



An Investigation into the Dimensional Structure of ESL Academic Writing Skills on TOEFL iBT Independent Essays

Youn-Hee Kim

Daegu Catholic University, Korea

This two-phase study examines the ways in which the construct of ESL academic writing is dimensionally determined on Test of English as a Foreign Language™ Internet-based test (TOEFL® iBT) independent essays, using both substantive and statistical methods. In the first phase, four ESL academic writing experts sorted 35 descriptors of ESL academic writing into dimensionally distinct ESL writing skills. In the second phase, a series of conditional covariance-based nonparametric dimensionality tests were conducted using the ratings awarded by 10 ESL teachers on 480 TOEFL iBT independent essays. The results from both substantive and statistical analyses indicated that (a) five writing skills (content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics) represent the construct of the TOEFL iBT independent writing; (b) each of the five writing skills is a statistically significant distinct dimension; and (c) writing skills associated with content fulfillment and organizational effectiveness are dimensionally different from those related to grammatical knowledge, vocabulary use, and mechanics. These findings are used as the basis for a discussion of the need for an analytic scoring system and diagnostic assessment in ESL academic writing.

Keywords: ESL academic writing, ESL writing assessment, ESL writing construct, TOEFL iBT independent writing, dimensionality analysis, conditional covariance-based nonparametric dimensionality test

Introduction

Writing in a second language (L2) is a multi-faceted and complicated language skill. A variety of linguistic and non-linguistic components constitute the construct of L2 writing, and text- and writer-related variables directly or indirectly interact with both writing processes and products. Numerous attempts have been made to define the construct of L2 writing and to assess L2 writing ability, but no all-encompassing framework has yet been described (Cumming, 1998, 2001, 2002; Cumming, Kantor, Powers, Santos, & Taylor, 2000; Grabe, 2001). As Cumming (2001) noted:

Unfortunately, as we all know, there is no generally agreed-on definition of this construct, let alone any substantiated model that is vying for this status. I know all too well myself, from having tried over several years to start to construct, with little empirical success, such a model in one setting (see Cumming & Riazi, 2000). Moreover, in recently reviewing the past 5 years' published research, ... I was only able to affirm that research has recently highlighted the multidimensionality of L2 writing... (p. 214)

This view on the multidimensional nature of L2 writing was highlighted in the development of a framework for the writing subtest of the 2000 Test of English as a Foreign Language (TOEFL). Cumming et al. (2000) framed the test's guiding principle by exploring multiple facets of writing *conception* rather than writing *construct*, so that they can realistically approach what L2 writing ability is.

Despite increasing interest in defining and assessing the construct of L2 writing, few studies have used both substantive and statistical methods to examine the latent dimensional structure of academic writing in English as a second language (ESL). Similarly, researchers have made few attempts to use multiple research techniques to investigate what assessment criteria constitute the construct of ESL academic writing. Most early empirical studies focused on analyzing discourse features of written compositions (e.g., Wolfe-Quintero, Inagaki, & Kim, 1998) or rater perceptions and behaviors in an assessment context (e.g., Cumming, 1990; Cumming, Kantor, & Powers, 2001, 2002; Lumley, 2002; Milanovic, Saville, & Shen, 1996; Sakyi, 2000; Smith, 2000; Vaughn, 1991), and gave little explanation about the ways in which the construct elements are statistically characterized and structured. Indeed, no research to date has taken a statistical approach to examining the underlying factor structure of ESL academic writing ability. This study fills this research gap by examining how the latent dimensional structure of ESL academic writing is determined on Test of English as a Foreign Language™ Internet-based test (TOEFL® iBT) independent essays, using (a) substantive analysis from ESL academic writing experts and (b) a series of conditional covariance-based nonparametric dimensionality techniques. This research is important because if the construct of the TOEFL iBT independent writing can be reliably and validly defined and operationalized, valid inferences can be made about test-takers' ESL academic writing ability. Toward this end, two research questions were constructed and addressed:

- 1) How do ESL writing experts substantively conceptualize the underlying dimensional structure of ESL academic writing on TOEFL iBT independent essays?
- 2) How do statistical analyses determine the underlying dimensional structure of ESL academic writing on TOEFL iBT independent essays?

Although these two research questions are equally important for identifying the dimensional structure of ESL academic writing on TOEFL iBT independent essays, the first research question is somewhat preliminary in that it partially provides substantive data for statistically analyzing the second research question. Therefore, the study's findings will be discussed and synthesized after both of the research questions have been answered.

Literature Review

The Construct of L2 Writing

Attempts to define and characterize the construct of writing have been made from both theoretical and empirical perspectives. In first language (L1) writing research, Grabe and Kaplan (1996) proposed a detailed taxonomy of writing skills, knowledge bases, and processes built on two theories: communicative competence (Bachman, 1990; Canale & Swain, 1980) and ethnography of writing. The taxonomy was developed by identifying such situation variables as settings, tasks, tests and topics, and integrating them with such writer variables as linguistic, discourse, sociolinguistic skills and strategies. It focuses not only on linguistic discourse skills and strategies but also on sociolinguistic aspects of writing ability, with detailed accounts provided for each skill and strategy category. Grabe and Kaplan suggested that this taxonomic approach could provide valuable insights to other researchers, since most writing research is conducted without full consideration of factors that could affect writing processes and outcomes.

Although Grabe and Kaplan's (1996) theoretical model contributed greatly to a general understanding

of how writing is organized and conceptualized, it originated in L1 writing development, a context with limited applications to L2 writing (Grabe, 2001). Acknowledging the absence of L2-specific models of writing, Silva (1990) suggested that (a) L2 writing theory, (b) research on the nature of L2 writing, (c) research on L2 writing instruction, (d) L2 writing instruction theory, and (e) L2 writing instruction practice should all be integrated into such model building.

Cumming (1997) and Leki, Cumming, and Silva (2008) looked at the problem from a somewhat different perspective. Instead of relying on unsubstantiated theories, they presented several empirical approaches for defining and validating L2 writing ability. One approach was to analyze the characteristics of written compositions by utilizing such discourse analytic measures as morphological and syntactic features or lexical and grammatical errors. Another approach focused on rater perceptions and behaviors in order to verify existing rating scales or empirically explore evaluation criteria. These two approaches originated in two different areas of research: second language acquisition (SLA) and language testing (LT), respectively.

Discourse analytic measures or objective measures (such as the number of T-units, error-free clauses per T-unit, etc.) have been increasingly used as a means of assessing the quality of L2 writing and are believed to be a reliable indicator of L2 writing proficiency. These measures have been conceptualized to quantitatively describe observable characteristics or qualities of writing performance by tallying the frequencies or calculating the ratios of certain linguistic features that occur in a written corpus. For example, Wolfe-Quintero et al. (1998) conducted a comprehensive analysis investigating the relationship between L2 writing development and the frequencies, ratios, and indexes of accuracy, fluency, and complexity measures. Along the same lines, other studies have also focused on ways in which textual features that extend across sentence boundaries can be quantified, particularly the extent to which textual structure is tied together in extended discourse (e.g., Halliday & Hasan, 1976; Lautamatti, 1978, 1987; Vande Kopple, 1985; Watson Todd, 1998). Most studies of this type have focused on (a) accuracy, (b) fluency, (c) complexity, (d) cohesion, and (e) coherence in order to conceptualize L2 writing ability, but the research findings indicate that little attention was paid to such non-linguistic aspects of L2 writing as content relevance, effectiveness, originality, or creativity. The results also suggest that a discourse analytic approach focusing on a narrowly defined aspect of grammatical or morphological rules may not be the best way of explaining all the factors that affect L2 writing competence.

Another way to examine L2 writing ability is to look at rater perceptions and rating scales. Most studies in this line of research utilized think-aloud verbal protocols to determine rater scoring behaviors or processes, to empirically explore the assessment criteria that they use, or to verify the accuracy of existing rating scales (e.g., Cumming et al., 2002; Lumley, 2002; Milanovic et al., 1996; Sakyi, 2000; Smith, 2000; Vaughan, 1991). The research findings indicate that there is some consensus about which aspects of L2 writing ability should be evaluated; by and large, raters consistently focused on three elements when assessing ESL essays: (a) content, (b) language use, and (c) organization. However, the ways in which the substance of the construct of L2 writing was referred to varied across research groups. For example, when written content was the focus of raters' assessments, it was called quality of content (Vaughan, 1991), quantity of ideas (Lumley, 2002), or task realization (Milanovic et al., 1996). Organization was also variably referred to as structure (Milanovic et al., 1996) or use of explicit cohesive devices (Lumley, 2002). Language use might be the case that showed a wide range of assessment criteria. It generally included grammatical, lexical, and mechanical features, but the grain size of the features differed drastically. For example, Smith (2000) was less specific than Vaughan (1991), who broke grammatical errors into smaller units such as tense and verb problems.

The construct of L2 writing has also been approached by examining existing rating scales. Rating scales represent the underlying construct of a test and help raters to focus on the skills or abilities intended to be assessed (Luoma, 2004; McNamara, 1996; Weigle, 2002). In a large-scale testing setting, the Test of English as a Foreign Language (TOEFL) is perhaps the best-known of all ESL academic tests. It assesses the writing ability required in an academic setting, while its rating scale scores the overall quality of the writing based on (a) development, (b) organization, and (c) appropriate and precise use of

grammar and vocabulary (Educational Testing Service, 2007).¹ Another well-known ESL proficiency test is the International English Language Testing System (IELTS), in which academic writing tasks are scored based on (a) task achievement, (b) coherence and cohesion, (c) lexical resource, and (d) grammatical range and accuracy (University of Cambridge, British Council, & IELTS Australia, 2007). The Michigan English Language Assessment Battery (MELAB) has similar evaluation criteria: (a) clarity and overall effectiveness, (b) topic development, (c) organization, and (d) range, accuracy, and appropriateness of grammar and vocabulary (University of Michigan, 2003). In a classroom assessment context, the rating scale created by Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981) is the best-known and most comprehensive instrument. It evaluates ESL written compositions based on (a) content, (b) organization, (c) vocabulary, (d) language use, and (e) mechanics with each criterion including fine-grained subcriteria. For example, effectiveness of language use is assessed by the elements associated with syntactic structure, errors of agreement, tense, number, word order/function, articles, pronouns, and prepositions.

In summary, despite different orientations, the three approaches (theoretical taxonomy, discourse analysis, and rater perceptions and rating scales) provide convergent evidence as to how the construct of writing is defined and operationalized. They consistently focus on content, organization, and language use (vocabulary, grammar, and mechanics) as important assessment criteria. Still, future research is warranted that uses a psychometric approach to examine the extent to which these construct elements (i.e., assessment components) are dimensionally distinct. In this regard, this study used conditional covariance-based nonparametric dimensionality procedures to empirically identify a factor structure that can account for the construct of ESL academic writing represented on TOEFL iBT independent essays.

Methodology

Participants

ESL academic writing teachers

Ten experienced ESL teachers were recruited from a language institute in Canada. All ESL teachers were native English speakers with varying experience (2 to 25 years; average 9.2 years) of teaching ESL writing to adult learners. All teachers assessed themselves as familiar with and competent in assessing the written English of non-native English speakers. Seven teachers also reported that they had been trained to assess ESL writing. All ESL teachers received monetary compensation for the time and effort they dedicated to the study.

ESL academic writing experts

Four doctoral students specializing in ESL writing (hereafter referred to as ESL academic writing experts) were recruited from a Second Language Education Program at a research-intensive university in Canada. The ESL writing experts included three males and one female. Two of the experts were native English speakers, while the other two were a native Korean and native Arabic speaker. All of the ESL writing experts had extensive research experience related to feedback on ESL writing, ESL writing conferencing, and the process and assessment of ESL writing. They also had varying experience (3 to 12 years; average 5.5 years) of teaching ESL writing to non-native English speakers at the university level.

¹ Although the TOEFL iBT contains two types of writing tasks (integrated and independent), it is the rating scale for independent writing tasks that is discussed in this article. A discussion of the integrated writing tasks is beyond the scope of the study.

Instruments

TOEFL iBT independent essays

The writing samples used in this study were requested from the Educational Testing Service (ETS) in New Jersey, U.S. ETS administered two forms of the retired TOEFL iBT at various domestic (i.e., United States or Canada) and international test centers in the fall of 2006 and the spring of 2007. ETS provided me with 480 TOEFL iBT independent essays written on two different prompts (240 essays \times 2 prompts) along with additional test-taker background information. The independent writing section of the TOEFL iBT required test-takers to write an essay based on their knowledge and experience with a 30-minute time limit. The two essay prompts were:²

- (a) Do you agree or disagree with the following statement? It is more important to choose to study subjects you are interested in than to choose subjects to prepare for a job or career. Use specific reasons and examples to support your answer.
- (b) Do you agree or disagree with the following statement? In today's world, the ability to cooperate well with others is far more important than it was in the past. Use specific reasons and examples to support your answer.

The TOEFL iBT independent writing test-takers consisted of 480 ESL learners who took the test at domestic or international test centers. Half of the test-takers participated in the TOEFL iBT administration in the fall of 2006, and the other half participated in the spring of 2007. Test-takers were 14 to 51 years of age ($M = 23.61$, $SD = 6.40$), and approximately the same percentage of male and female test-takers participated in each test administration. Test-takers came from 76 different countries and spoke 52 different languages as a first language. Test-takers who spoke Chinese as a first language accounted for the largest number of the test-takers, followed by Korean, Spanish, and Japanese. Test-takers' primary reason for taking the TOEFL was to enter a college or a university as either an undergraduate student (18.13%) or a graduate student (21.04%).

The ESL academic writing assessment checklist

An ESL academic writing assessment checklist was developed for the study. The purpose of the checklist was to assess non-native English-speaking students' ESL writing ability in an academic context. The checklist was developed using an empirical method to include concrete, fine-grained descriptors characterizing ESL academic writing ability. Nine ESL teachers participated in individual think-aloud sessions in which they verbalized their thought processes while providing detailed feedback on 10 TOEFL iBT independent essays. Before the session, I trained the teachers to think aloud *constantly* while they were reading and providing feedback on the essays, instead of planning what they were going to say or explaining to me what they were saying. I emphasized that it was important for them to keep talking with no long intervals of silence. When they were silent for any length of time, I reminded them to keep thinking aloud (for more information, see Kim, 2010).

The teachers' verbal accounts provided rich descriptions of ESL academic writing ability represented on TOEFL iBT independent essays and served as the base for constructing the pool of descriptors. Recorded verbal data were transcribed in full and then reviewed iteratively to identify the distinct ESL academic writing subskills and textual features that would constitute the descriptors. Grounded theory (Glaser & Strauss, 1967) was the principle methodology used to determine the emerging descriptors with varied properties and dimensions. Transcripts ranged from 5,422 to 8,504 words (i.e., from 9 to 17 single-spaced typed pages) per teacher. Each transcript was read, categorized, and segmented into meaningful

² TOEFL[®] test materials are reprinted by permission of Educational Testing Service, the copyright owner.

units using the computer program *NVivo 8* (QSR, 2008). The unit of analysis was one distinct evaluation theme that characterized ESL writing subskills and textual features, and each evaluation theme represented one distinct descriptor. The analysis of the transcripts resulted in a final total of 1,715 segments represented by 39 descriptors.

Four ESL writing experts participated in a focus group meeting to review the 39 descriptors elicited from the teachers' think-aloud verbal data. They were provided with six TOEFL iBT independent essays (3 essays \times 2 prompts) along with the descriptors. Once the experts had read the essays and had a general understanding of the writing context, they were asked to review each descriptor and to discuss whether it was clear/not clear, redundant/non-redundant, useful/useless or relevant/irrelevant to ESL academic writing. When necessary, the teachers' think-aloud transcripts were made available to them to have a better understanding of the ways in which the descriptors were elicited. After examining each descriptor, they were also asked whether the descriptor pool was comprehensive enough to cover all aspects of ESL academic writing. Any missing theoretical aspects were added to the descriptor pool based upon existing theories of ESL academic writing. The review and refinement process of the four ESL academic writing experts eliminated three descriptors and merged two descriptors into one, resulting in a final total of 35 descriptors (see Table 1). These 35 descriptors constituted the ESL academic writing assessment checklist accompanied by a *yes* or a *no* response option (see Appendix).

Descriptor index cards

A set of 35 index cards was prepared with one descriptor reproduced on each card. The descriptor number was not shown on the index card so that the order the descriptors were listed in the checklist would not affect the writing experts' sorting behavior.

The sorting scheme

A sorting scheme was developed based on both empirical and theoretical grounds. The teachers' think-aloud verbal protocols were used as guiding empirical sources, complemented by theories that define and assess the construct of ESL academic writing. I read each descriptor iteratively in order to identify the writing skills that best represented the characteristics of the descriptors. When the skills were empirically identified, they were sequentially compared and confirmed according to theories of ESL writing and a variety of existing ESL writing assessment schemes. During the sorting scheme finalization process, the following was taken into consideration: (a) writing skills should be conceptually distinct from each other; (b) each writing skill should have a minimum number of descriptors (i.e., more than three descriptors) in order to be considered for the statistical dimensionality analysis; and (c) writing skills should be comparable to those specified in existing ESL writing assessment schemes for cross-validation purposes. The sorting scheme created through this process included five skills: content fulfillment (CON), organizational effectiveness (ORG), grammatical knowledge (GRM), vocabulary use (VOC), and mechanics (MCH). These writing skills were consistent with the assessment components discussed in the previous section (see Table 1) and consistent with the assessment criteria described in Jacobs et al.'s (1981) scale.

TABLE 1
Descriptors of ESL Academic Writing

Descriptor
1. This essay answers the question.
2. This essay is written clearly enough to be read without having to guess what the writer is trying to say.
3. This essay is concisely written and contains few redundant ideas or linguistic expressions.
4. This essay contains a clear thesis statement.
5. The main arguments of this essay are strong.
6. There are enough supporting ideas and examples in this essay.
7. The supporting ideas and examples in this essay are appropriate and logical.
8. The supporting ideas and examples in this essay are specific and detailed.
9. The ideas are organized into paragraphs and include an introduction, a body, and a conclusion.
10. Each body paragraph has a clear topic sentence tied to supporting sentences.
11. Each paragraph presents one distinct and unified idea.
12. Each paragraph is connected to the rest of the essay.
13. Ideas are developed or expanded well throughout each paragraph.
14. Transition devices are used effectively.
15. This essay demonstrates syntactic variety, including simple, compound, and complex sentence structures.
16. This essay demonstrates an understanding of English word order.
17. This essay contains few sentence fragments.
18. This essay contains few run-on sentences or comma splices.
19. Grammatical or linguistic errors in this essay do not impede comprehension.
20. Verb tenses are used appropriately.
21. There is consistent subject-verb agreement.
22. Singular and plural nouns are used appropriately.
23. Prepositions are used appropriately.
24. Articles are used appropriately.
25. Pronouns agree with referents.
26. Sophisticated or advanced vocabulary is used.
27. A wide range of vocabulary is used.
28. Vocabulary choices are appropriate for conveying the intended meaning.
29. This essay demonstrates facility with appropriate collocations.
30. Word forms (noun, verb, adjective, adverb, etc) are used appropriately.
31. Words are spelled correctly.
32. Punctuation marks are used appropriately.
33. Capital letters are used appropriately.
34. This essay contains appropriate indentation.
35. Appropriate tone and register are used throughout the essay.

Data Collection Procedure

ESL academic writing experts' sorting activity

Four ESL academic writing experts were invited to individual meetings to sort the 35 descriptors into dimensionally distinct ESL writing skills. The purpose of this activity was to substantively identify the underlying structure of the ESL academic writing represented on TOEFL iBT independent essays. Two phases preceded this sorting activity: first, I asked each writing expert to come up with his own skill identification scheme while sorting the descriptors; second, I asked him to sort the descriptors using the predetermined sorting scheme consisting of CON, ORG, GRM, VOC and MCH. The first sorting activity took place before the second so that the predetermined sorting scheme might not influence the expert's

own skill identification.

During the first phase, each writing expert received a set of index cards on which the 35 descriptors were reproduced. I asked the experts first to skim through these cards and then to sort them freely into piles that they thought represented distinct ESL writing skills. This first sorting activity was based solely on the experts' own skill configuration, without being influenced by a predetermined set of categories. When they thought a descriptor was associated with multiple writing skills, they labeled them as primary or secondary. In the second sorting activity, the experts used the predetermined sorting scheme to assign the descriptors to appropriate skills. I explained the sorting scheme to the experts and then asked whether they thought the five writing skills (content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics) were comprehensive enough to represent the content of all the descriptors. I did not provide detailed definitions or descriptions of each skill category, so that the experts' mapping of the descriptors onto skills was not restricted. I also asked the experts to mark those descriptors that matched with multiple or none of the skills. The sorting activity lasted approximately one hour for each expert, and their verbal accounts were tape recorded with their permission. I did not transcribe all the recorded verbal data because their primary purpose was to confirm the outcomes of the experts' sorting activities. I consulted the verbal data only when it was unclear how an expert had assigned a specific descriptor to a specific skill.

ESL academic writing teachers' essay assessment

The purpose of essay assessment was to generate data that would be used for statistical dimensionality analyses. Ten ESL teachers assessed 480 TOEFL iBT independent essays using the ESL academic writing assessment checklist. Rater training for the teachers was held prior to their assessment of the essays. I set up an individual meeting with each of the teachers to explain the purpose of the study and to outline the checklist development procedure in greater detail. Each descriptor was explained using concrete examples; the *yes* or *no* option was also explained, and the difficulty of determining a cut-off for *yes* or *no* was acknowledged. The general rule of thumb was that if a teacher thought that an essay *generally* had the feature specified by the descriptor, it was considered a *yes*; otherwise it was considered a *no*. The training was informal in order to minimize potential psychological pressure that might affect the teachers' assessment. The number of essays assigned to each teacher was determined by the individual's availability. Four teachers assessed 48 essays; three teachers assessed 56 essays; two teachers assessed 24 essays; and one teacher assessed 72 essays. When the assessment was completed, I collected all of the assessment materials for security purposes. The teachers' score data were compiled in Microsoft® Excel spreadsheets, with a *yes* response entered as a 1 and a *no* response as a 0.

Data Analysis Procedure

Substantive dimensionality analysis

The substantive dimensionality analysis was carried out based upon the outcomes of the ESL academic writing experts' descriptor sorting activities. I collected each of the expert's sorting outcomes and created a matrix that could represent the skill-by-descriptor relationships. When creating this matrix, I constantly identified and compared the ways in which each of the four experts categorized the 35 descriptors into dimensionally distinct ESL writing skills. Because the experts' sorting activity consisted of two phases, it resulted in two final matrices. The skills-by-descriptors correspondence was then used for statistical dimensionality analyses.

Statistical dimensionality analysis

The statistical dimensionality analysis was conducted in both exploratory and confirmatory manners

using the ratings awarded by 10 ESL teachers on 480 TOEFL iBT independent essays. In an exploratory DIMTEST analysis, AT items were selected using its built-in program, ATFIND, and were tested against the remaining PT items several times until the DIMTEST failed to reject the null hypothesis. Each time the null hypothesis was rejected, the initial AT items were removed from the next run. The magnitude of multidimensionality was also examined in an exploratory DETECT analysis. An exploratory CCPROX/HCA procedure further informed the dimensional structure of the data. The results from the CCPROX/HCA analysis were used to develop a hypothesis of dimensionality and to identify items that could be used as an AT set. DIMTEST was then conducted iteratively with varying AT sets in a confirmatory manner. The findings from the three different methods were synthesized in order to determine the dimensional structure of ESL academic writing ability.

Results

How Do ESL Writing Experts Substantively Conceptualize the Underlying Dimensional Structure of ESL Academic Writing on TOEFL iBT Independent Essays?

The results of the first sorting activity showed that each ESL academic writing expert had a unique conceptualization of the ESL writing skills represented on TOEFL iBT independent essays and created a unique skill categorization scheme. Anthony categorized descriptors into (a) idea development, (b) organization, (c) language use, (d) vocabulary, and (e) punctuation, noting that idea development is associated with the meaning of written text, while organization is focused on the form of written text. Anthony also suggested subdividing the language use category into global and local levels. Jane's categorization was particularly interesting, as she conceptualized ESL writing skills from a hierarchical perspective, with a skill identification scheme layered according to its (a) word, (b) sentence, (c) paragraph, and (d) essay components. From a slightly different perspective, Gary divided the descriptors into five categories: (a) organization, (b) grammar, (c) vocabulary, (d) style, and (e) formatting, commenting that vocabulary knowledge is closely related to writing style, and that all ESL writing skills are intertwined with each other. Gary also argued for a holistic interpretation of writing, pointing out that it was difficult to analytically distinguish one skill from the others, and that the unitary nature of writing ability cannot be accounted for by the sum of its parts. Alex's categorization scheme was similar to those of Anthony and Gary, incorporating (a) content, (b) organization, (c) grammar, and (d) mechanics.

Table 2 shows the ways in which the experts related the descriptors to their own skill categories. Overall, their skills-by-descriptors correspondence indicates that the experts related the first half of the descriptors (D01-D14) to such global assessment components as idea development, organization, content, essay, and paragraph, and the second half of the descriptors (D15-D35) to such local assessment components as language use, grammar, vocabulary, mechanics, punctuation, sentence, and word. This suggests that the experts' skill configurations were comprehensive and comparable to the sorting scheme used in this study.

TABLE 2
Result of the First Sorting Activity

Descriptor	Anthony	Jane	Gary	Alex
D01	Idea development	Essay	Organization	Content
D02	Language use	Essay	Organization	Organization
D03	Idea development	Essay	Style	Organization
D04	Organization, Idea development	Essay	Organization	Organization
D05	Idea development	Essay	Organization	Content
D06	Idea development	Essay	Organization	Content
D07	Idea development	Essay	Organization	Organization
D08	Idea development	Essay	Organization	Content
D09	Organization	Essay	Organization	Organization
D10	Organization, Idea development	Paragraph	Organization	Organization
D11	Organization	Paragraph	Organization	Organization
D12	Organization	Paragraph	Organization	Organization
D13	Idea development	Paragraph	Organization	Content
D14	Organization	Sentence	Organization	Organization, Grammar
D15	Language use	Sentence	Style	Grammar
D16	Language use	Word	Grammar	Grammar
D17	Language use	Sentence	Grammar	Grammar
D18	Language use	Sentence	Grammar	Grammar
D19	Language use	Essay	Grammar	Grammar
D20	Language use	Word	Grammar	Grammar
D21	Language use	Word	Grammar	Grammar
D22	Language use	Word	Grammar	Grammar
D23	Language use	Word	Grammar	Grammar
D24	Language use	Word	Grammar	Grammar
D25	Language use	Word	Grammar	Grammar
D26	Vocabulary	Word	Vocabulary	Grammar
D27	Vocabulary	Word	Vocabulary	Grammar, Organization
D28	Vocabulary	Word	Vocabulary	Organization
D29	Vocabulary	Word	Grammar	Grammar
D30	Language use	Word	Grammar	Grammar
D31	Punctuation	Word	Grammar	Mechanics
D32	Punctuation	Word	Grammar	Mechanics
D33	Punctuation	Word	Grammar	Mechanics
D34	Punctuation	Word	Formatting	Mechanics
D35	Idea development	Essay	Style	Grammar

Note 1. The notation, “D + number” indicates “Descriptor + number.” Therefore, D01 indicates Descriptor 01.

Note 2. When multiple skills are assigned to a descriptor, a primary skill appears before a secondary skill.

Table 3 shows how the experts related the descriptors to a specific writing skill using the predetermined sorting scheme during the second sorting activity. Prior to the second sorting task, I asked the four writing experts whether they thought the five predetermined writing skills (content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics) were comprehensive enough to represent all the descriptors’ content. They generally agreed that the five writing skills accurately described the descriptors’ characteristics, and that they represented the construct of the TOEFL iBT independent writing. I examined the experts’ agreement on the skills-by-descriptors relationship and found that they agreed completely on 20 out of the 35 descriptors. The areas in which the experts showed the most discrepancy were D02, D03, D14, and D35, descriptors that focused on a holistic assessment of an essay’s general quality. In particular, they disagreed considerably on D35, which assessed an essay’s tone and register, because they could have been mastered by appropriate use of vocabulary or the consistent interplay of all aspects of ESL writing skills throughout the essay.

TABLE 3
Result of the Second Sorting Activity

Descriptor	Anthony	Jane	Gary	Alex
D01	CON	CON	CON	CON
D02	CON, GRM,ORG	CON, ORG	ORG	ORG
D03	CON	CON,VOC,ORG	VOC	GRM,ORG,CON
D04	ORG	CON, ORG	CON	ORG,CON
D05	CON,ORG	CON	CON	CON
D06	CON	CON	CON	CON
D07	CON	CON	ORG	ORG
D08	CON	CON	CON	CON
D09	ORG	ORG	ORG	ORG
D10	ORG	ORG	ORG	ORG
D11	ORG,CON	ORG	ORG	ORG
D12	ORG	ORG	ORG	ORG
D13	CON	ORG	ORG	CON
D14	ORG	ORG	ORG	ORG,GRM,VOC
D15	GRM	GRM	GRM	GRM
D16	GRM	GRM	GRM	GRM
D17	GRM	GRM,MCH	GRM	GRM
D18	GRM	GRM,MCH	GRM	GRM
D19	GRM	GRM	GRM	GRM
D20	GRM	GRM	GRM	GRM
D21	GRM	GRM	GRM	GRM
D22	GRM	GRM	GRM	GRM
D23	GRM	GRM	GRM	GRM
D24	GRM	GRM	GRM	GRM
D25	GRM	GRM	GRM	GRM
D26	VOC	VOC	VOC	VOC
D27	VOC	VOC	VOC	VOC
D28	VOC	VOC	VOC	VOC
D29	VOC	VOC	GRM	GRM
D30	GRM	GRM	VOC	GRM
D31	GRM	MCH	MCH	MCH
D32	MCH	MCH	MCH	MCH
D33	GRM	MCH	MCH	MCH
D34	MCH	MCH	MCH	MCH
D35	VOC,ORG	CON,GRM,VOC	VOC	GRM,MCH

Table 4 presents each skills-by-descriptors correspondence and the skill definitions. I took all of the experts' opinions when determining each skills-by-descriptors correspondence. If the experts assigned a different skill or multiple skills to a descriptor, I consulted relevant ESL writing literature as a final judgment call. The literature provided useful theoretical grounds for identifying the most appropriate skills-by-descriptors correspondence. As Table 4 shows, grammatical knowledge included the greatest number of descriptors, followed by content fulfillment, organizational effectiveness, vocabulary use, and mechanics. Since students greatly desire feedback on grammatical problems in their writing (Cohen & Cavalcanti, 1990; Ferris, 1995, 2002; Hedgecock & Lefkowitz, 1994; Leki, 1991), the large number of grammar descriptors is reasonable; however, the relatively small number of descriptors for vocabulary use and mechanics was somewhat problematic because the small number of descriptors per skill could cause instability of dimensionality analysis. After completing the skills-by-descriptors correspondence, I

determined a definition of each skill based upon the characteristics of the descriptors assigned to that skill (see the last column of Table 4).

TABLE 4
Description of the Skills-by-Descriptors Relationship

Writing skill	No. of descriptors	Descriptor	Description
Content fulfillment (CON)	8	1, 2, 3, 4, 5, 6, 7, 8	Content fulfillment assesses the degree to which a writer satisfactorily addresses a given topic. A writer who shows strength in this area generally demonstrates an excellent understanding of the topic by presenting clear and substantial arguments supported by specific examples.
Organizational effectiveness (ORG)	6	9, 10, 11, 12, 13, 14	Organizational effectiveness assesses the way in which a writer organizes and develops his or her ideas. A writer who is competent in this area generally demonstrates the ability to construct and develop a paragraph effectively and to connect textual elements well within and between paragraphs using appropriate cohesive and transitional devices.
Grammatical knowledge (GRM)	12	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30	Grammatical knowledge assesses the extent to which a writer demonstrates consistent ability to properly apply the rules of English grammar. A well-written essay adheres to English grammar rules with full flexibility and accuracy, and displays a variety of syntactic structures and few linguistic errors.
Vocabulary use (VOC)	5	26, 27, 28, 29, 35	Vocabulary use assesses the extent to which a writer demonstrates great sophistication and variety of vocabulary knowledge. A writer who is strong in this area generally uses a broad range of sophisticated words, knows how to combine words, and displays accurate knowledge of word form and usage.
Mechanics (MCH)	4	31, 32, 33, 34	Mechanics assesses the extent to which a writer follows the conventions of English academic writing. A writer who is strong in this area generally demonstrates correct use of spelling, punctuation, capitalization, and indentation.

How Do Statistical Analyses Determine the Underlying Dimensional Structure of ESL Academic Writing on TOEFL iBT Independent Essays?

An exploratory DIMTEST analysis resulted in the rejection of the null hypothesis of unidimensionality with an extremely small p -value, $T = 7.28$, $p < .00$. Twelve descriptors were selected as an initial AT set by the program, including six CON descriptors (D01, D03, D04, D05, D07, and D08) and six ORG descriptors (D09, D10, D11, D12, D13, and D14). The subsequent exploratory DIMTEST analysis failed to reject the null hypothesis, suggesting that CON and ORG skills are dimensionally distinct from GRM, VOC, and MCH skills.

I then performed DETECT in an exploratory manner in order to estimate the number of dimensions present in the data and the magnitude of the multidimensionality. As Table 5 shows, the descriptors were separated into four clusters maximizing the DETECT index. Consistent with the results of the exploratory DIMTEST analysis, the first cluster was composed of the CON and ORG descriptors (D01-D14). The second cluster consisted of two GRM descriptors (D15 and D17) and two VOC descriptors (D26 and

D27). The third cluster was composed of a mixture of GRM, VOC, and MCH descriptors. Finally, the fourth cluster included only two descriptors (D34 and D35). The DETECT value was noticeably large (DETECT index = 1.25), indicating strong evidence of multidimensionality. In addition, the IDN and r indices were close to 1 (IDN index = 0.82 and r index = 0.79), indicating that the approximate simple structure held true for the data.

TABLE 5
Descriptor Clusters Identified by DETECT

Cluster	Descriptor
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
2	15, 17, 26, 27
3	16, 18, 19, 20, 21, 22, 23, 24, 25, 28, 29, 30, 31, 32, 33
4	34, 35

I also performed an exploratory CCPROX/HCA analysis in order to visually examine the most interpretable cluster solution in the data. Figure 1 displays a part of the CCPROX/HCA output from Levels 18 to 34. Visual inspection suggests that the five-cluster solution is likely to be the most interpretable. From Level 22, the CON descriptors began to form one large cluster without being disjointed by other skill descriptors. The ORG and VOC descriptors also formed two distinct clusters from the early stage of the HCA solution. Although the GRM and MCH descriptors showed some fuzzy areas within their clusters, they appeared to possess separate dimensions.

I tested the dimensional hypothesis developed using exploratory methods with a confirmatory DIMTEST, which examined whether the data were multidimensional and manifested the five identified writing skills (CON, ORG, GRM, VOC, and MCH). As Table 6 shows, the five DIMTEST runs had five rejections of the null hypothesis of unidimensionality, indicating that the five skills are statistically significant distinct dimensions, $p < .01$. The dimensionality test statistics, T , further indicated that CON, ORG, and GRM had a greater magnitude of multidimensionality than VOC and MCH.

TABLE 6
Confirmatory DIMTEST Results

Writing skill	No. of descriptors	T	p
CON	8	6.0456	0.00
ORG	6	8.5905	0.00
GRM	12	6.4971	0.00
VOC	5	3.3222	0.00
MCH	4	5.3902	0.00

A set of exploratory and confirmatory dimensionality analyses showed that the ESL academic writing ability represented on TOEFL iBT independent essays could comprise five distinct skills, CON, ORG, GRM, VOC, and MCH, and provided some support for the claim that the underlying dimensional structure of CON and ORG differs distinctly from that of GRM, VOC, and MCH. This result is consistent with theoretical accounts of ESL academic writing that characterize writing ability as a constellation of multiple skills. The research findings therefore support the use of an analytic scoring system for ESL academic writing. Although analytic rating scales have been used widely to assess ESL students' writing, little statistical evidence supports the division of assessment components. Most analytic rating scales provide a set of assessment criteria with little explanation of whether each criterion is independent of the others. In this regard, these research findings are important because they provide statistical evidence that accounts for the underlying structure of ESL academic writing assessment criteria. The statistical dimensionality results also confirmed that the descriptor classification provided by the ESL academic writing experts was accurate, which supports the results of the substantive analysis.

Level of hierarchical cluster:

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
CON																
CON																
CON																
***	***	***	***	CON												
CON																
CON																
***	***	CON	CON	***	***	***	***	***	ORG							
CON	CON	***	***	ORG												
***	***	ORG														
ORG																
ORG																
ORG																
ORG	***	***	***	***	***	***	***	VOC								
ORG	ORG	ORG	ORG	***	***	***	***	***	VOC							
ORG	ORG	***	***	VOC												
***	***	VOC	GRM													
VOC	***	***	GRM	GRM	GRM	GRM	GRM	GRM								
VOC	VOC	VOC	VOC	***	***	***	***	***	GRM	GRM	***	***	GRM	GRM	GRM	GRM
VOC	VOC	***	***	GRM	GRM	GRM	GRM	GRM	***	***	GRM	GRM	GRM	GRM	GRM	CON
***	***	GRM	GRM	***	***	***	***	***	GRM	GRM	GRM	GRM	CON	CON	CON	GRM
GRM	GRM	***	***	GRM	CON	CON	GRM	GRM	GRM	CON						
***	***	GRM	CON	CON	GRM	GRM	***	***	CON	GRM						
GRM	GRM	GRM	GRM	***	***	***	CON	CON	GRM	GRM	***	***	CON	CON	GRM	VOC
GRM	GRM	***	***	CON	CON	CON	GRM	GRM	***	***	CON	CON	GRM	GRM	VOC	GRM
***	***	CON	CON	GRM	GRM	GRM	***	***	CON	CON	GRM	GRM	VOC	VOC	GRM	GRM
CON	CON	GRM	GRM	***	***	***	CON	CON	GRM	GRM	VOC	VOC	GRM	GRM	GRM	GRM
GRM	GRM	***	***	CON	CON	CON	GRM	GRM	VOC	VOC	GRM	GRM	GRM	GRM	GRM	GRM
***	***	CON	CON	GRM	GRM	GRM	***	VOC	GRM							
CON	CON	GRM	GRM	***	***	***	VOC	GRM								
GRM	GRM	***	***	VOC	VOC	VOC	GRM	GRM	***	GRM						
***	***	VOC	VOC	GRM	GRM	GRM	GRM	GRM	***	GRM	GRM	GRM	GRM	GRM	GRM	MCH
VOC	VOC	GRM	***	GRM	GRM	GRM	GRM	GRM	GRM	MCH						
GRM	GRM	***	GRM	***	***	***	GRM	GRM	GRM	GRM	GRM	GRM	***	MCH	MCH	MCH
***	***	GRM	***	GRM	***	***	MCH	MCH	MCH							
GRM	GRM	***	GRM	***	MCH	MCH	MCH	MCH	MCH	VOC						
***	***	GRM	***	MCH	MCH	MCH	MCH	MCH	VOC							
GRM	GRM	GRM	GRM	***	***	GRM	GRM	***	MCH	MCH	MCH	MCH	MCH	MCH	VOC	
GRM	GRM	GRM	***	GRM	GRM	GRM	***	MCH	MCH	MCH	***	MCH	VOC			
***	GRM	***	GRM	GRM	GRM	***	MCH	MCH	MCH	***	MCH	VOC				
GRM	***	GRM	GRM	***	***	MCH	MCH	MCH	***	MCH	VOC					
***	GRM	GRM	***	MCH	MCH	MCH	MCH	***	MCH	VOC						
GRM	GRM	***	MCH	MCH	MCH	MCH	***	MCH	VOC							
***	MCH	MCH	MCH	***	***	MCH	VOC									
MCH	MCH	***	MCH	***	VOC											
MCH	***	MCH	***	VOC												
***	MCH	***	VOC													
MCH	***	VOC														
***	VOC															
VOC																

Figure 1. CCPROX/ HCA results.

Discussion and Conclusions

This study used substantive and statistical procedures to examine how the latent dimensional structure of ESL academic writing is determined on TOEFL iBT independent essays. A substantive analysis from four ESL academic writing experts indicated that although the experts exhibited a different conceptualization of the TOEFL independent writing skills when creating their own skill taxonomies, they generally agreed that the five skills (content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics) represented the multidimensional nature of ESL writing. The experts also assigned the same skills to a majority of the descriptors (20 out of 35), showing a high agreement rate on the skills-by-descriptors relationship; however, there were some discrepancies with regard to descriptors that focused on holistic assessment of an essay's general quality, possibly because it is difficult to determine exactly which skills the descriptors associated with the global aspect of writing assess.

A series of conditional covariance-based nonparametric dimensionality techniques also supported the claim for the multidimensionality of ESL academic writing. Not only did the DETECT analysis provide strong evidence of multidimensionality, but the CCPROX/HCA and confirmatory DIMTEST analyses further indicated that each of the five writing skills is a statistically significant distinct dimension. The exploratory DIMTEST and DETECT analyses also reported that skills related to content fulfillment and organizational effectiveness are dimensionally different from skills associated with grammatical knowledge, vocabulary use, and mechanics. This result can be interpreted such that content fulfillment and organizational effectiveness are more closely related to the global aspect of ESL writing, while grammatical knowledge, vocabulary use, and mechanics are associated with more specific, local, and linguistic aspects of ESL writing.

Overall, the findings of this study suggest that ESL academic writing can be approached from both macro- and micro-level perspectives. From a macro-level perspective, ESL academic writing is bi-divisible, characterizing global (content fulfillment, organizational effectiveness) and local (grammatical knowledge, vocabulary use, mechanics) writing skills. From a micro-level perspective, it is also possible to argue that ESL academic writing is multi-divisible, encompassing such varied writing skills as content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics.

The findings of this study support the need for an analytic rating system and diagnostic assessment in ESL academic writing. Analytic rating scales assume that the sum of the separate scores awarded to subcomponents of writing is equal to a single score awarded for the entire written piece (Goulden, 1992). In this scoring system, raters take note of several aspects of writing and produce multiple ratings or subscores, which are then weighed according to theoretical considerations or the test developer's specifications. In a holistic scoring system, on the other hand, raters assign a single score that best reflects their general impression of written text assuming that language ability is a single unitary ability (Bachman & Palmer, 1996) and that a score for the whole is not equal to the sum of separate scores for the parts (Goulden, 1992). While a single, holistic score can obscure variations in writing ability, analytic ratings make them visible, enabling the construction of writing profiles (Hamp-Lyons & Henning, 1991). Profile scores are particularly helpful for L2 writers, who are more likely than their L1 counterparts to show an uneven or marked profile across different areas of writing ability (Hamp-Lyons, 1991; Weigle, 2002). The ratings assigned to each component of writing can also be used to diagnose the relative strengths and weaknesses of written texts (Bachman & Palmer, 1996; Hamp-Lyons, 1991; Hamp-Lyons & Henning, 1991; Weigle, 2002). The dimensionality information gathered from this study could therefore be useful to scale developers who wish to know what needs to be assessed in ESL writing assessment, and could then be used to make informed decisions about the implementation of diagnostic assessment in ESL academic writing.

Care must be taken in interpreting the study's results, however. While the ESL academic writing experts were conducting a substantive analysis, they were provided with a sorting scheme outlining the skills that were believed to account for the construct of the ESL academic writing represented on TOEFL

iBT independent essays. It is not known how the latent dimensional structure of the TOEFL iBT independent writing would be determined if a different skill categorization scheme was used. For example, if Jane's skill taxonomy describing word, sentence, paragraph, and essay components of ESL academic writing was used, a different dimensional structure might have resulted. Future research is recommended in order to examine variability in the dimensional structure according to different skill categorization schemes.

It would also be interesting to statistically validate the evaluation criteria elicited from think-aloud verbal protocols in other studies. Previous studies in this line of research (e.g., Cumming et al., 2002; Lumley, 2002; Milanovic et al., 1996; Sakyi, 2000; Smith, 2000; Vaughan, 1991) have suggested that raters consider a varying range of assessment criteria when assessing ESL essays. For example, Milanovic et al. (1996) identified a wide range of evaluation elements, including (a) length, (b) legibility, (c) grammar, (d) structure, (e) communicative effectiveness, (f) tone, (g) vocabulary, (h) spelling, (i) content, (j) task realization, and (k) punctuation, while Sakyi (2000) sought more global assessment criteria, focusing on (a) content and organization, (b) grammatical and mechanical errors, and (c) sentence structure and vocabulary. Although these studies qualitatively verified the skills on which ESL raters focused while assessing essays, little statistical evidence has been provided as to whether assessment components are dimensionally distinct from each other and can function as stand-alone criteria. If they do not provide a unique contribution to defining and assessing ESL academic writing, the validity of such qualitative research findings would be threatened. It would therefore be worthwhile to examine the extent to which the assessment criteria that were qualitatively validated can be supported by additional quantitative evidence.

Several methodological limitations should be noted. First, this study's results cannot be overgeneralized for use in other ESL academic writing contexts, although they can be useful to ESL writing teachers in general. The ESL writing samples utilized in this study included only TOEFL iBT independent essays, and limiting the research findings to a specific writing context (i.e., a large-scale assessment context) would make interpretations of the study more valid. If different types of ESL academic writing samples had been used, different kinds of writing skills might have been identified. It should also be noted that the construct of ESL academic writing can be conceptualized differently from different perspectives. As the four ESL academic writing experts' own skill categorizations showed, the construct of ESL academic writing can be characterized in different ways, focusing on different assessment components. It is therefore still possible to construct different sets of ESL academic writing skills.

Acknowledgements

This work was supported by research grants from Daegu Catholic University in 2017. The responsibility for any errors that remain is solely mine, and any opinions expressed in this article are not necessarily those of Educational Testing Service.

The Author

Youn-Hee Kim is an associate professor in the Department of English Education at Daegu Catholic University, South Korea. Her primary research interests are second language assessment and testing. She has published numerous papers in *Applied Linguistics*, *Language Learning*, and *Language Testing*, and has been working as a reviewer of international journals such as *Applied Linguistics*, *Language Assessment Quarterly*, *Language Testing*, and *Language Learning*.

Department of English Education

Daegu Catholic University
13-13 Hayang-ro, Hayang-eup
Kyungsan-si, Kyungpook
South Korea, 38430
Tel: +82-53-850-3133
Email: younkim@cu.ac.kr

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Cohen, A. D., & Cavalcanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155-177). Cambridge: Cambridge University Press.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7, 31-51.
- Cumming, A. (1997). The testing of writing in a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Volume 7 Language testing and assessment* (pp. 51-63). Dordrecht, Netherlands: Kluwer.
- Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics*, 18, 61-78.
- Cumming, A. (2001). The difficulty of standards, for example in L2 writing. In T. Silva & P. Matsuda (Eds.), *On second language writing* (pp. 209-229). Mahwah, NJ: Lawrence Erlbaum.
- Cumming, A. (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas. *Assessing Writing*, 8, 73-83.
- Cumming, A., Kantor, R., & Powers, D. E. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework*. TOEFL Monograph Series 22. Princeton, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal*, 86, 67-96.
- Cumming, A., Kantor, R., Powers, D. E., 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., & Riazi, A. M. (2000). Building models of adult second language writing instruction. *Learning and Instruction*, 10, 55-71.
- Educational Testing Service (ETS). (2007). *TOEFL iBT Tips: How to prepare for the TOEFL iBT*. Princeton, NJ: Educational Testing Service. Retrieved July 3, 2008, from http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf
- Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29, 33-53.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor, MI: University of Michigan Press.
- Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago; IL: Aldine.
- Goulden, N. R. (1992). Theory and vocabulary for communication assessments. *Communication Education*, 41, 258-269.
- Grabe, W. (2001). Notes toward a theory of second language writing. In T. Silva & P. Matsuda (Eds.), *On second language writing* (pp. 39-57). Mahwah, NJ: Lawrence Erlbaum.

- Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing*. New York: Longman.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-276). Norwood, NJ: Ablex.
- 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.
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing, 3*, 141-163.
- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Kim, Y-H. (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.
- Lautamatti, L. (1978). Observations on the development of the topic in simplified discourse. In V. Kohonen & N. E. Enkvist (Eds.), *Text linguistics, cognitive learning, and language teaching* (pp. 71-104). Turku, Finland: Afinfla .
- Lautamatti, L. (1987). Observations on the development of the topic of simplified discourse. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley.
- Leki, I. (1991). The preferences of ESL students for error correction in college level writing classes. *Foreign Language Annals, 24*, 203-218.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. New York, NY: Routledge.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing, 19*, 246-276.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- Milanovic, M., Saville, N., & Shen, S. (1996). A study of the decision making behavior of composition markers. In M. Milanovic & N. Saville (Eds.), *Studies in language testing 3: Performance testing, cognition and assessment* (pp. 92-111). Cambridge: Cambridge University Press.
- QSR. (2008). *NVivo 8: Getting started*. QSR International.
- Sakyi, A. A. (2000). Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium* (pp. 129-152), Orlando, Florida. Cambridge: Cambridge University Press.
- Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing* (pp. 11-23). Cambridge: Cambridge University Press.
- Smith, D. (2000). Rater judgments in the direct assessment of competency-based second language writing ability. In G. Brindley (Ed.), *Studies in immigrant English language assessment* (pp. 159-189). Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.
- University of Cambridge, British Council, & IELTS Australia. (2007). *IELTS Handbook 2007*. University of Cambridge, British Council, and IELTS Australia: UK.
- University of Michigan. (2003). *Michigan English Language Assessment Battery: Technical manual 2003*. Testing and Certificate Division, English Language Institute, University of Michigan, Ann Arbor.
- Vande Kopple, W. J. (1985). Some exploratory discourse on meta discourse. *College Composition and Communication, 36*, 82-93.
- Vaughan, C. (1991). Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 111-125). Norwood, New Jersey:

Ablex Publishing Corporation.

Watson T. R. (1998). Topic-based analysis of classroom discourse. *System*, 26, 303-318.

Weigle, S. C. (2002). *Assessing writing*. New York: Cambridge.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity* (Technical Report No. 17). Honolulu, HI: University of Hawai'i Press.

Appendix

ESL Academic Writing Assessment Checklist

Essay number: _____

1. This essay answers the question.	<input type="checkbox"/> Yes <input type="checkbox"/> No
2. This essay is written clearly enough to be read without having to guess what the writer is trying to say.	<input type="checkbox"/> Yes <input type="checkbox"/> No
3. This essay is concisely written and contains few redundant ideas or linguistic expressions.	<input type="checkbox"/> Yes <input type="checkbox"/> No
4. This essay contains a clear thesis statement.	<input type="checkbox"/> Yes <input type="checkbox"/> No
5. The main arguments of this essay are strong.	<input type="checkbox"/> Yes <input type="checkbox"/> No
6. There are enough supporting ideas and examples in this essay.	<input type="checkbox"/> Yes <input type="checkbox"/> No
7. The supporting ideas and examples in this essay are appropriate and logical.	<input type="checkbox"/> Yes <input type="checkbox"/> No
8. The supporting ideas and examples in this essay are specific and detailed.	<input type="checkbox"/> Yes <input type="checkbox"/> No
9. The ideas are organized into paragraphs and include an introduction, a body, and a conclusion.	<input type="checkbox"/> Yes <input type="checkbox"/> No
10. Each body paragraph has a clear topic sentence tied to supporting sentences.	<input type="checkbox"/> Yes <input type="checkbox"/> No
11. Each paragraph presents one distinct and unified idea.	<input type="checkbox"/> Yes <input type="checkbox"/> No
12. Each paragraph is connected to the rest of the essay.	<input type="checkbox"/> Yes <input type="checkbox"/> No
13. Ideas are developed or expanded well throughout each paragraph.	<input type="checkbox"/> Yes <input type="checkbox"/> No
14. Transition devices are used effectively.	<input type="checkbox"/> Yes <input type="checkbox"/> No
15. This essay demonstrates syntactic variety, including simple, compound, and complex sentence structures.	<input type="checkbox"/> Yes <input type="checkbox"/> No
16. This essay demonstrates an understanding of English word order.	<input type="checkbox"/> Yes <input type="checkbox"/> No
17. This essay contains few sentence fragments.	<input type="checkbox"/> Yes <input type="checkbox"/> No
18. This essay contains few run-on sentences or comma splices.	<input type="checkbox"/> Yes <input type="checkbox"/> No
19. Grammatical or linguistic errors in this essay do not impede comprehension.	<input type="checkbox"/> Yes <input type="checkbox"/> No
20. Verb tenses are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
21. There is consistent subject-verb agreement.	<input type="checkbox"/> Yes <input type="checkbox"/> No
22. Singular and plural nouns are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
23. Prepositions are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
24. Articles are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
25. Pronouns agree with referents.	<input type="checkbox"/> Yes <input type="checkbox"/> No
26. Sophisticated or advanced vocabulary is used.	<input type="checkbox"/> Yes <input type="checkbox"/> No
27. A wide range of vocabulary is used.	<input type="checkbox"/> Yes <input type="checkbox"/> No
28. Vocabulary choices are appropriate for conveying the intended meaning.	<input type="checkbox"/> Yes <input type="checkbox"/> No
29. This essay demonstrates facility with appropriate collocations.	<input type="checkbox"/> Yes <input type="checkbox"/> No
30. Word forms (noun, verb, adjective, adverb, etc) are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
31. Words are spelled correctly.	<input type="checkbox"/> Yes <input type="checkbox"/> No
32. Punctuation marks are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
33. Capital letters are used appropriately.	<input type="checkbox"/> Yes <input type="checkbox"/> No
34. This essay contains appropriate indentation.	<input type="checkbox"/> Yes <input type="checkbox"/> No
35. Appropriate tone and register are used throughout the essay.	<input type="checkbox"/> Yes <input type="checkbox"/> No