

# Investigating Linguistic Differences in Adult EFL Learners' Writing Test Performance\*

Kim, Hyun Jung · Lee, Jooseung

## Contents

- I. Introduction
- II. Literature Review
- III. Methodology
- IV. Results and Discussion
- V. Conclusion

## I. Introduction

Second language (L2) testers have adapted different theoretical models of communicative competence or communicative language ability (e.g., Bachman, 1990; Bachman & Palmer, 1996, 2010; Canale 1983; Canale & Swain, 1980) to define various aspects of L2 ability, including writing ability. Specifically, language knowledge/competence, which is theoretically defined as part of the communicative language ability models, has been drawn upon to operationalize the construct definition of a language test. Such operational definitions of language ability may differ across diverse language tests. One of the most common components, however, which is often included or considered as part of L2 writing ability across different

---

\* This work was supported by Hankuk University of Foreign Studies Research Fund of 2015.

tests, is test-takers' textual knowledge, or organizational control (knowledge about how to structure written discourse with coherence and cohesion in mind).

Previous empirical research has reported that L2 learners often find it difficult to organize written texts in their L2 through the course of language learning in general, and in a testing situation in particular (Bitchener & Basturkmen, 2006; Lee, 2002; Riazantseva, 2012). Cultural differences in textual organization and a lack of genre awareness have been reported as contributing factors of such difficulty. Although L2 learners' ability to organize texts has been examined in relation to their scores on organizational control (e.g., validation studies), there is a lack of research that analyzes actual writing responses with regard to the linguistic features used to organize written discourse. That is, in addition to the quantitative analysis of test scores, an actual analysis of the linguistic features in L2 writing is needed to better understand learners' experience in L2 writing and the accompanying difficulties.

A more practical motivation of the present study comes from the findings of earlier research on two types of L2 writing tests (Kim, 2015). In the study, two types of expository writing (i.e., compare/contrast writing and problem/solution writing) administered as a high-stakes admissions test were compared to examine the validity of the writing admissions test. A many-facet Rasch measurement analysis of 143 test-takers' test performance found that the test-takers who selected the compare/contrast writing type had greater difficulties in organizing their response/essay (organizational control) than in elaborating the topic with supporting details (content control) using accurate and diverse grammatical forms (grammatical control). Contrary to the compare/contrast writing group, the problem/solution writing group did not necessarily represent different levels of difficulty in L2 writing across different aspects of L2

writing ability (organization, content, and grammatical control). Therefore, it is necessary to further examine such different levels of difficulty in organizational control found in the two different types of writing by analyzing and comparing the test-takers' actual responses. To this end, the present study aimed to examine the linguistic features used in the two types of writing to organize written discourse as a follow-up to Kim (2015). The comparison of linguistic features in the two types could explain a possible source of different levels of difficulty in organizational control and further support the use of two types of writing for the high-stakes admissions test.

In addition, the raters' perceptions about the linguistic differences they expected or noticed while scoring were examined. In the testing context of Kim (2015), the raters evaluated both types of writing, contrary to the test-takers, who selected and responded to only one writing type out of two choices. Therefore, it was expected that the raters could explain differences in overall or certain linguistic features between the two types of writing from their experience in scoring organizational control. The analysis of the raters' perceptions could support the results of the analysis of the linguistic features included in the two types of writing. For these purposes, the following two research questions were addressed in the study:

- 1) To what extent do adult EFL learners' linguistic features used to organize written discourse differ between the two types of writing?
- 2) How do the raters perceive the linguistic differences between the two types of writing while evaluating test-takers' discourse?

## II. Literature Review

### 1. Second Language Ability and Writing Ability

There have been several models explaining language ability in the field of L2 research (e.g., Bachman, 1990; Bachman & Palmer, 1996; Canale 1983; Canale & Swain, 1980). Of these, the model of Bachman and Palmer (2010) is one of the most frequently adopted or adapted language ability models. According to Bachman and Palmer (2010), language ability is "a capacity that enables language users to create and interpret discourse" (p. 33). In their model, language ability includes two areas (language knowledge and strategic competence) as a whole, and language knowledge comprises its own subcomponents (organizational knowledge and pragmatic knowledge). Organizational knowledge has two distinguishable subcomponents (grammatical knowledge and textual knowledge), and so does pragmatic knowledge (functional knowledge and sociolinguistic knowledge). In addition, there are other attributes affecting language ability, which are known as personal attributes, topical knowledge, affective schemata, and cognitive strategies.

While the model mentioned above gives plentiful insight into understanding language ability, language skills could or should be differently appreciated from context to context. In defining L2 writing ability, di Gennaro (2011), for instance, considered studies conducted by other researchers, concluding that writing ability for the study should include (1) language use features targeting lexico grammatical accuracy and appropriacy, (2) language use context features characterizing participants and tasks, (3) discourse knowledge for organization, coherence, and genre conventions, (4) sociolinguistic knowledge describing the expression of claims and contextual appropriacy, and (5) pragmatic

knowledge specifically for writing.

Discourse knowledge, which is the focus of the present study, refers to the ability to produce or comprehend the sequence of informational units in either written or spoken text (Bachman & Palmer, 2010), and the knowledge includes cohesion and coherence. In di Gennaro (2011), cohesion and coherence were separately defined, since a cohesive text may not be coherent, and vice-versa. It is, however, more common to consider these linguistic features in the same category. For example, Bachman and Palmer (2010) place cohesion and coherence under textual knowledge.

## **2. L2 Learners' Difficulties in L2 Writing**

It is well known that differences between first language (L1) and L2 writing can bring about potential problems in L2 writing. These differences include (1) differences in linguistic proficiency, (2) differences in intuitive knowledge of language, (3) different preferences in ways to organize composition, and (4) differences in knowledge on how different text types are socially valued (Hyland, 2003). Bitchener and Basturkmen (2006), for instance, conducted interviews to examine the perceptions of L2 graduate students and their supervisors on the students' difficulties in writing a discussion of results section (DRS) in English. The study found that the student participants' responses were not as consistent as the supervisors' responses were. The students tended to believe that general English proficiency was a threat to their L2 academic writing. They also could not explicitly explain how to write a DRS (i.e., the functions and contents of a DRS). This conforms to the comment by the supervisors, who pointed out that "students lacked a full enough understanding of the DRS as a genre" (p. 10). In addition to a lack of language ability, insufficient knowledge about the genre appeared to have an impact on their academic writing.

Another difficulty L2 learners face is the organization of written discourse, which is closely related to cultural factors. In a study by Reid (1992), participants from four different language backgrounds (Arabic, Chinese, Spanish, and English native speakers) were chosen to compare L2 learners' organization of a text with English native speakers' textual organization. Reid found that the use of organizational and cohesive devices is different between L1 and L2 writers. Discourse analysis and analysis of contrastive rhetoric of four cohesive variables found that organizational knowledge in L1 writing affected textual organization in L2. Kobayashi and Rinnert (1996) also demonstrated the effects of culture on L2 writing. L2 writers under an EFL context used their L1 discourse features when writing in English. Of these discourse features, rhetorical organization was selected for comparison, and the findings revealed that the use of preferred rhetorical patterns, which are based on one's own culture, affected the scores of L2 writers' compositions.

In relation to the organization of a text, L2 writers differ from L1 writers in their use of linguistic features. Crossley and McNamara (2012) examined linguistic features in order to predict L2 writing proficiency. One reason they focused on cohesion was that cohesion has been a traditional predictor of proficiency in L2 writers. In the study, a computational tool, Coh-Metrix, was adopted for analysis, and indices related to cohesion were examined. The study suggested two cohesion-related features (lexical diversity and aspect repetition) as partially characterizing L2 writers' proficiency.

Until recently, researchers in the field of writing assessment have tended to draw upon either rater behavior or rating scales and the results of rating (i.e., scores) rather than the actual product of L2 writers when analyzing L2 learners' writing ability or development (e.g., Bae & Lee, 2012; di Gennaro, 2009; Huhta, Alanen, Tarnanen, Martin, & Hirvela,

2014; Knoch, 2009; Li & He, 2015; Lumley, 2002; Schoonen, 2005; Shi, 2001). Bae and Lee (2015), for instance, set five analytic ratings as dependent variables to investigate young learners' quantitative development of writing skills under an EFL setting. Another study (Li & He, 2015) comparing newly developed analytic scales with existing global scales gathered data from raters through think-aloud protocols. Then, the data were statistically analyzed in order to find out difference in essay-rating processes. di Gennaro (2009) analyzed composite ratings derived from analytic rating scales yielding one total score to determine the writing features differentiating two types of L2 writers (i.e., generation 1.5 and international college students). With reference to qualitative-based research, Plakans (2009) gathered data from three sources, including the written products of examinees. However, the products were marked by the raters instead of being linguistically analyzed. Thus, little is known about L2 learners' actual use of linguistic features, including discourse and cohesive features, in writing.

### **3. Study of Linguistic Features Using Coh-Metrix**

In searching for the linguistic features affecting L2 learners' writing, researchers can benefit from computer software, as computer programs help not only save time and effort (e.g., counting and indexing linguistic features), but also return the quantitative results of those features. In light of this, Coh-Metrix is an attractive tool for those who are interested in studying linguistic features. It is a web-based software having several modules for measuring cohesion-related indices on different levels (Grasser, McNamara, Louwerse, & Cai, 2004). As is illustrated in Figure 1, the users input basic information about a text (title, source of the text, user code, categories/genre of the text) and enter the text in the text

window. Once they click the “submit” button, the results of the text analysis on the text appear on the screen. (Categories of measures are introduced in the data analysis section.)

Figure 1. Screen Shot of Coh-Metrix with Text Input and Measures of Language and Cohesion (Grasser et al., 2004, p. 3)

**Coh-Metrix**

Welcome agrasser!

Title: The Needs of Plants

Source: Research

User Code: GrasserTest Genre: Science LSA Space: CollegeLevel CollegeLevel

**What Are the Needs of Plants?**

Like all living things, plants have certain needs. Plants need sunlight, water, and air to live. Plants also need minerals (MIN+uhr+uhlt). A mineral is a naturally occurring substance that is neither plant nor animal.

The parts of plants help them to get or make what they need. All plants get water and minerals from the soil. The root is the part of the plant that grows underground. Roots help hold the plant in the ground. Roots also help take in water and minerals that the plant needs.

The stem is the part that supports the plant. It helps the plant stand upright. It carries minerals and water from the roots. It also carries food from the leaves to other parts of the plant.

Some plants, such as mosses, are simple plants. They don't have real roots or stems. These plants do not grow tall. Instead, they form low-growing mats in damp places to get water directly from the soil.

Other plants, such as the redwood tree, have many roots and a large stem. They can grow very tall.

How Does a Plant Move Food?

**Primary Measures**

Coreference Cohesion Global 1	0.589
Coreference Cohesion Global 2	0.647
Coreference Cohesion Local 1	0.75
Coreference Cohesion Local 2	0.818
Causal Cohesion	0.75
LSA Global	0.576
LSA Local	0.582
Reading Ease	85.812
Reading Grade Level	3.84
Word Frequency	2.316
Number of Words	462
Type-Token Ratio	0.6
Connectives	73.583

**Help Links**

- Causal Cohesion
- Concept Clarity
- Connectives
- Coreference Cohesion
- Densities
- Logical Operators
- LSA
- Part of Speech
- Polysemy Hypemym
- Readability
- Syntactic Complexity
- Type Token Ratio
- Word Frequency
- Word Information

Clear Submit

Coh-Metrix has been introduced and used in L2 research in order to analyze learners' use of various linguistic features. For example, Crossley, Salsbury, and McNamara (2010) used Coh-Metrix when analyzing four lexical cohesive indices from non-native speakers' English writing in order to understand the relationship between the use of cohesive devices and the coherence of written discourse. Another study by Crossley and McNamara (2011) also used Coh-Metrix for an analysis of cohesive linguistic features (e.g., connectives, word overlap, and semantic co-referentiality) to see whether the same L1 backgrounds shared the same linguistic features. While the contribution of the aforementioned studies cannot be neglected, the use of Coh-Metrix has been somewhat limited in L2

research in general, and in L2 writing research in particular. For instance, L2 writers' use or choice of linguistic features has rarely been examined in relation to different types of genres of writing. Therefore, it is suggested that the use of Coh-Metrix be more broadly exploited across various fields of TESOL and applied linguistics (e.g., L2 writing teaching and testing).

### III. Methodology

#### 1. Participants

The participants of the present study included 143 test-takers (119 females and 24 males) who were applicants of the M.A. TESOL program at a university in Seoul. The test-takers took a writing test as part of the admissions process across five occasions from 2014 to 2015. They were between the ages of 23 and 56; the majority of the test-takers were in their twenties and thirties (128 test-takers, approximately 90%). The test-takers also included seven applicants from an English-speaking country (i.e., the U.S. and Canada). In addition, 15 Korean applicants earned their B.A. degree in an English-speaking country.

The other group of participants included two raters who scored the 143 test-takers' writing responses at the time of the admissions selection process. Both raters were native speakers of English who had been teaching various TESOL subject courses in the M.A. TESOL program for approximately seven years. They had also participated in the interview process regularly for the purpose of admission for those seven years. Therefore, they already had a good understanding of the applicants' language ability. In addition to their teaching experience in the context of

the current study, they had extensive teacher training experience in various contexts for over 15 years.

## 2. Instruments

### 1) Writing Tests

As part of the admissions selection process, all applicants to the M.A. TESOL program took a writing test. The writing test for each of the five administrations, which was used as the test instrument for the present study, presented two types of writing of the expository essay genre: compare/contrast writing and problem/solution writing. Applicants were instructed to choose one of the two writing types and to write a 300-word essay for thirty minutes. Therefore, they had a chance to select a type of writing, possibly a topic of writing, that they liked. The topics of the writing prompts were not repeated, although the same two types of writing were used for each administration. The number of test-takers who selected the compare/contrast type ( $N = 70$ ) was almost identical to the number of test-takers who chose the problem/solution type ( $N = 73$ ).

The compare/contrast and problem/solution writing types were used for the purpose of admission in order to make accurate predictions about test-takers' future writing performance in the M.A. TESOL program, by presenting writing prompts whose characteristics corresponded to the features of real-life academic writing (e.g., making comparisons on the basis of evaluation of both the pros and cons of a given issue, and expressing opinions to suggest a possible solution to a given/potential problem). The first type of writing (i.e., compare/contrast writing) required test-takers to evaluate and discuss the advantages and disadvantages, or strengths and weaknesses, of a given issue. The second type (i.e., problem/solution writing) asked test-takers to identify the

problems of a given issue and to make suggestions to solve the problems. Both writing types included topics that were related to general issues of English education in Korea. In order to give an equal opportunity to all applicants, regardless of their prior knowledge of TESOL, general issues that had been introduced oftentimes through various types of media (e.g., newspapers, TV, and radio) were included in the writing prompts.

The tests were delivered in paper format. There was no separate time given for preparation before the actual writing; thus, the test-takers were supposed to make a plan, write an essay, and proofread their writing within 30 minutes. Dictionary use was not allowed.

## 2) Scoring Rubric

Test-takers' written responses were evaluated independently by two raters using an analytic scoring rubric. It included three rating scales of grammatical control, organizational control, and content control. The grammatical control scale considered the degree to which test-takers used accurate, diverse, and complex lexical, syntactic, and graphical forms in their writing. The organizational control scale measured both the coherence of written discourse and the use of cohesive devices used to connect sentences and paragraphs. The last content control scale evaluated the extent to which test-takers elaborated the given topic and provided details to support their arguments or ideas. A six-point scale (0, 2, 4, 6, 8, and 10) was used for each rating scale. A score of 0 was assigned when it was impossible to evaluate the response due to a lack of evidence, while a score of 10 was assigned for a response that represented full control of each scale (e.g., completely coherent organization with accurate use of cohesive devices for organizational control). (The actual scoring rubric cannot be presented due to confidentiality issues.)

### 3. Data Collection Procedure

Three sets of data were collected in the current study: test-takers' writing scores, their actual written responses, and the two raters' interview data. First, 143 test-takers who applied to the M.A. TESOL program took a writing test from one of the five administrations between 2014 and 2015. As described above, they were presented with two writing prompts and were required to choose one prompt and write a 300-word essay for 30 minutes. Their responses were evaluated at the time of the admissions test by two raters who assigned ratings using an analytic scoring rubric.<sup>1)</sup> The averages of the two raters' ratings, which had been reported as the results of the writing portion of the admissions test, were used as the score data for the present study. In addition to the test-takers' writing scores, their written responses were obtained for linguistic analysis using Coh-Metrix. All handwritten responses were entered into the computer verbatim for analysis. That is, the responses were transcribed as they were, without any errors (e.g., spelling and grammar) corrected.

In order to examine the two raters' perceptions about the test-takers' ability to organize their responses, a semi-structured interview was conducted with each of the two raters. Since the interviews were conducted after all five administrations had been completed, the raters were provided with random sample responses and the scoring rubric to refresh their memory about the scoring. The interview questions included

---

1) The inter-rater reliability was calculated using Pearson product-moment correlations. The correlation coefficient was .66 ( $p < .001$ ), showing a moderate level of agreement in scoring. However, further analysis of the rater facet in many-facet Rasch measurement indicated that the two raters were used the three rating scales consistently within their own ratings although they differed somewhat in severity. (For a detailed discussion of rater behavior, refer to Kim, 2015.)

the raters' own definition of organizational control; the subcomponents of organizational control to which they paid special attention while scoring a description of the types of responses to which they assigned a high score of organizational control; differences in organization they perceived while evaluating the two types of written responses; and expected difficulties test-takers might have felt when organizing written discourse for each of the two writing types. Each interview lasted for approximately 40 minutes. Both interviews were recorded and transcribed for later analysis.

#### 4. Data Analysis

For the data analysis, the linguistic features of the test-takers' responses were first analyzed using Coh-Metrix version 3.0, which is a computation tool that evaluates the indices of the linguistic and discourse features of a text (Graesser et al., 2004; McNamara & Graesser, 2012). Among over 100 indices, only four groups of indices, which were closely related to the descriptors of the organizational control scale in the rubric, were analyzed for the present study. These four groups included (1) *descriptive indices*, which provide basic information of a text (e.g., number of paragraphs, sentences, and words); (2) *referential cohesion*, which evaluates local cohesion (overlap between adjacent sentences) and global cohesion (overlap between all sentences in a paragraph); (3) *latent semantic analysis*, which measures semantic overlap between sentences and between paragraphs; and (4) *connectives*, which presents frequencies of different types of connectives (e.g., causal, logical, adversative/contrastive, temporal, and additive connectives). To answer the first research question (i.e., differences in the linguistic features used to organize written discourse between the two types of writing), a series of independent samples *t*-tests were computed in which the two writing

groups (i.e., two types of writing) were compared for each of the indices included in the analysis.

Following the analysis of the linguistic features of the test-takers' written responses, the recorded interview data were qualitatively analyzed to answer the second research question, which examined the two raters' perceptions about the similar/different linguistic features that the test-takers used to organize the two types of writing. The data were first sorted out for each interview question. A content analysis of each segment and cross-story analysis (between the two raters' perceptions) followed in order to supplement the findings of the linguistic feature analysis (Heigham & Croker, 2009).

## IV. Results and Discussion

### 1. Linguistic Differences in the Two Types of Writing

In order to answer the first research question, which examined the differences in the linguistic features used to organize written discourse between the compare/contrast (N = 70) and the problem/solution writing (N = 73) groups, 143 test-takers' written responses were analyzed using Coh-Metrix. The four groups of the Coh-Metrix indices that were relevant to the analysis of this study included a total of 36 individual indices (11 *descriptive indices*, 10 *referential cohesion*, 8 *latent semantic analysis*, and 7 *connectives*), for which a separate independent samples *t*-test was conducted to compare the two types of writing.

The results indicated that the two writing groups (compare/contrast and problem/solution writing) did not show any significant differences in terms of most of the indices. That is, the differences in the incidence scores for

the indices were not statistically significant in most analyses. More specifically, the two groups did not show differences at all in the *descriptive indices* or *latent semantic analysis*. In other words, the two types of written responses presented similarities in general with respect to the length of the paragraphs, sentences, and words; the mean length of the paragraphs; the mean number of words in each sentence; and the mean number of syllables and letters in all of the words in the response. Moreover, they were similar in the measures of semantic overlap between the sentences and between the paragraphs in the responses (e.g., measures of conceptual similarities between adjacent sentences; measures of conceptual similarities between each sentence and every other sentence in the response). Different from the *descriptive indices* and indices from the *latent semantic analysis*, there was a significant difference between the two writing groups in the following two indices of *connectives*: CNCLoGic (the incidence score of logic connectives) and CNCTempx (the incidence score of temporal connectives). The results of the *t*-tests showed that the compare/contrast writing group used more logical connectives (e.g., and, and then, or, either... or, if... then, but, however, neither... nor) than the problem/solution group ( $t = 2.72, p = 0.007$ ), while the problem/solution writing group included more temporal connectives (e.g., first, next, lastly, finally, then) in their responses than the compare/contrast group ( $t = -4.18, p < 0.001$ ). It appears that logical connectives might have been used more often in the compare/contrast writing group for idea development, which included contrary perspectives on an educational issue. On the contrary, the problem/solution writing did not involve multiple perspectives, but rather a single perspective to identify the problems of an issue and involved making suggestions for the problems. Thus, the test-takers might have used temporal connectives more often to maintain the flow of the argument and logical connections between ideas.

Contrary to the three groups of indices, in which the two writing groups showed similar use of linguistic features to form an essay (except for the logical and temporal connectives), the two groups represented rather different uses of linguistic features in terms of *referential cohesion*. Among the 10 indices of *referential cohesion*, the mean incidence scores were significantly different in six indices. The compare/contrast writing responses presented higher incidence scores in all six indices in which the difference between the two writing groups was statistically significant. Table 1 summarizes the mean incidence scores of the *referential cohesion* indices, accompanying standard deviations, and the results of the *t*-tests.

Table 1. Summary of *t*-test Results for the *Referential Cohesion* Indices

Index description	Compare/ contrast	Problem/ solution	<i>t</i>	<i>p</i>
	Mean (SD)	Mean (SD)		
Noun overlap (adjacent sentences)	0.45 (0.21)	0.39 (0.19)	2.10	0.04*
Argument overlap (adjacent sentences)	0.57 (0.18)	0.53 (0.17)	1.52	0.13
Stem overlap (adjacent sentences)	0.56 (0.20)	0.51 (0.19)	1.66	0.10
Noun overlap (all sentences)	0.41 (0.18)	0.33 (0.15)	2.58	0.01*
Argument overlap (all sentences)	0.51 (0.16)	0.47 (0.14)	1.84	0.07
Stem overlap (all sentences)	0.52 (0.18)	0.45 (0.16)	2.44	0.02*
Content word overlap (adjacent sentences)	0.13 (0.05)	0.11 (0.04)	2.81	0.01*

Content word overlap (adjacent sentences, SD)	0.12 (0.04)	0.11 (0.02)	1.88	0.06
Content overlap (all sentences)	0.12 (0.04)	0.10 (0.03)	3.72	0.00*
Content overlap (all sentences, SD)	0.12 (0.03)	0.11 (0.02)	2.97	0.00*

\*  $p < .05$

Overall, the analysis of the linguistic features in the written responses suggests that the two writing groups did not show significant differences in the use of linguistic features for cohesion and coherence in writing. However, the compare/contrast writing group used more linguistic features for *referential cohesion* by repeating nouns and words between adjacent sentences and within the whole text. Such results of the comparative analysis of the two writing groups revealed somewhat contradictory results from the findings of a previous study (Kim, 2015). Kim (2015) found that test-takers had greater difficulty organizing a coherent and cohesive essay when responding to a compare/contrast writing prompt, whereas the analysis of linguistic features in the present study found that the compare/contrast writing group included more linguistic features that could help test-takers organize a coherent and cohesive response.

In order to explain these contradictory findings, the relationship between the use of the linguistic features of *referential cohesion* and test-takers' scores on the organizational control scale was first examined. Correlations between the incidence scores of each of the six referential cohesion indices and analytic scores on organizational control were calculated to examine whether the raters considered *referential cohesion* (i.e., overlap, repetition) while scoring the two types of writing. An analysis of Pearson product-moment correlations reported that none of the six linguistic features were correlated with the test-takers' scores. Differences in the

use of the linguistic features of *referential cohesion* might have primarily resulted from the different characteristics of the two types of writing. However, the result of the correlation analysis suggests that the raters might not have considered *referential cohesion* while evaluating the test-takers' ability to organize an essay. In spite of rater training, rater effects are necessarily involved in scoring, which is likely to threaten the validity of test scores (AERA, APA, & NCME, 1999; Bachman, 2004; Reed & Cohen, 2001). As a result, the relative difficulty with the organization of the responses, which was found on the basis of a score analysis in Kim (2015), might not have reflected test-takers' attempts to integrate the features of *referential cohesion*.

## 2. Raters' Perceptions about Discourse Organization of the Two Types of Writing

In order to answer the second research question, which concerned how the two raters perceived linguistic differences revealed in the organization of the two types of writing, interview data were qualitatively analyzed. This analysis was used to supplement the findings from the analysis of linguistic features for the first research question. The content analysis of the raters' perceptions represented two themes in relation to (1) the raters' understanding of organizational control and (2) the task effect on test-takers' organization of the two types of writing.

First, when the raters were asked to define organizational control and explain the subcomponents of organizational control, both raters mentioned only the descriptors of the organizational control scale, as stated in the rubric. For instance, Rater 2 in the following excerpt stuck to the two subcomponents of organizational control in the rubric (coherence and cohesion) when explaining how he understood the scale while scoring it.

(Rater 2)

When I see students' writing samples, and I'm grading it in this aspect of the rubric, what I'm looking for in terms of coherence is how well the big ideas match together. That is, is there a logical sequence to what the student wishes to express? We are talking about cohesive devices. We are talking [about] pieces of language that students may use to connect these thoughts. So, cohesive devices might be connectors, transitional words and expressions that help link sentences together. I also look at very easily how well they are able to write paragraphs, how well are they able to write topic sentences and supporting sentences, some kind of concluding sentence at the end. So that shows some sort of logical sequence to what they wish to express.

His description was very similar to the descriptors of the rubric, and he did not add any other features that were irrelevant to the descriptors when asked a few more times to explain what he focused on while scoring for the organizational scale. In addition to focusing on the two subcomponents (coherence and cohesion), the raters also provided examples of certain features to which they paid attention to determine test-takers' ability to organize an essay. The same Rater 2 added examples of cohesive devices he looked for in test-takers' responses.

(Rater 2)

When it comes to cohesive devices, I will look at the terms such as *first, second, third, or next, then, after that, finally*.

Although the raters displayed a good, appropriate understanding of the scale and its descriptors, neither mentioned one aspect of cohesion, *repetition*, which was included as part of the rubric. It was interesting that

neither of the raters paid attention to this aspect because it was the aspect directly related to the *referential cohesion* indices, based on what differentiated the two writing groups in the earlier analysis of the written responses. When asked explicitly whether the raters considered overlaps or repetitions of words in scoring organizational control, both gave a negative response. Therefore, it was confirmed that the raters did not take different types of *referential cohesion* into account when scoring; consequently, the compare/contrast group of test-takers, who used more overlaps for cohesion, did not necessarily receive a high score on organizational control.

The other theme that emerged from the interview responses was the different effects of the two types of writing prompts on the test-takers' development of written discourse and the raters' evaluation of organizational control. When the two raters were asked to evaluate the relative difficulty in organizing the discourse by comparing between the two types of writing, Rater 1 expected the problem/solution writing type to be more difficult, while Rater 2 had an opposite opinion for the following reasons, respectively.

(Rater 1)

With compare and contrast, you have to think for both sides of the issue. You have to think, you know, not only about this side that supports in this case making an easier task; you also have to think about pros and cons in order to make more difficult tasks. So that's like four things, pros and cons of two things, whereas here [it's a] little bit more straightforward; you describe the problem, you offer three potential solutions to [it].

(Rater 2)

[The] first prompt [is] when you ask students to discuss or state the advantages or disadvantages; I find that task actually to be quite easy. Yes, easier than the other one because it's simply compare and contrast, and so students could maybe make a little graphical image.

Although the two raters displayed opposite opinions at the beginning, they stated the same difficulties that the test-takers might have had while taking the test or similar expectations they had as raters when reading the test-takers' responses. That is, both raters mentioned the time limit, which could make it difficult to compare the advantages and disadvantages of an issue and, at the same time, present the discussion in a coherent and cohesive manner (the compare/contrast writing type). Due to such a heavy cognitive load, the test-takers might not have been able to pay attention to the organization of the compare/contrast writing within the restricted time allowed.

(Rater 2)

Was that because the students couldn't give a summary or conclusion or final opinion, or was their time restricted or they're just able to put down positive and negative and then [they] ran out of time, because honestly, 30 minutes is not a lot of time to write 300 words. I think you'll find [with] native speakers, that time will be a little tight, and for EFL students, it would be very tight. Some students' samples in the past, they don't even finish their writings.

The other aspect both raters agreed on was the problem of the task itself, which related to the way the prompt was written for the compare/contrast writing type. They argued that they expected

test-takers' opinion or perspective while reading the compare/contrast writing responses, although the prompt simply required test-takers to discuss the advantages and disadvantages of an issue.

(Rater 1)

If I were to read an essay that is in, you know, had an introduction paragraph here the advantages, here the disadvantages and then a conclusion paragraph, I don't think I would be very impressed with that organizational structure. I'd want to see some discussion about whether or not the advantages over are stronger than disadvantages, so whether the disadvantages are stronger than the advantages, yeah, some comments about it to really get that A+ answer, you have to make a comment about how the advantages outweigh the disadvantages or how the disadvantages outweigh [the advantages]. Here are the advantages, here are the disadvantages, and here is what we should do, because one outweighs the other.

The second rater also pinpointed raters' expectations for the conclusion of the compare/contrast writing type, regardless of the instruction of the prompt.

(Rater 2)

It's not even, it's not asked; it might be expectations of the evaluators, that there would be some sort of final assessment that should be given by students. But if it's not asked [of] them in the prompts, then the students just do what they're told; perhaps they think they met the expectations.

The analysis of the raters' perceptions (or possibly native English

speakers' perceptions) about the organization of the two types of writing contributed to a better understanding of the test-takers' difficulty in organizing the compare/contrast written responses in Kim (2015). Their difficulty might have been related to the effects of the tasks, which included the time factor and the gap between the instruction of the prompts and the raters' expectations when reading a compare/contrast writing in real life.

## V. Conclusion

The results of the current study suggest that cohesive-related linguistic features of the two types of writing (i.e., compare/contrast and problem/solution writing) differed in terms of *referential cohesion*. However, correlations between *referential cohesion* and analytic scores on organizational control were not significant because repetitions or overlaps, which are related to *referential cohesion*, did not seem to be considered in the rating process. The raters did not pay attention to repetitions for coherence and cohesion, although repetitions were specified under the organizational control scale in the scoring rubric. From the interviews with the raters, it was also found that besides the time constraint, the prompt was likely to bring about different understandings by the raters and the examinees, which might have led to relative difficulty in organizing discourse in the compare/contrast writing in Kim (2015). It is hoped that such findings of the current study (the time factor and the task effect) will have implications for the development of a high-stakes L2 writing test and the evaluation of organization for written discourse. Moreover, the study implies the need for rater training in which raters and test developers have the same understanding of the prompt, the rubric, and the accompanying

descriptors.

In spite of the implications of the study, it has a number of shortcomings. First, the test-takers' perceptions of organizational control were not included or analyzed in this study. It is recommended to compare between L2 learners' perceptions concerning the two types of writing and the raters' perceptions. This might provide new insights about test-takers' difficulty in organizing a compare/contrast writing. Another limitation is that each test-taker chose and responded to only one writing task, which, in turn, creates a limitation in the data analysis. Future studies, therefore, are needed to collect various types of writing from the same L2 writers for a more comprehensive analysis of L2 writers' production and difficulties. Finally, qualitative analysis and comparisons of linguistic features used in different types of writing, in addition to the quantitative comparisons of the present study, will broaden our understanding of how L2 writers organize texts in L2 for different types of writing.

## Works Cited

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. *Standards for educational and psychological testing*. 2nd ed. Washington, DC: American Educational Research Association, 1999.
- Bae, J., & Y.-S. Lee. "Evaluating the development of children's writing ability in an EFL context." *Language Assessment Quarterly* 9.4 (2012): 348-74.
- Bachman, L. F. *Fundamental considerations in language testing*. Oxford: Oxford UP, 1990.
- \_\_\_\_\_. *Statistical analyses for language assessment*. Cambridge: Cambridge UP, 2004.
- \_\_\_\_\_. & A. S. Palmer. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford UP, 1996.
- \_\_\_\_\_. *Language assessment in practice*. Oxford: Oxford UP, 2010.
- Bitchenor, J., & H. Basturkmen. "Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section." *Journal of English for Academic Purposes* 5.1 (2006): 4-18.
- Canale, M. "On some dimensions of language proficiency." *Issues in language testing research*. Ed. J. W. Oller, Jr. Rowley, MA: Newbury House, 1983. 333-42.
- \_\_\_\_\_. & M. Swain. "Theoretical bases of communicative approaches to second language teaching and testing." *Applied Linguistics* 1 (1980): 1-47.
- Crossley, S. A., & D. S. McNamara. "Shared features of L2 writing: Intergroup homogeneity and text classification." *Journal of Second Language Writing*, 20.4 (2011): 271-85.
- \_\_\_\_\_. "Predicting second language writing proficiency: The role of

- cohesion and linguistic sophistication.” *Journal of Research in Reading* 35.2 (2012): 115–35.
- \_\_\_\_\_ and T. L. Salsbury. “The role of lexical cohesive devices in triggering negotiations for meaning.” *Applied Linguistics* 18.1 (2010): 55–80.
- di Gennaro, K. “Investigating differences in the writing performance of international and generation 1.5 students.” *Language Testing* 26.4 (2009): 533–59.
- \_\_\_\_\_. *An exploration into the writing ability of generation 1.5 and international second language writers: A mixed methods approach*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York, 2011.
- Grasser, A. C., D. S. McNamara, M. M. Louwse, & Z. Cai. “Coh–Metrix: Analysis of text on cohesion and language.” *Behavioral Research Methods, Instruments, and Computers* 36.2 (2004): 193–202.
- Heigham, J., & R. A. Croker. *Qualitative Research in Applied Linguistics: A practical introduction*. London: Palgrave Macmillan, 2009.
- Huhta, A., R. Alanen, M. Tarnanen, M. Martin, & T. Hirvela. “Assessing learners’ writing skills in a SLA study: Validating the rating process across tasks, scales and languages.” *Language Testing* 31.3 (2014): 307–28.
- Hyland, K. *Second language writing*. Cambridge: Cambridge UP, 2003.
- Kim, H. J. “Investigating the construct validity of two types of second language writing tests.” *The Journal of Linguistic Science* 74 (2015): 87–108.
- Knoch, U. “Diagnostic assessment of writing: A comparison of two rating scales.” *Language Testing* 26.2 (2009): 275–304.
- Kobayashi, H., & C. Rinnert. “Factors affecting composition evaluation in an EFL context: Cultural rhetorical pattern and readers’

- background." *Language Learning* 46.3 (1996): 397–437.
- Lee, I. "Teaching coherence to ESL students: A classroom inquiry." *Journal of Second Language Writing* 11.2 (2002): 135–59.
- Li, H., & L. He. "A comparison of EFL raters' essay-rating processes across two types of rating scales." *Language Assessment Quarterly* 12 (2015): 178–212.
- Lumley, T. "Assessment criteria in a large-scale writing test: What do they really mean to the raters?" *Language Testing* 19.3 (2002): 246–76.
- McNamara, D. S., & A. C. Graesser. "Coh-Metrix." *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Ed. P. M. McCarthy & C. Boonthum. Hershey, PA: IGI Global, 2012. 188–205.
- Plakans, L. "Discourse synthesis in integrated second language writing assessment." *Language Testing* 26.4 (2009): 561–87.
- Reed, D. J., & A. D. Cohen. "Revisiting rater and ratings in oral language assessment." *Studies in language testing 11: Experimenting with uncertainty: Essays in honor of Allan Davies*. Ed. T. McNamara, K. O'Loughlin, C. Elder, & A. Brown. Cambridge: Cambridge UP, (2001). 82–96.
- Reid, J. "A computer text analysis of four cohesion devices in English discourse by native and nonnative writers." *Journal of Second Language Writing* 1.2 (1992): 79–107.
- Riazantseva, A. "'I ain't changing anything': A case-study of successful generation 1.5 immigrant college students' writing." *Journal of English for Academic Purposes* 11.3 (2012): 184–93.
- Schoonen, R. "Generalizability of writing scores: An application of structural equation modeling." *Language Testing* 22.1 (2005). 1–30.

- Shi, L. "Native-and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing." *Language Testing* 18.3 (2001): 303-25.

Abstract

## **Investigating Linguistic Differences in Adult EFL Learners' Writing Test Performance**

Kim, Hyun Jung

Lee, Jooseung (Hankuk University of Foreign Studies)

Previous research on L2 writing assessment has tended to focus on quantitative analyses of ratings assigned by raters/teachers. Different from previous research, the present study aimed to examine the linguistic features in test-takers' written responses across different writing task types, focusing on those used for coherence and cohesion in writing. A total of 143 adult EFL learners were presented with two types of writing—compare/contrast and problem/solution—and they were required to write a 300-word essay on one type of their choice. Their written responses (70 compare/contrast and 73 problem/solution) were analyzed using Coh-Metrix to compare the linguistic differences between the two types of writing. Two raters who evaluated both types of writing were further interviewed to examine their perceptions about the discourse organization of the two types. The results indicated that the two types of written responses showed linguistic differences mostly in referential cohesion (e.g., overlap). However, neither of the raters considered referential cohesion when scoring discourse organization; consequently, the use of referential cohesion by the compare/contrast writing group was not reflected in the ratings. The study also suggests the need to consider the time restrictions and task effects on test-takers' relative difficulty in organizing discourse.

**Key Words:** writing test, expository writing, organizational control, linguistic features, Coh-Metrix

쓰기 시험, 설명문, 조직 능력, 언어 특성, 코메트릭스

논문접수일: 2015.11.23

심사완료일: 2015.12.25

게재확정일: 2015.12.27

이름: 김현정 (제 1저자)

소속: 한국외국어대학교 TESOL 대학원

주소: 130-791 서울특별시 동대문구 이문로 107 한국외국어대학교

이메일: hkim@hufs.ac.kr

이름: 이주승 (교신 저자)

소속: 한국외국어대학교 일반대학원

주소: 130-791 서울특별시 동대문구 이문로 107 한국외국어대학교

이메일: leejs@hufs.ac.kr