significant difference between the two proficiency groups in their textual borrowing in terms of distortion or sentence replication. However, other studies have found that proficiency impacts source use (Campbell, 1990; Currie, 1998; Cumming et al., 2005).

2. STATEMENT OF THE PROBLEM

Since writing is one of the important skills and tasks to learn, researchers seek for a way to improve EFL learner's ability to write well-formed texts. To examine the scope of source use and discourse features in integrated academic writing among Iranian intermediate EFL learners, the researcher have done a survey. The researchers tried to teach article writing to the intermediate students of one of the Malayer institutes, and asked them to read the article in their books, and write an article like that and use it as a source, and then examined the three criteria mentioned here. There are other surveys done about this subject, but because of importance of this issue, the researchers assumes that it's necessary to do another one, but this time about Iranian intermediate English learners.

3. RESEARCH QUESTIONS

To build on previous research with integrated writing tasks, this study explored how discourse features and source text use interact with writing proficiency in reading-writing tasks for Iranian intermediate EFL learners. The following research questions guided this study of integrated tasks:

1. Does writing proficiency level have a relationship with the use of discourse features?
2. Does writing proficiency level have a relationship with the use of source texts and verbatim source use?

4. RESEARCH HYPOTHESES

Based on the objectives determined for the current research, the following research hypotheses were examined:

1. Writing proficiency level has a relationship with the use of discourse features.

2. Writing proficiency level has a relationship with the use of source texts and verbatim source use.

5. INSTRUMENTS

In order to teach the students the writing skill throughout the program, *Touchstone Book 2* was used, which was also the students' textbook of their English class. However, for the testing session, two reading passages about "texting and communication" also adopted from *Touchstone 2*, were used as the source.

The task for the study was developed to have some characteristics of authentic academic writing. Thus, an argumentative essay prompt and two short reading passages were selected from *Touchstone Book 2*, 2nd edition. The argumentative mode was selected based on the recommendations of a number of researchers (Plakans, 2008; Gebril, 2006; Melenhorst, 2006; Cumming et al., 2005). For example, Melenhorst (2006) argues that argumentative texts are approached successfully by students. The topic was addressed in this prompt, two topics were selected with the idea that they would hold students' interest, hold clear positions that writers can choose, and has possible source texts that included some solid evidence. Two passages were selected following Lewkowicz (1994) recommendations to use more than one text.

6. DESIGN AND PRECEDURE

The present study used an intact class so it was not possible to randomly assign subjects to the groups. The design consisted of three levels, no control group and arranged as quasi experimental as the three levels were interacting groups. However, we attempted to collect qualitative data to check the possibility of any improvement dimension which might be missing. Mixed methods have been selected for this study. By using more than one method within a research study, the researchers were able to obtain a more complete picture of human behavior and experience. Thus, it was a better way to be able to hasten the understanding and achievement of the research goals more quickly. The data collection also involves gathering both numeric information (e.g., on instruments) as well as text information (e.g., on questionnaires) so that the final database represents both quantitative and qualitative information.

Several testing sessions were held during the summer of 2015 in a classroom at the institute. Participants were first provided an information sheet describing the study, the procedures for the writing session, and information on their rights as participants. This information also was given verbally. Of course, the students were informed of the general nature of the test, that is, the researchers who was also the teacher of the class, informed the students that their scores would not affect their final scores and that they should rely on their own knowledge to complete the test. In addition the instructions to the participants were clear enough to understand what was expected of them.

Then the writing task was presented to the students. Verbal instructions were also given with the written prompt. Then they received the process questionnaire, which they were

asked to complete after they finished writing their essays. They had one hour to complete the task. Each writer was given an identification number; no names were included with their work to assure confidentiality.

Essays were rated using a holistic scale adopted from the integrated tasks, but revised for this study based on a rating session with the pilot data to incorporate the following changes: (a) clarifying development and organization in the rubric to match the prompt evaluation criteria, (b) providing instructions for raters on scoring essays with plagiarism or with no source use, and (c) removing any references to listening features. From the pilot rating, essays were chosen to represent each score as profile and training essays.

Two raters conducted the rating. Both were experienced ESL teachers who had also taught academic writing. They attended a rater training session conducted by the researchers. In the session, the study's purpose was discussed, and they were acquainted with the task and the rubric. Then the raters and the trainer rated one essay together followed by a discussion of scoring. The same process was followed for three more essays. Then raters scored five essays independently followed by discussion and calibration. This procedure was continued until both raters feel comfortable with the rating scale. Each essay was scored by both raters. No ratings were more than one level apart; if there were essays where the raters disagree, a third rating had been done by the researcher.

7. DATA ANALYSIS

The essays and questionnaires were analyzed to answer the two research questions. The data were analyzed both qualitatively and quantitatively. First, discourse features were analyzed in the essays, followed by a comparison across the three levels of writing proficiency. Then, essays were analyzed for source text use and again were compared across proficiency levels. Each of these steps will be described in more detail in this section.

Discourse Features

Writing research uses many different methods to identify discourse features (Hinkel, 2002). In this study, guidelines by Cumming et al. (2005: 8–9) were adopted which state that indicators for such analysis should meet the following requirements:

1. Include a range of discourse features including lexical, syntactical, rhetorical, and pragmatic characteristics
2. Be applied reliably and meaningfully
3. Possibly show differences between compositions scored at different scale levels

Several discourse features were chosen as a focus for the analysis: lexical sophistication, syntactic complexity, grammatical accuracy, and fluency. However, before analyzing these separate features, some general characteristics of the essays need tabulation: T-units, the number of sentences, the number of words, and characters. T-units were marked for all essays by two raters, while the other characteristics were calculated using a Microsoft Word feature.

Lexical Sophistication

This feature of discourse has been defined as the size of a writer's productive vocabulary based on the written essay (Wolf-Quintero, Inagaki, & Kim, 1998). Average word length was used to establish lexical sophistication based on prior research (Biber, 1988; Cumming et al., 2005; Engber, 1995; Frase, Faletti, Ginther, & Grant, 1999; Grant & Ginther, 2000).

Syntactic Complexity

This feature measures “the range of forms that surface in production and the degree of sophistication of such forms” (Ortega, 2003: 492). This feature particularly has had many different methods employed in previous research. In this study the Ortega's advice (2003) was followed to include several measures for this discourse feature and adapted methods used by Cumming et al. 2005; Henry, 1996; Homburg, 1984; Perkins, 1980; Tedick, 1990:

1. The mean number of T-units per sentences
2. Mean length of T-units

Grammatical Accuracy

This feature is a standard assessment of language proficiency and has gained increased interest in second language acquisition and L2 writing assessment (Polio, 1997). However, it remains somewhat elusive because of the difficulty in precisely measuring linguistic accuracy in a quantifiable way. Most of the quantitative measures used to judge linguistic accuracy, such as error counts (Fischer, 1984; Zhang, 1987; Carlisle, 1989; Kepner, 1991) and error count with classification (Bardovi-Harlig & Bofman, 1989; Chastain, 1990; Frantzen, 1995; Kobayashi & Rinnert, 1992; Kroll, 1990) have been problematic due to either the unreliability of coding schemes or not accounting for the error severity (Polio, 1997). For these reasons, this study used a holistic rating of grammatical accuracy adopted from Cumming et al. (2005), that was based on a scale used by Hamp-Lyons and Henning

(1991):

1. Many severe errors, often affecting comprehensibility
2. Some errors but comprehensible to a reader
3. Few errors, and comprehensibility seldom obscured for a reader

Two raters used this scale to rate all essays for grammatical accuracy.

Fluency

As is common in most studies of writing features, fluency was determined through word count, which has been successful in differentiating proficiency levels in many studies of second language writing (e.g., Cumming et al., 2005; Hirano, 1991; Larsen-Freeman, 1978; Tedick, 1990).

Source Text Use

To the second answer question, the use of source texts was analyzed in the essays. This variable required defining kinds of source text use and verbatim source use. Initially, a set of ten essays was selected and each T-unit was coded for summary, paraphrase, and direction quotation. A commercial Internet program called Turnitin.com was used to identify verbatim source use across all the essays. Based on this analysis, several decisions were made. First, it was that the distinction between summary and paraphrase was difficult for raters and often caused disagreement. As a solution, the two kinds of source use were combined as "indirect" source use and considered in contrast to direct use. There was some concern with source use, and it searched the entire Internet as well as its essay base to find verbatim source use.

To resolve this problem, the raters coded "direct" source use as "no quotation" or "with quotation" to identify verbatim source use. The raters defined direct use as others have done (Cummings et al., 2005), but counting strings of three or more words from the original as verbatim source use. Following these guidelines, raters retrained on a small set of essays to achieve an inter-rater reliability of coefficient, and then the raters coded the rest of the essays.

Statistical Analysis of Discourse Features and Source Use

In order to analyze the discourse features and source use, the researchers checked these characteristics of the student's writings:

1. Average word length
2. The mean number of T-units per sentences
3. The mean number of words per T-unit
4. Total number of words
5. Grammatical accuracy
6. Indirect source use
7. Direct source use with quotation marks
8. Direct source use no quotation marks
9. Total source use

8. RESULTS AND DISCUSSION

In this part, the researchers' main goal is to give the results for the qualitative data and quantitative data, to infer them and then to discuss the results. To do this, the researchers presented procedure results summarized in their observation and the accuracy was provided according to the rubric.

8.1. Quantitative Results

In order to do the quantitative analysis, that is, preparing the profiles of each of the participants for the repeated measures of his/her performance across the research data collection, a graphic representation is prepared as follows. In this representation each participant's performance across the three patterns of developments has been rated by the two raters. Finally, the mean of these performances is presented. Also, the students were divided into three levels of proficiency according to their final score.

Table 1. Students' ID numbers, their final scores and levels

Students' ID Numbers	Final Score	Level
1	84	2
2	80	2
3	81	2
4	61	3
5	95	1
6	70	3
7	85	2
8	93	1
9	70	3
10	96	1
11	70	3
12	82	2
13	70	3
14	90.5	1
15	93	1

Table 2. The names of groups and students' ID numbers

GROUP	STUDENTS' ID NUMBERS
1	5-8-10-14-15
2	1-2-3-7-12
3	4-6-9-11-13

Table 3. Rater one's and rater two's scores on the subjects' writings

Students' ID Numbers	RATER1	RATER2	MEAN
1	88	80	84
2	70	65	67.5
3	80	76	78
4	75	70	72.5
5	85	78	81.5
6	84	80	82
7	79	70	74.5
8	91	90	90.5
9	85	78	81.5
10	75	68	71.5
11	80	74	77
12	68	65	66.5
13	65	60	62.5
14	89	80	84.5
15	93	91	92

In order to answer the proposed research questions, the quantitative section is intended to provide further explorations into the nature of the collected data. This will consist of the following parts: the descriptive section and the inferential section which focuses on testing the hypotheses.

8.1.1. Descriptive Section

There were 15 students invited to the current study who participated in both the qualitative and the quantitative sections. Table 4 and 5 present the descriptive statistics.

Table 4. Descriptive statistics for Rater One

	Statistics	Standard Error
Mean	80.4667	2.20144
95% Confidence Interval for Mean	Lower Bound	75.7450
	Upper Bound	85.1883
5% Trimmed Mean	80.6296	
Median	80.0000	
Variance	72.695	
Std. Deviation	8.52615	
Minimum	65.00	
Maximum	93.00	
Range	28.00	
Interquartile Range	13.00	
Skewness	-.359	.580
Kurtosis	-.823	1.121

Table 5. Descriptive statistics for Rater Two

	Statistics	Standard Error
Mean	75.0000	2.28869
95% Confidence Interval for Mean	Lower Bound	70.0913
	Upper Bound	79.9087
5% Trimmed Mean	74.9444	
Median	76.0000	
Variance	78.571	
Std. Deviation	8.86405	
Minimum	60.00	
Maximum	91.00	
Range	31.00	
Interquartile Range	12.00	
Skewness	.229	.580
Kurtosis	-.343	1.121

Table 6. The Descriptive Statistics for the Two Raters

	Mean	Std. Deviation	N
Rater One	80.4667	8.52615	15
Rater Two	75.0000	8.86405	15

Table 7. The Correlations

		Rater One	Rater Two
Rater One	Pearson Correlation	1	.962**
	Sig. (2-tailed)		.000
	N	15	15
Rater Two	Pearson Correlation	.962**	1
	Sig. (2-tailed)	.000	
	N	15	15

Table 8. Descriptive Statistics

	Mean	Std. Deviation	N
Rater One	80.4667	8.52615	15
Rater Two	75.0000	8.86405	15
wring Process	44.2667	7.29449	15
General Instructions	18.6000	5.02565	15
Writing Task	56.1333	6.41278	15

Table 9. The Result of Correlation across the Components of Writing Proficiency

	Rater One	Rater Two	Wring Process	General Instructions	Writing Task
Rater One					
Pearson Correlation	1	.962**	-.261	-.120	-.249
Sig. (2-tailed)		.000	.348	.669	.370
N	15	15	15	15	15
Rater Two					
Pearson Correlation	.962**	1	-.224	-.014	-.276
Sig. (2-tailed)	.000		.422	.959	.319
N	15	15	15	15	15
Wring Process					
Pearson Correlation	-.261	-.224	1	.445	.729**
Sig. (2-tailed)	.348	.422		.096	.002
N	15	15	15	15	15
General Instructions					
Pearson Correlation	-.120	-.014	.445	1	.343
Sig. (2-tailed)	.669	.959	.096		.211
N	15	15	15	15	15
Writing Task					
Pearson Correlation	-.249	-.276	.729**	.343	1
Sig. (2-tailed)	.370	.319	.002	.211	
N	15	15	15	15	15

8.1.2. Inferential Statics

The current research has a separate quantitative data collection procedure in which the researcher has attempted to test the hypotheses on the correlation coefficients across the independent variable of the research with the nature of writing proficiency. Table 10 shows

the results of the correlation across the components of writing proficiency and the rater's allocated scores.

Table 10. The Result of Correlation Across The Components of Writing Proficiency

		Rater one	Rater two	Wring process	General instructions	Writing task
Spearman's rho	Sig. (2-tailed)	.369	.292	.002	.366	.
	N	15	15	15	15	15
	Correlation Coefficient	1.000	.974**	-.305	-.187	-.307
	Rater One Sig. (2-tailed)	.	.000	.269	.504	.266
	N	15	15	15	15	15
	Correlation Coefficient	.974**	1.000	-.279	-.061	-.328
	Rater Two Sig. (2-tailed)	.000	.	.313	.828	.233
	N	15	15	15	15	15
	Correlation Coefficient	-.305	-.279	1.000	.484	.802**
	Wring Process Sig. (2-tailed)	.269	.313	.	.068	.000
	N	15	15	15	15	15
	Correlation Coefficient	-.187	-.061	.484	1.000	.270
	General Instructions Sig. (2-tailed)	.504	.828	.068	.	.331
	N	15	15	15	15	15
	Correlation Coefficient	-.307	-.328	.802**	.270	1.000
	Writing Task Sig. (2-tailed)	.266	.233	.000	.331	.
	N	15	15	15	15	15

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

Table 11. Descriptive statistics

		Statistic	Std. Error	
Wring Process	Mean	44.2667	1.88343	
	95% Confidence Interval for Mean	Lower Bound	40.2271	
		Upper Bound	48.3062	
	5% Trimmed Mean	44.0185		
	Median	44.0000		
	Variance	53.210		
	Std. Deviation	7.29449		
	Minimum	34.00		
	Maximum	59.00		
	Range	25.00		
	Interquartile Range	13.00		
	Skewness	.468	.580	
	Kurtosis	-.369	1.121	
	General Instructions	Mean	18.6000	1.29762
95% Confidence Interval for Mean		Lower Bound	15.8169	
		Upper Bound	21.3831	
5% Trimmed Mean		18.4444		
Median		17.0000		
Variance		25.257		
Std. Deviation		5.02565		
Minimum		12.00		
Maximum		28.00		
Range		16.00		
Interquartile Range		10.00		
Skewness		.288	.580	
Kurtosis		-1.098	1.121	

	Mean	56.1333	1.65577
	Lower Bound	52.5821	
	95% Confidence Interval for Mean	Upper Bound	59.6846
	5% Trimmed Mean	55.9259	
	Median	53.0000	
	Variance	41.124	
Writing Task	Std. Deviation	6.41278	
	Minimum	48.00	
	Maximum	68.00	
	Range	20.00	
	Interquartile Range	11.00	
	Skewness	.617	.580
	Kurtosis	-.957	1.121

As the sample group has just 15 members and it's too small to be quantitatively significant in order to be verified through quantitative analysis, the qualitative analysis was done as follows.

8.2. Qualitative Results

The researchers taught the writing instruction to the students for 10 sessions, and collected numerous writings, but finally 15 of them were analyzed. The analysis of the writings in order to the writing's qualification aspects, the following result is gained. Almost all of the students used the words from the reading parts of the book, the number of sentences and t-units was almost the same between them and there was no special difference between three levels of proficiency. The number of words which they use was were between 64 to 240 words, and it seemed that students in level 1, used more words than students in level 2, and students in level 2 used more words than level 3. The investigation of complexity of the words showed that they used simple to medium words in terms of syllables and word complexity. The number of sentences is almost the same between three levels. Cohesion and coherence of texts was studied, and came to this result that the students in level 1 were more capable in using connectors in their sentences and the unity of the texts are better than level 2 and 3. The samples were analyzed based on different criteria as follows:

Fluency: The researchers counted the number of words of the essays, and realized that there is no special difference between the numbers of words between three groups.

Table11. The Table of the Number of Words in Each Essay

Students' ID Numbers	Number of Words
1	240
2	64
3	95
4	73
5	130
6	162
7	155
8	122
9	115
10	90
11	96
12	72
13	96
14	102
15	99

Grammatical Accuracy: The scores of two raters to the students' writings show that the grammatical accuracy of students in level 1 was higher than students in other two levels.

Lexical Sophistication: As the researchers investigated the students' papers, it showed that students in level one used more complex words than levels two and three, but there was no special difference between levels two and three.

Syntactic Complexity: Syntactic complexity was calculated based on the average number of words in each T-unit and the mean number of T-units per sentences. The average number of T-units in each sentence is reported as following tables:

Table 12. The Mean Number of T-units of Level 1

ID number	T-units	Sentences	T-units/sentence
5	4	11	0.36
8	7	5	1.4
10	6	4	1.5
14	5	7	0.71
15	5	4	1.25

M=1.04

Table 13. The Mean Number of T-Units of Level 2

ID number	T-units	Sentences	T-units/sentence
1	11	16	0.68
2	4	5	0.8
3	5	7	0.71
7	10	6	1.6
10	6	4	1.5
M=1.05			

Table 14. The Mean Number of T-Units of Level 3

ID number	T-units	Sentences	T-units/sentence
4	4	4	1
6	8	7	1.14
9	8	9	0.88
11	7	6	1.16
13	5	8	0.62
M=0.96			

These tables show that, the approximate amount of T-units to the sentences is almost near in three levels, but there is no significant difference between the mean of level one and three, wherever level two is distinct.

Source Use: In order to investigate the source use patterns in reading-to-write tasks across different proficiency levels, the following indicators were used:

1. Direct source use with quotation
2. Direct Source use without quotation
3. Indirect source use
4. Total number of source use in each essay

According to the collected data it is found that none of the participants of this study used direct source use with or without quotation. However in level one and two indirect sources use was found. It also shows that the total source use in essays of levels one and two were more than level three.

Writers' Reported Source Use and Proficiency Level

In addition to exploring writers' differences in discourse features and source text use, a questionnaire was used to delve into how their processes differed. Particularly this study focused on the use of the source texts. In this section, the results from analyzing the

questionnaire will be detailed in terms of source text use for ideas and organization; integration processes; knowledge of integrating reading and writing; task difficulties; and integration style.

According to the questionnaire and question number 42, about the required time to complete the task, all students in three levels had similar responses, that were three of them in each level thought they had enough time, but two of them in each level said they need more time to complete the task.

Based on the question number 43, that is about the difficulties that they had with reading or writing in the task, three out of five students of level one said they didn't know the meaning of some words, but the last two students said they didn't have any difficulties. In level 2 and 3 students also said they didn't know the meaning of some words, but one of students said she didn't have enough time and the last one didn't answer the question. In level 3 again three students said they had difficulties with the vocabulary, but the last two students didn't have any ideas.

About the question 44 that asked students to write how they used reading for writing, 4 of the students in level one said that they hadn't used this method before but one of them said she usually read the passage and try to write an article based on it. Four students in level 2 also said they hadn't used this method before, but one of the students said she used this method to get ideas from the passage. All of the students in level 3 said they hadn't used this method before.

About the question number 45 that was "Would you prefer writing with reading on attest to a writing rest without reading? Why or why not?" four students in level 1 said they prefer writing with reading, because they have some ideas, but one of the students preferred writing without reading. All of the students in level 2 said they prefer writing with reading on attest, because it gives them some ideas and it's easier in their opinions. 4 if the students in level 3 said they prefer writing with reading on a test because it's easy, and one of them said "with reading is better because I have problems with spelling and words." However one of the students thought writing without reading is better but didn't mention "why?".

Question 46 was "What would you tell a teacher who is thinking about using reading to write tasks in his/her classes?" All of the students in level 1 said it's good and useful to use this method in writing classes. Four students in level 2 said they agreed with using this method in writing classrooms because it's "helpful and interesting", but one of them didn't answer this question. Four students also in level 3 said using this method is "useful and good" but one of them thought it was hard for her.

Question 47 was "is there anything you would like to comment on about this task?" two students in level 1 said they thought it's good and useful. One of students said "I like use it more in class." And one of them didn't answer this question. All of the students in level 2 didn't have any comments or ideas. Three students in level 3 didn't answer this question, one of the students said it was hard for her, and one said it's good for writing.

About background information, three students in level 1 said they had been two years at the high school and the last two students said they had been 3 years. Two students in level 2

said they had been three years at university; one of them had been three years at high school, the other one had been 4 years at university and the last one had been 2 years at university. Two students in level 3 said they had been two years in high school; two of them had been three years at high school. Finally, one of them had been one year at the university.

The question 49 was about the major of the participants. Four students in level 1 were majoring in mathematics, and the last one was majoring in science. Two students in level 2 were majoring in math, one was a chemical engineering student one of them was majoring in biology and the last one's major was computer. Three students in level 3 were majoring in experiential science. One of the participants was majoring in art and last one's major was math.

The next question was "have you had previous writing courses?" all of the participants in three levels said they did not have writing courses, but one of them in level 2 said she had a very short writing course in high school.

Question number 51 was "what is your level of experience with academic writing?" three students of level 1 said they had a little experience, but two students chose the option "some" for this question. Three students in level 2 said they had some experience with academic writing and two students said they had a little experience. Four students in level 3 said they had no experience with academic writing but one of them said she had "a little experience".

The last question was "What is your level of English language skill?" three students in level 1 said their level in English language is medium and the two students chose the option "so-so". All of the students in level two said their level of English language skill is medium. Two students in level 3 said their level is medium. Two students said they are in low level, and one chose the answer "so-so". However, all of the 15 students said they have not taken IELTS courses.

9. CONCLUSION

Integrated tasks provide an interesting, innovative, and authentic means to assess academic writing. Writers and test developers see them as a potential solution to a number of dilemmas with impromptu independent tasks. This study investigated several aspects of the products and processes from writers completing integrated tasks in order to learn more about them. The findings show interesting differences between low and high proficiency writers. More questions arose in this study and will require attention for the field to continue using integrated tasks and for test users to understand the resulting scores.

This study holds implications for developers and researchers of integrated tasks as well as those using the tasks or the scores with L2 writing. Firstly, we need to continue investigation into the skills in the tasks to fully understand how comprehension impacts the use of source texts and whether, as our study found, only a certain degree of comprehension is needed. This research may inform the notion of a threshold for integration of reading and writing, meaning writers need to reach a certain level of proficiency in order to comprehend the source texts used in completing these tasks.

Regardless of comprehension issues, materials developers should carefully consider texts chosen, adapted or written for integrated tasks, as the source texts are facilitating writers' processes both as academic writing/thinking as well as in providing language support. In addition, tasks should have instructions that clarify the expected uses for the source texts. While writers gravitate to the texts for ideas and words, they may need firm direction to use the sources to support their ideas on the topic, if the task is persuasive writing.

REFERENCES

- Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11, 17–34.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communication*, 8, 533–556
- Campbell, C. (1990). Writing with other's words: Using background reading texts in academic compositions. In B. Kroll (Ed.), *Second language writing* (pp. 211–230). Cambridge UK: Cambridge University Press.
- Carlisle, R. (1989). The writing of Anglo and Hispanic elementary school students in bilingual, submersion, and regular programs. *Studies in Second Language Acquisition*, 11, 257–280.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.
- Chastain, K. (1990). Characteristics of graded and ungraded compositions. *Modern Language Journal*, 74, 10–14.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10, 5–43.
- Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing*, 7, 1–18.
- Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139–155.
- Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based writing assessment. *College ESL*, 6 (1), 73–84.
- Fischer, R. (1984). Testing written communicative competence in French. *Modern Language Journal*, 68, 13–20.
- Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal*, 79, 329–344.
- Frase, L., Faletti, J., Ginther, A., & Grant, L. (1999). *Computer Analysis of the TOEFL Test of Written English* (TOEFL Research Report No. 64). Princeton, NJ: Educational Testing Service.

- Gebril, A. (2006). *Independent and integrated academic writing tasks: A study in generalizability and test method*. Unpublished doctoral dissertation, The University of Iowa, Iowa City.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9, 123–145.
- Hamp-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. *Language Learning*, 41, 337–373.
- Henry, K. (1996). Early L2 writing development: A study of biographical of essays by university-level students of Russian. *The Modern Language Journal*, 80, 309–326.
- Hirano, K. (1991). The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students. *Annual Review of English Language Education in Japan*, 2, 21–30.
- Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly*, 18, 87–107.
- Johns, A., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, 11 (3), 253–271.
- Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal*, 75, 305–313.
- Kobayashi, H. & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning*, 42, 183–215.
- Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 140–154). Cambridge: Cambridge University Press.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12(4), 75–84.
- Lewkowicz, J. A. (1994). Writing from sources: Does source material help or hinder students' performance? In M. Bird et al. (Eds.), *Language and learning: Papers presented at the Annual International Language in Education Conference*. ERIC Document (ED 386 050).
- Manchon, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing: A study of foreign language writers. *Language Learning*, 57 (4), 549–593.
- Melenhorst, M. (2006). *Highlighting professional writing: On screen note-taking as part of writing-from-sources by professionals*. Unpublished doctoral dissertation, The University of Twente, The Netherlands.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.
- Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly*, 14, 61–69.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13 (2), 111–129.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101–143.

- Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, 9 (2), 109–121.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9, 259–291.
- Tedick, D.J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes*, 9, 123–143.
- Weigle, S. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(1), 27–55.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second language development in writing: Measures of fluency accuracy and complexity*. Honolulu: University of Hawai'i at Manoa.
- Zhang, S. (1987). Cognitive complexity and written production in English as a second language. *Language Learning*, 37, 469–481.