

# Analysis of Linguistic and Discourse Feature of the Essays Written for the York English Language Test

Ms. Nadia Tarique Haque

Lecturer, Department of English & Modern Languages,  
North South University,

Dhaka, Bangladesh

nadiansu051@yahoo.com; nadia.siraj@northsouth.edu

**Abstract** – This study builds on the Explanation inference of the Interpretative Argument (Chapelle et al., 2009) and on the work of Cumming et. al (2006) to address the question of whether and how the linguistic and discourse features of essays written in response to the reading-to-write tasks of the York English Language Test (YELT) vary across proficiency levels. For this investigation, a sample of 50 YELT essays, with YELT score between 2 and 7, were analyzed for the features of essay length, lexical sophistication, syntactic complexity, grammatical accuracy, density of interactive and interactional metadiscourse markers, and quality of argument structure. The results showed marked differences across proficiency levels in terms of lexical, syntactic, pragmatic and rhetorical features. In sum, it can be concluded that with the increase in proficiency levels, as depicted by YELT scores, the essays were longer and improved in terms of syntactic complexity, density of interactive and interactional metadiscourse markers, and quality of argument structure.

**Keywords:** *L2 writing, language proficiency, proficiency levels, L2 texts, discourse analysis*

## 1. INTRODUCTION

Although research has provided us with important insights on L2 writers, their writing behaviors (e.g., Raimes, 1985; Zamel, 1985) and the structure of their texts (Connor & Kaplan, 1987), an important question regarding *how language proficiency of a L2 writer may influence or manifest itself through the different linguistic and discourse features in their texts* still deserves to be addressed. A major reason for this being its direct relevance to the challenging situation that many ESL students encounter while pursuing higher education in English-medium universities, where they not only have to deal with a developing language<sup>1</sup> but also have to meet the demands of competing with NES.

For anyone in this context a certain level of linguistic competence is necessary, and amongst all the language skills' *writing is considered as one of the best indicators and predictors of L2 proficiency* and therefore, their academic success. For these students just the way good listening and reading skills are integral to the reception of academic knowledge, good speaking and writing skills are essential for the incorporation and production of that knowledge; the testimony to the importance given to writing is the fact that most of the gate-keeping practices at English-medium universities require the production of at least one written sample by prospective applicants.

Although writing abilities are considered crucial for success in academic contexts, a number of studies on language pedagogy have concluded that writing skills<sup>2</sup> are THE most complex to acquire and master in, both, a student's L1 and L2 (e.g., Brown & Yule, 1983; Krashen, 1984; Nunan, 1989).

Studies on contrastive analysis have established that problems experienced by L2 writers are more serious than those encountered by their L1 counterparts (Grabe, 2001) because their writing is significantly influenced by their culture, L1 instruction, (Hinkel, 2003; Yang, 2001) and their perception about L2. According to Hamp-Lyons (2001) and North & Pillay (2002), the fact that assessment practices often emphasize writing, and that the writing test scores are an indicator of academic progress in university means that failure to comply with the aforementioned may signify unpreparedness for university-level studies. This points to an urgent need to understand L2 writing proficiency in academic contexts, and with reference to the range and types of research done so far, one of the ways this can be done is by *developing a better understanding of the linguistic and discourse features that are attributed to good academic writing* (Witte & Faigley, 1981).

The present study attempts to pursue this goal by coding a sample of 50 ESL essays<sup>3</sup> - which were written in response to a reading-to-write task of the York English Language Test (YELT) – to understand how the use of different linguistic and discourse features<sup>4</sup> vary with the quality<sup>5</sup> of the essays and the language proficiency of the test-takers. The current study accomplishes this goal by addressing the following research question: ***To what extent and how do essays at different proficiency levels on the YELT rating scale differ in terms of their linguistic and discourse features?***

It is worth mentioning that this study is a replication of a study done by Cumming, Kantor, Baba, Eouanzoui, Erdosy, and James (2006) that compared the differences in written discourse for TOEFL prototype integrated<sup>6</sup> tasks with independent tasks, across the different proficiency levels. The rationale for the current study stems from previous research and arguments which suggested that integrated writing tasks –in this context, reading-to-write tasks – have advantages over independent writing tasks because the former assess writing abilities better, “improve the washback effects of the test on teaching and learning practices internationally” (Cumming et. al., 2006, p.1), and are more authentic and resemble the writing tasks that test-takers will encounter in English medium universities (e.g., Cumming, Kantor, & Powers, 2001; Cumming, Kantor, Powers, Santos, & Taylor, 2000; Hamp-Lyons & Kroll, 1997; Lee, Kantor, & Mollaun, 2002; Rosenfeld, Leung, & Oltman, 2001).

<sup>2</sup> despite being the most important

<sup>3</sup> between the YELT score levels of 2 to 7

<sup>4</sup> essay length, lexical sophistication, syntactic complexity, grammatical accuracy, density of metadiscourse features, and quality of argument structure

<sup>5</sup> Essay quality is operationalized as the YELT essays scores that the essays received based on the standardized YELT scoring rubric

<sup>6</sup> tasks involving responding to print or audio sources

<sup>1</sup> A developing L2

## II. CONTEXT OF THE STUDY

For admissions at York University, ESL students are required to take either TOEFL or IELTS or YELT as a *measure of academic English language proficiency*. The *York English Language Test (YELT)*, a standardized test, is administered by the York University and evaluates the academic English language proficiency in reading, writing, listening, and speaking in two phases. The first is a 90-minute writing component that evaluates reading and writing skills, and the second phase involves a 15–to-20 minute oral component that assesses listening and speaking skills; only those who pass the first phase are allowed to take the second phase of the test. Once a student has taken both components of YELT, s/he receives an overall band score between 1 and 7 based on her/his overall performance; according to the YELT rating scale, 1 is the highest overall band score while 7 is the lowest.

The written component of the YELT comprises of a reading and a writing test and requires the examinees to answer, in their own words, five reading-comprehension questions based on two reading passages within the first 60 minutes of the test. In the next 30 minutes, they have to write a well-organized essay that requires them to provide their thoughts and opinions on a topic related to the content of the reading passages. The YELT writing tasks are integrated-timed writing tasks in which the test-taker has to write an essay in response to a question/ topic that draws, both, from the test-taker's background knowledge and the reading texts. Two independent raters scores these YELT essays using the standardized YELT scoring rubric, which emphasizes two broad criteria namely: structure and organisation, and grammar and conventions, and they encompass features like, essay length, presence of clear explanations and elaborate examples, display of an understanding of the genre and the text, reflection of personal opinions, a clear thesis, variation in syntactic structure, and coherence in the essays.

## III. THEORETICAL BASIS

Both test takers and researchers hold test designers responsible for providing valid and clear interpretations of test scores along with authorizing the proper use of these scores. For this study, the *YELT scores are interpreted as the ability of the test takers to use and understand English language in English medium colleges' and universities*. Since *YELT scores are responsible for distinguishing between test-takers from different proficiency levels*, and consequently *facilitating admissions decisions at York University*, it is obligatory that the test scores be used accurately to make valid interpretations.

### A. Second Language Proficiency

According to Messick (1994), a language proficiency theory should be the basis of *score interpretation* for all the language tests. However, it is also true that pinning down a suitable theory of language proficiency, “is a divisive issue in language assessment, and agreement does not exist on a single

best way to define language proficiency to serve as a defensible basis for score interpretation” (Chapelle et al., 2008, p. 2). Instead of inferring *language proficiency* as encompassing the knowledge of grammar, vocabulary and other discrete elements only, it is now being accepted as the ability to use a complex combination of linguistic and strategic knowledge, and processes to achieve communicative goals (Chapelle, Enright & Jamieson, 2008). According to Cummins (1983), language proficiency must also consider the context in which it is being used because the latter can significantly influence the linguistic performance of a speaker. Therefore the knowledge of linguistic forms – although important – is considered inadequate in language assessment because performance in a test is inferred as involving “more than direct reflection of knowledge” (Chapelle et al., 2008, p. 2) of the language and its rules.

The discussions of the complexity of defining language proficiency has required test designers to consider a more comprehensive approach to account for test score interpretations and uses; this has led to a *unified view of validity*, which requires different types of evidences to justify the interpretation and uses of test scores (Chapelle et al., 2008). However, only very recently has the concept of a unified view of validity with multiple types of evidence been considered for synthesis “into an integrated evaluative judgment<sup>7</sup>”, or into a “validity argument<sup>8</sup>” (Chapelle et al., 2008, p. 5). This conceptual approach is being developed by researchers such as Mislevy, Steinberg, and Almond (2002, 2003) and Kane (Kane, 2001; Kane, Crooks & Cohen, 1999) and it entails developing

- an interpretative argument that justifies test score interpretation and use, and
- a validity argument developed through “critical analysis of the plausibility of the theoretical rationales and empirical data that support the inferences of the interpretative argument” (Chapelle et al., 2008, p. 5).

### B. The Interpretative Argument

A “validity argument supporting test score interpretation and use should be based on an overall interpretative argument” (Chapelle, 2008, p.6), which provides the basis for test validation as a scientific inquiry into score interpretation and use. It is based on six inferences, which are summarized in Table. 1 below. Since this paper focuses on the YELT writing tasks, the constituent inferences of the interpretative argument are explained using a hypothetical example of a test taker who received a band score of 6 on the YELT writing task.

#### Inferences

1. *Domain description* – connects “performances in the target domain” (Chapelle et al., 2008, p. 14) with the observed performance in the test domain. This implies that the observation of the performance in the test reveals knowledge of the skills and abilities in situations representative of the actual target settings. Since the primary use of the YELT is for admission to an English-medium

<sup>7</sup> According to Messick

<sup>8</sup> According to Cronbach

university, the test score users would demand evidence that supports the inference of domain description. For example, an important performance of a student in the target domain may be responding to or incorporating information from print sources, and in that case, the observation of an examinee's performance on YELT task can help predict how well the student will perform across similar tasks in the target setting. A student, who has a score of 6, provided the task is authentic, will have difficulty performing satisfactorily across similar tasks in the target setting.

2. *Evaluation* – assumes that “the observations of performance on test tasks are evaluated” in order to give observed scores that clearly reflect the target language abilities of the test taker (Chapelle et al., 2008, p. 15). According to the current context, evaluation occurs when an examinee's observed performance receives a score – which the raters agree – is reflective of the test taker's language abilities. In regards to the hypothetical example, this happens when the test-takers performance on the test containing major grammatical errors, incoherent response and unclear propositions receives a score of 6.

3. *Generalization* – is the link between the scores assigned to the student's performance; for instance, on the writing component of the actual test and his/ her expected score on similar tasks under similar conditions. Therefore, if a student receives a score of 6 in an integrated writing task, which is a valid representative of the tasks the student is expected to encounter in the test, according to the generalization inference, there is a high possibility that the student will receive similar scores across similar writing tasks in the test. The generalization inference is of critical importance in standardized assessments because it is necessary to be able to claim that an examinee's test scores are comparable “no matter which test form they take, where they take the test, or who scores the responses” (p. 16).

4. *Explanation* – links the expected test score to the construct of academic language proficiency<sup>9</sup> (Chapelle et al., 2008); here academic language proficiency is operationalized as the knowledge of language, usage and strategies that the writers employ in their communication and in their language usage. With relevance to the present study, the evidence pertaining to this inference is obtained by analyzing the linguistic and discourse characteristics of language tasks and the strategies used in performing those tasks. Hypothetically, the student's performance in the test – in which s/he has received a score of 6 – will be interpreted by drawing on a construct of writing ability in academic contexts which in turn encompasses the linguistic, pragmatic and strategic competence that is entailed when the students write in the target settings.

5. *Extrapolation* – links the construct of language proficiency with the target score. This means performance on the test – as measured by test scores – is related to the criteria of academic language proficiency which is in turn defined as the linguistic knowledge, processes and strategies that test takers use to respond to the tasks. For example, it can be said that the score a test taker receives on YELT allows score users to make claims regarding the test-taker's actual “quality of performance in the real-world domain of interest” (Chapelle, et al., 2008, p. 16) or the target language setting. So, if a student has some ability in writing essays but lacks concrete knowledge of sentence construction and of English language rules, and/ or lacks the ability to integrate information from prompts to support claims in an academic task, s/he may encounter substantial difficulty in integrating information from the course readings and academic lectures in his writing in the target setting.

6. *Utilization* – relates the target score to test use and includes decisions regarding admissions and course enrolment. Unfavorable conclusions based on performance on the writing tasks in YELT (in this case a score of 6) may imply that the test taker will have difficulty in courses that require students to integrate information from one or more sources. This may in turn contribute to a decision such as, the test-taker should be out in an intensive ESL program before pursuing university studies. The actual decision will be dependent on the “cut score set by an institution” which in turn means that “test scores will reveal distinctions in the test takers that

will allow institutions to set cut scores and make decisions on the basis of the cut scores that they set” (Chapelle et al., 2008, p. 17).

### C. The Explanation Inference – Linguistic Knowledge, Processes and Strategies

The current study builds on the understanding that *academic language proficiency of different test-takers can be assessed in terms of the linguistic knowledge and strategies they use in their communication*; hence, an understanding of the connection between language proficiency, linguistic knowledge and strategies that the test takers use in their writing can work well as the theoretical basis for this study. The proposition that the academic language proficiency of a test-taker is manifested in the linguistic knowledge, processes and strategies one uses to respond to tasks is supported by studies involving the analysis of the discourse characteristics of test takers' written or spoken responses, task characteristics and item difficulty, concurrent correlational studies, and the comparison of groups with different levels of proficiency (Chapelle, et al., 2008). Additionally, the explanation inference of the interpretative argument is supported by and forms the basis of many studies that investigated discourse characteristics of L2 learners' texts (e.g., Cumming et al., 2006; Banerjee, Franceschina & Smith, 2007; McNamara, Crossley & McCarthy, 2010).

## IV. REVIEW OF RELATED LITERATURE

With the increasing heterogeneity of students in North American universities, the latter is facing a lot of the challenges regarding designing courses and curriculum to meet the diverse demands. These dilemmas stem directly from the assumption<sup>10</sup> that the *writing process and features of L2 writers differ significantly from their L1 counterparts*. L2 composition is considered a less fluent process that often results in significantly shorter texts (Cummings, 1990; Mahmoud, 1982) that have more errors (Benson, Deming, Denzer & Valeri-Gold, 1992; Silva, 1990), and are less effective in communicating meaning (Reid, 1988; Xu, 1990; Yu & Atkinson, 1998). Numerous studies on the structure of L2 texts (e.g., Campbell, 1987a, 1987b; Kamel, 1989; Oi, 1984; Silva, 1993; Xu, 1990) have also concluded that “at the discourse level” L2 texts “often exhibited distinct patterns of exposition, argumentation, and narration” and that “their responses to two particular types of academic writing tasks – answering essay exam questions and using background reading texts – [tend to be] different and less effective”, and their orientation of readers tend to be less appropriate (Silva, 1993, p. 668).

One way we can map the distinctness of L2 writing practices and features is through *text/ discourse analysis studies* which, according to the interdisciplinary fields of applied linguistics and education, can be carried out through the analysis of a corpus of texts. Although a new way of analysis, this approach provides insights into language structure and

<sup>9</sup> academic language proficiency is operationalized as the knowledge of language, usage and strategies that the writers employ in their communication and language usage

<sup>10</sup> This assumption is also validated by many text analysis studies done till date.

use, occurrences of specific textual features and structures and their variation across texts and writers (Kaplan & Grabe, 2002). A number of studies on the linguistic and discourse features of writing tasks for assessing ESL learners proficiency have been conducted for understanding how different task types differ in the nature of the discourse and linguistic features produced (e.g., Cumming et al., 2006; Kamel, 1989), how different discourse and linguistic features vary across different proficiency levels (e.g., Cumming et al., 2006; Banerjee et al., 2007; McNamara et al., 2010), how different essay topics put varying demands on the types of linguistic and discourse features used by L2 writers (e.g., Frase, Faletti, Ginther & Grant, 1999), and how writers from different language groups use the different linguistic and discourse features differently (e.g., Connor, 1984; Oi, 1984; Kamel, 1989; Frase et al., 1999; Banerjee et al., 2007; Zare-ee & Farvardin, 2009).

Cumming et al. (2006) assessed whether and how written discourse for TOEFL prototype integrated tasks differed from that of the independent essays. 216 essays, representing score levels 3-5, was coded for lexical and syntactic complexity, grammatical accuracy, argument structure, orientations to evidence, and verbatim uses of source texts; the analysis showed that the discourse produced for integrated writing tasks differed noticeably at the lexical, syntactic, pragmatic and rhetorical levels from that produced for the independent writing tasks. Differences in the frequency of some of the other features were also noticed across proficiency levels; for example, more proficient examinees wrote longer compositions, used a variety of different words, wrote more and longer clauses, demonstrated greater grammatical accuracy, had better claims in their arguments, and made more summaries of the source evidence. For integrated writing tasks, the examinees indicated information sources other than self, paraphrased/ repeated verbatim or summarized from source texts, used longer and a wider range of words, wrote longer and more clauses yet, wrote shorter and less argumentatively oriented texts. Since the frequency of the aforementioned discourse features differed noticeably with respect to the proficiency levels of the test takers, it could be inferred that “these discourse features are integral to distinguishing these score levels attributed to examinees’ writing” (p.44).

Another study explored how the competence levels, operationalized by the IELTS rating scale, might be related to the L2 developmental stages (Banerjee, et al., 2007). Texts of 275 IELTS test takers from two language groups and across six different proficiency levels<sup>11</sup> were analyzed for cohesive devices, vocabulary richness, grammatical accuracy, and syntactic complexity, and the findings suggested that all the features, except for syntactic complexity, were informative of increasing proficiency levels. Furthermore, the gains in vocabulary were salient at lower IELTS band levels but other criteria became increasingly salient at higher band levels. The results of the analysis of vocabulary richness have supported

researchers’ expectations that essays at higher IELTS band score levels “displayed greater lexical variation and sophistication” (Banerjee et al., 2007, p. 301), and that L1 of the test taker affected the lexical output, lexical variation and lexical density but not the lexical sophistication of the essays. Different task types also appeared to affect the use and quality of the linguistic and discourse features used; essays written in response to task 1 of the IELTS writing module had greater lexical density and generated the use of fewer high-frequency words as a proportion of total words unlike those written in response to task 2, which were lexically more varied.

Frase et al., (1999) compared the linguistic properties of a large sample of TOEFL essays written by test takers from five language groups, on a variety of topics, and for 106 variables. The study also tried to determine how the TWE<sup>12</sup> performance of the writers of different language groups related to the essay styles. Discriminant function analysis of the data suggested, three features that can distinguish the language performance of the different language groups are directness, expressiveness, and academic stance; the ‘number of words’ in an essay and ‘average word length’ were shown to be more predictive of TWE scores than linguistic complexity. The outcomes further suggested that differences between topics could affect performance of language groups in similar way<sup>13</sup>.

McNamara et al., (2010) computationally<sup>14</sup> analyzed a corpus of expert-graded essays at different score levels. The goal of the study was to examine the linguistic differences related to “cohesion and linguistic sophistication” (McNamara et al., 2010, p. 65) and the essays were tagged for: cohesion<sup>15</sup>, syntactic complexity<sup>16</sup>, the diversity of words used by the writer, and characteristics of words<sup>17</sup>. According to the authors, the three most predictive indices of essay quality in this study were syntactic complexity<sup>18</sup>, lexical diversity<sup>19</sup> and word frequency; based on the findings, more-skilled writers used more sophisticated language.

All the aforementioned studies point to substantive discrepancies in the writing features and writing qualities emerging across different proficiency levels; a greater writing proficiency – determined by the test takers’ scores on the writing tasks – was often associated with attributes like *longer responses, greater lexical sophistication, higher syntactic complexity, substantial grammatical accuracy, coherence and cohesion of the texts, and effective argument structures*.

For selecting the linguistic and discourse features for the current study, relevant published literature on

<sup>12</sup> Test of Written English

<sup>13</sup> For instance, in terms of the diversity of grammatical structures used and so on

<sup>14</sup> Coh-Metrix, an automated text analysis tool, was used to examine the degree to which different linguistic features of the essays can be predicted and better understood.

<sup>15</sup> co reference and connectives

<sup>16</sup> number of words before the main verb, sentence structure overlap

<sup>17</sup> frequency and concreteness

<sup>18</sup> number of words before the main verb

<sup>19</sup> Measure of Textual Lexical Diversity

<sup>11</sup> as signified by the IELTS bands 3 to 8,

written discourse analysis, text analysis, and L2 writing assessment were extensively reviewed. Table 2 below gives a brief description of the selected linguistic and discourse features for which the essays will be tagged by defining them with respect to the analytical framework each of them have been derived from, along with the procedures for analysis, reasons behind the choice of the particular features and the findings from relevant studies regarding them. Besides the fact that these features have been in most text analysis studies, these (selected) features also allowed the analysis of both intra-sentential and extra-sentential aspects of the written texts (Cumming et al., 2000; Cumming et al., 2001; Cumming et al., 2006).

1. *Essay Length*: the total number of words written within the allocated time (Polio, 1997), is believed to be a strong predictor of fluency and of the scores assigned to essays (Grant & Ginther, 2000; Polio, 1997; Ried, 1986). Larsen-Freeman (1978) reported, “subjects with a higher proficiency tended to write longer compositions – perhaps because of their fluency” (p. 44).

2. *“Lexical Sophistication”* (Cumming et al., 2006, p.4): is indicative of sophisticated writing (Grant & Ginther, 2000), and includes, average word length and type/token ratio. Average word length is the number of characters divided by the number of words per essay, and type/ token ratio is the summation of different lexical items or segments divided by the summation of lexical items or segments (Engber, 1995).

Several studies (e.g., Connor, 1990; Cumming et al., 2006; Engber, 1995; Ferris, 1994; Frase et al., 1999; Reid, 1986) have consistently shown that: a well-written composition and essays with higher scores comprised of correct, long and variety of words. Grant and Ginther (2000), Engber (1995), and Reid (1986) confirmed that with the increase in the proficiency level of the test taker, there is a steady increase in the lexical sophistication and “flexibility in vocabulary usage” (Frase et al., 1999, p. 2). L2 writers use shorter (Reid, 1988) and vague (Carlson, 1988) words and their overall writing exhibited less lexical variety and sophistication (Hu, Brown & Brown, 1982; Linnarud, 1986).

According to Raimes (1985), lexical proficiency is a major problem for L2 writers with lower language proficiency and can impede their composing processes; moreover Engber (1995) suggested, “an efficient retrieval of vocabulary is especially important in the timed writing tasks” (p.140) as most students quite frequently encounter them in academic contexts. Other than correlating lexical proficiency with academic achievement, understanding the lexical proficiency of L2 learners is important because L2 errors can often result in global errors that can severely impede communication (de la Fuente, 2002; Ellis, 1995; Ellis, Tanaka, & Yamakazi, 1994).

3. *“Syntactic Complexity”* (Cumming et al. 2006, p.4): can be measured by computing the number of clauses per T-unit, and the number of words per T-unit<sup>20</sup> (Polio, 1997).

A large number of studies reported that L2 writers often produced sentences with more (Gates, 1978; Silva, 1990) but shorter (Cumming, 1990; Dennett, 1985, 1990; Gates, 1978; Kamel, 1989) T-units, and fewer (Gates, 1978; Hu et al., 1982) but longer (Gates, 1978; Hu et al., 1982; Silva, 1990) clauses. Usually, L2 writing appeared to be less complex (Park, 1988), less mature, less stylistically appropriate (Yu,

1988), less consistent, and less academic as far as language, style, and tone are concerned (Campbell, 1987a; 1987b; 1990).

T-unit length or number of words per T-unit, is one of the most often used measures of L2 development and has turned out to have a significant positive effect on proficiency for most of the studies (Larsen-Freeman, 1983; Hirano, 1991; Homburg, 1984; Ishikawa, 1995; Kawata, 1992) because longer production units are assumed to be an indicator of the presence of more complex structures in writing (Wolfe-Quintero et al., 1998).

4. *“Holistic rating of grammatical accuracy”* (Cumming et al., 2006): can be evaluated impressionistically on the scale of 0 to 3 according to Hamp-Lyons and Henning (1991). The rating is holistic, takes the whole composition into account and considers the frequency, range and gravity of errors involving grammar, punctuation, spelling, and word choice” in the T-units.

Studies (e.g., Corrigan & Upshur, 1982; McNamara, 1990; Cumming & Mellow, 1996) have suggested that grammatical ability correlates highly with the overall proficiency of the test-takers. Analyzing essays for grammatical accuracy can have important implications because L2 writers generally make more errors in general (Silva, 1990; Yu & Atkinson, 1988), more morphosyntactic errors (Campbell, 1987b; Hu et al., 1982; Silva, 1990), more lexicosyntactic errors (Dennett, 1985; Linnarud, 1986; Yu & Atkinson, 1988), and more errors with verbs (Hu et al., 1982; Silva, 1990), prepositions, articles and nouns (Silva, 1990). Error rate is not only a good predictor of the proficiency level of the student (Hawkey and Barker, 2004; Wolfe-Quintero et al., 1998), but more generally has been of immense help in acting as a yardstick of development in first and second language acquisition (e.g., Brown, 1973; Zobl & Liceras, 1994; Goldschneider & DenKeyser, 2001).

5. *Metadiscourse*: and its view, according to Hyland and Tse (2004) and Hyland (2005), divides metadiscourse markers into interactive and interactional metadiscourse markers.

According to Hyland (2000), with the judicious addition of metadiscourse a writer is able to transform a dry and a difficult text into coherent, reader-friendly prose and relate it to a given context and convey his or her personality, credibility, audience-sensitivity, and relationship to the message. Metadiscourse markers also allow: writers to directly interact with their audience and develop a dialogue, and “explicitly mark the structure of the text,” (Intaraprawat & Steffensen, 1995, p.254), and increase cohesion by making relationships between the different textual units explicit. Literature suggests that better ESL and L1 texts tend to have higher levels and a wider variety of metadiscourse forms than poorer texts (Intaraprawat & Steffensen, 1995; Chaudron, 1987; Cheng and Steffensen, 1996). According to Crismore and Farnsworth (1990) and Hyland (1998a), metadiscourse markers are particularly regarded as an essential component of persuasive and argumentative written discourse.

6. *“Quality of argument structure”* (Cumming et al., 2006, p.5): One of the most commonly used operational definitions of the quality of argument structure has been adapted from the work of Knudson (1992), McCann (1989), Toulmin (1958), and Toulmin, Rieke, & Janik, (1984); it evaluates argument structure based on a 3-point rating scale, of, claims, data, warrants, oppositions, and responses to oppositions (e.g., Connor, 1990, 1991; Crammond, 1998; Yeh, 1998).

Argumentative writing empowers students and enables them to produce texts in which they can express their opinions and justify them, and also promote critical and analytical thinking, production of new knowledge, (Perelman, 1982) and creation of new meanings (Enos & Lauer, 1992).

An ongoing concern has been the minimal or the below-standard competency of students in argumentative writing; National Assessment of Educational Progress (1980) concluded that students comparatively have more difficulty with writing tasks involving arguments. Several studies on understanding the properties of L2 arguments (e.g., Mahmoud, 1982; Connor, 1984; Oi, 1984; Ouiaouicha, 1986; Choi, 1988; and Yu and Atkinson, 1988) concluded that L2 writers exemplify less, less often state and support their position fully,

<sup>20</sup> The shift from the sentence to the T-unit as the basic unit of syntax and hence, the main indicator of syntactic complexity and sophistication, occurred as the need to probe the internal attributes or qualities of sentences was considered more important for differentiating the performance of learners. In fact, this is one of the reasons why Hunt (1965) abandoned counting the number of sentences in children’s writing as there often is a preponderance of run-on sentences and sentences with punctuation errors.

often tend to develop arguments by restating their positions (Mahmoud, 1982), have less justifying support for their claims, have less linking of concluding inductive statements to the prior subtopics of problems (Connor, 1984), and have arguments that are more mixed, tentative and alternating (Oi, 1984). Given the importance and prevalence of arguments and argumentative essays in academic writing, analyzing ESL essays in terms of their argument structures is crucial for understanding how ESL writers understand and express an argument and why they are weaker in developing effective arguments.

### III RESEARCH DESIGN

The sample essays were selected and given by the YELT administration, and were analyzed to understand whether and how they differed in terms of the selected linguistic and discourse features namely: essay length, lexical sophistication, syntactic features, grammatical accuracy, density of interactive and interactional metadiscourse features, and quality of argument structure. The essays had no background information on the test takers, and the only information evident was that the test takers were all L2 speakers of English and were taking the test as York University's admission requirements. All the essays were also written in response to the same task and topic. The reading-to-write YELT task, for which the sample essays were composed, consisted of two reading passages of approximately the same length on the topic of "Folk Medicine". The instructions for completing the task were explicitly provided in writing on the test paper. The question, in response to which the sample essays were written, is as follows: *Write a well-organized essay (with a clear beginning, middle and an end) on the following topic: Do you agree that traditional remedies and procedures of folk medicines are of value in the treatment of disease these days? Support your opinion by providing reasons for it.* The sample essays were stratified into four different proficiency levels based on the scores they had received (see Table 3 below).

Proficiency Level	YELT Score	Sample Size
1	0 – 2.0	4
2	3.0 – 4.0	22
3	5.0	14
4	6.0 and above	10
Total Number of Essays		50

### IV. ANALYTICAL SCHEME & DATA ANALYSIS

Since the analysis of linguistic and discourse features of the YELT essays was the first of its kind, before the actual analysis, no concrete predictions regarding how the selected textual features will vary across different score levels were made. However, by taking the related literature into consideration, it can be assumed that there will be noticeable differences in the frequency and features of the essays, and that the essays written at higher score levels would demonstrate more proficient and adequate uses of the linguistic and discourse features that essays at lower score levels.

The essays were, at first, analyzed for the essay length and lexical sophistication via a linguistic analysis website called *VocabProfile*. Syntactic

complexity was analyzed manually. Each essay was also assigned a holistic score on a 3-point scale for grammatical accuracy. A quantitative count of the metadiscourse markers was done using a word processing program in order to find out the density (ratio) of the different metadiscourse markers in the essays. The quality of the argument structure of the essays involved identifying and rating the presence and quality of claims, data, warrants, opposition and response to opposition on a 3-point scale.

In order to establish inter-coder reliability in the analysis, six essays – representing 12% of the total sample of 50 compositions – were randomly selected and given to a co-rater for rating. The co-rater was neither provided with the YELT rubric nor informed about the YELT scores assigned to the essays; he rated the essays for 'holistic grammatical accuracy' and 'quality of argument structure'. There was an overall percentage agreement of 61.11%, and individual percentage agreements of 66.67% for grammar, 33.33% for claims, 66.67% for data, 0% for warrants, 100% for oppositions, and 83.33% for response to opposition.

For this study, the six linguistic and discourse features were the dependent variables and the proficiency levels of the test takers was the independent variable. After the quantitative counts of the discourse features were done, descriptive statistics were used to summarize the data and compare variation in the frequency of features across different proficiency levels (see Table.3).

### V. RESULTS & DISCUSSION

There were marked differences amongst essays at different proficiency levels with regards to the frequency, range and quality of the selected linguistic and discourse features except for that of lexical sophistication. More proficient test-takers wrote better and longer essays that contained more clauses, longer T-units, less grammatical and linguistic errors, wider variety of metadiscourse markers, and better argument structure. These findings are blatantly consistent with most of the previous studies, for example, Banerjee et al., 2007; Cumming et al., 2006; Grant & Ginther, 2000; Intaraprawat & Steffensen, 1995; Mahmoud, 1982; Silva, 1990).

Although the small sample size may not be representative of the general ESL population and the nature of the writing task and topic may be specific only to the YELT, findings of this study resonate with those of many discourse and text analysis studies involving a much larger pool of learners, essays and tasks. The present study was also successful in identifying syntactic, pragmatic and rhetorical features that were used more frequently and effectively by more proficient ESL writers. The findings from the analysis are summarized in Table 3 below.

	Level	Min.	Max	Mean	Standard Deviation
Essay Length	1	49	268	155	90.79
	2	91	367	259	89.92
	3	175	363	249	51.55

	4	184	473	307	81.11
<b>TOTAL</b>		49	473	307	85.26
<i>Average Word Length</i>	1	5.67	6.73	6.01	0.48
	2	5.28	6.20	5.83	0.28
	3	5.75	6.39	5.98	0.22
	4	5.51	6.33	5.96	0.31
<b>TOTAL</b>		5.28	6.73	5.9	0.29
<i>Type/Token Ratio</i>	1	0.4	0.84	0.59	0.20
	2	0.34	0.67	0.51	0.08
	3	0.41	0.61	0.53	0.06
	4	0.48	0.67	0.54	0.06
<b>TOTAL</b>		0.34	0.84	0.53	0.87
<i>Number of Clauses/ T-Unit</i>	1	1	2	1	0.50
	2	1	2	1.55	0.50
	3	1	2	1.60	0.50
	4	1	2	1.30	0.50
<b>TOTAL</b>		1	2	1.50	0.51
<i>Number of Words/ T-Unit</i>	1	10	16	13	2.50
	2	10	25	16	3.70
	3	12	22	16	2.60
	4	12	20	16	2.40
<b>TOTAL</b>		10	25	15.7	3.17
<i>Grammatical Accuracy</i>	1	1	1.0	1.0	0.0
	2	1	3.0	1.4	0.6
	3	1	3.0	1.8	0.8
	4	1	3.0	2.1	0.9
<b>TOTAL</b>		1.00	3.0	1.6	0.76
<i>Interactive Metadiscourse Markers</i>	1	0.03	0.1	0.07	0.04
	2	0.04	0.1	0.07	0.01
	3	0.05	0.1	0.07	0.02
	4	0.04	0.1	0.06	0.02
<b>TOTAL</b>		0.03	0.1	0.07	0.02
<i>Interactional Metadiscourse Markers</i>	1	0.02	0.04	0.02	0.01
	2	0.03	0.1	0.05	0.02
	3	0.02	0.1	0.05	0.03
	4	0.04	0.1	0.06	0.02
<b>TOTAL</b>		0.02	0.1	0.05	0.02
<i>Quality of Argument Structure</i>	1	0.0	3.0	1.8	1.30
	2	0.0	7.0	4.2	2.40
	3	1.0	11	6.1	2.60
	4	5.0	10	7.1	1.70
<b>TOTAL</b>		0.0	11	5.1	2.66

For all the 50 essays, the *essay length*<sup>21</sup> varied from 49 to 473 words with the mean length of the entire sample being 257. As Table. 3 shows, the mean number of words per composition was considerably higher for essays at higher proficiency levels in comparison to those at lower levels<sup>22</sup>, perhaps implying, the longer the essays are, the higher the score assigned to them and the better their quality is. Even

<sup>21</sup> Operationalized as the total number of words written within the allocated time (Polio, 1997), which in the current context was 30 minutes

<sup>22</sup> The length of the essays of Level 3 and 4 were noticeably more than those of Levels 1 and 2

though the mean essay length for essays in Level 4 (M= 307) was almost twice that of the essays in Level 1 (M= 155), between Levels 2 and 3 the average length did not vary significantly (M=259 and M=249, respectively). These findings clearly resonate with previous research, which has established that there is a noticeable increase in the essay length with the increase in proficiency levels of the test-takers (Connor, 1990; Frase et al., 1999; Grant & Ginther, 2000).

According to Table. 3, the *average word length* of the essays, irrespective of the proficiency levels, varied from 5.28 to 6.73; the average word length of the whole sample was 5.9 letters per word – which is neither very high nor very low – and the descriptive statistics did not seem to show discernible difference in average word length across different proficiency levels. That the average word length remained fairly unchanged across proficiency levels suggests that for the current sample, ‘average word length’ was not good at distinguishing between proficiency levels.

For the entire sample, the mean *type/ token ratio*<sup>23</sup> varied from 0.34 to 0.84, with the mean of the entire sample being 0.53 which means that the percentage of new words in the sample varied from 34% to 84% with the average being just above 50% of the total number of words written. Also evident from the descriptive statistics, there seemed to be a noticeable difference in the type/ token ratio of essays between Level 1 (M=0.59) and all the other levels, particularly level 2 (M=0.51); Level 1 essays showed a larger variability – indicated by a larger standard deviation – than those of the other levels. One possible explanation for the larger variability and higher mean for Level 1 essays is that the sample size is small (n= 4) compared to the other levels.

While previous research (e.g., Cumming et al., 2006; Frase et al., 1999; Grant & Ginther, 2000) has generally shown that academic and better quality texts are usually longer and contain correct, longer and a variety of words, research on L2 writing specifically reported that L2 texts are comprised of shorter (Reid, 1988) and vaguer (Carlson, 1988) words and displayed noticeably less lexical variety and sophistication (Hu et al., 1982; Linnarud, 1986). On the contrary, the findings from the analysis of lexical sophistication shows that *essays across all proficiency levels had almost similar level of lexical sophistication*. Although the findings were quite unanticipated, it is worth noting that the analysis of lexical sophistication was done using a computerized tagging program, which does not assess how appropriately or inappropriately the words were used (Grant & Ginther, 2000). Other explanations for this may be that the sample size for each proficiency level is not large enough, and probably the distinction between the proficiency levels is not substantially large enough to result in significant differences across proficiency levels. Another explanation for this may be that the feature of lexical

<sup>23</sup> Provides insight into the lexical diversity or the number of different words used by the test takers

sophistication is not sensitive to differences across proficiency levels.

For all the essays in the sample, regardless of the proficiency levels, the *number of clauses per T-unit* varied between 1 and 2 meaning that all the T-units written were either comprised of 1 or 2 clauses. The mean number of clauses per T-unit was the lowest (M=1) for Level 1 essays, noticeably higher for Level 2 (M=1.55), highest for Level 3 (M=1.6), and relatively a little lower for Level 4 (M=1.3) essays. Consequently, the overall trend was towards an increase in the mean number of clauses per T-unit or compound sentences across proficiency levels of the test-takers. Also evident from the raw data is that better quality essays had many more clauses or compound sentences than poorer essays; additionally, many essays in Levels 1 and 2 had several run-on sentences and missing commas or faulty or incomplete sentence structure which, according to the analytical model used in this study, had to be counted at separate clauses.

Similar to the number of clauses per T-unit, *number of words per T-unit*<sup>24</sup> also increased across proficiency levels suggesting that, more proficient test-takers wrote longer or more words per T-unit. The average number of words per T-unit increased from 13 to 16 across Level 1 and 2 but remained fairly constant from Level 2 to 4. However, since the goal was to examine whether there was variation in syntactic complexity across proficiency levels, it is clear that the length of the T-units in the text, as measured in terms of number of clauses and words per T-unit, increased with the quality of the essays. The findings are again in conformity with research done on L2 writing that suggests L2 texts to be less matured syntactically and stylistically, and comprised of sentences with fewer clauses (e.g., Gates, 1978; Hu et al., 1982). The finding that L2 sentences tend to be comprised of fewer clauses is consistent with the findings of the current study as the number of clauses per T-unit did not exceed 2 for all the essays.

For all sample essays, the holistic rating for *grammatical accuracy*<sup>25</sup> ranged from 1 to 3, and the mean score of the entire sample was 1.6 on the 3-point scale. The descriptive statistics for the analysis of grammatical accuracy also showed large differences across proficiency levels, that is, M=1.0 at Level 1, to M=1.4 at Level 2, to M=1.8 at Level 3, to M=2.1 at Level 4. Although the minimum and maximum ratings, for all the proficiency levels, ranged from 1 to 3, a closer examination of the raw scores indicated that essays at higher proficiency levels tended to obtain more ratings of 2 and 3 and less of 1. Levels 1 and 2 essays had higher number of major grammatical mistakes, which often affected the comprehensibility, and clarity of the sentences in the texts. In Level 1 and 2 essays, inconsistent use of tense, faulty sentence structures (i.e., run-on sentences, and sentences that

either did not have the SVO<sup>26</sup> pattern or had either the subject, verb or object missing), and, spelling and punctuation errors were significantly widespread and affected the quality of comprehensibility of the essays. However, due to the lack of any background information about the test-takers, it could not be understood whether these mistakes were due to L1 interference, inefficient English language proficiency, or both.

Level 3 and 4 essays, on the contrary, were noticeably better as far as the types, frequency and gravity of grammatical errors prevalent were concerned. Firstly, with the improvement in the overall quality of the essays, as indicated by the YELT scores, the number and type of major errors and errors that affected the comprehensibility of the texts decreased. These essays, in contrast to those from Levels 1 and 2, had considerably fewer errors, particularly for tense and sentence structures; run-on sentences, fragments, subject-verb agreement, and inconsistent and incorrect tense became less as the essays became better. In fact, syntactic errors were almost nonexistent in Level 4 essays despite being frequent even in many of the essays from Level 3.

As anticipated, *the grammatical accuracy of the sample essays increased with the proficiency level of the test-takers and their essays*, however, the overall grammatical competence of the sample essays were in general quite low as evident from the mean scores and the frequent spelling, punctuation, and word choice errors amongst essays across all proficiency levels. Research on L2 writing clearly resonates with the findings from this study in indicating that L2 writers tend to: be less proficient and make more linguistic errors in their writing in contrast to their L1 counterparts (Silva, 1990; Hu et al., 1982; Yu & Atkinson, 1988).

For the sample essays, the ratio of the *Interactive Metadiscourse Markers* ranged from 0.03 to 0.1 implying that 3% to 10% of the total number of words written in all the 50 essays was interactive metadiscourse markers. The mean ratio of the density of interactive metadiscourse markers remained relatively stable at 0.07 for levels 1, 2 and 3 before marginally declining to 0.06 for Level 4 essays. While the percentage of interactive metadiscourse markers never exceeded 10% of the total number of words, essays written by less proficient writers had relatively fewer interactive metadiscourse markers (Level 1 min=0.03) in contrast to those written by more proficient writers (i.e., Level 2 and 4 min=0.04; Level 3 min=0.05). Although the mean density of interactive metadiscourse markers may not be showing any noticeable variation across proficiency levels, the raw count specified that *better quality essays used a variety of interactive metadiscourse markers judiciously*.

Based on the raw data, transition markers tend to be the most frequently used, while endophoric<sup>27</sup> and evidentials<sup>28</sup> were the least frequently used; in fact,

<sup>24</sup> According to the definition and explanation provided by Polio (1997), was computed by dividing the total number of words in each essay by the total number of T-units in each essay

<sup>25</sup> Was holistically rated on a rating scale of 1 to 3 depending on the number of errors and their effect on the comprehensibility of the writing

<sup>26</sup> Subject-Verb-Object

<sup>27</sup> (In) Section X

<sup>28</sup> According to X

none of the essays in the sample used endophoric markers, and only a very few, surprisingly the ones at the lower proficiency levels, used evidentials as a support for their arguments. The higher frequencies of the transitional markers were possibly because of the fact that they are common words, have transparent functions (e.g., again/ also/ so/ but/ because), and serve a variety of essential functions like signifying: addition, comparison, contrast and consequence (Hyland, 2005). Some other common interactive metadiscourse markers were Code Glosses (e.g., called/ for example/ in fact) and Frame Markers (e.g., finally/ first of all/ in sum/ now/ so) however, their use did not demonstrate any noticeable patterns across proficiency levels except that, *more proficient writers wrote longer texts that employed a range of these features*, especially Frame Markers, to signal text boundaries, and show chronology and transition across different stages.

**Interactional metadiscourse markers** depict the relationship of the reader with the argument, and since metadiscourse markers are one of the facets of good writing, as anticipated, both the range and quantity of interactional metadiscourse markers increased as the quality of the essays got better. The ratio of interactional metadiscourse markers for the entire sample of essays ranged from 0.02 to 0.1, suggesting that interactional metadiscourse markers made up 2% to 10% of the total number of words in the entire sample; whereas, the mean ratio of interactional metadiscourse markers was 0.05, which implies that an average of 5% of the total number of words written in all the essays was interactional metadiscourse markers. According to Table.3, the mean density of interactional metadiscourse markers increased from 2% to 6% from Levels 1 to 4 suggesting that *the total number of interactive metadiscourse markers in the essay sample was significantly higher than that of interactional metadiscourse markers*. Furthermore, *better quality essays in the sample demonstrated more and a wider variety of interactional metadiscourse markers than interactive metadiscourse markers* suggesting that, more proficient test-takers wrote interactive essays that involved the reader/s in the argument or were effectively reader-based. In general, most of the better quality essays exhibited a judicious use of metadiscourse markers that made the texts coherent and less repetitive.

Hedges<sup>29</sup> were the most common interactional metadiscourse markers whereas, attitude markers<sup>30</sup> were relatively rare in entire sample. Essays from Levels 1 and 2 had proportionately more Self-Mention markers than those from Levels 3 and 4 suggesting that less proficient writers used them to force an authorial presence (Intaraprawat & Steffensen, 1995) on the reader to get their argument/ point across. Other than Self-Mention markers, Boosters (e.g., actually/ really), and Engagement markers (e.g., allow/ remember) were noticeably common in the essay sample. Better quality essays, indicated by the YELT score assigned to them,

contained more Hedges and Boosters. While Boosters validate the writer's certainty, the former are a depiction of the writers' willingness to produce a text by accurately acknowledging, "what they believe, what they know, and what they assume" (Intaraprawat & Steffensen, 1995, p. 6). According to Hyland (2005), a judicious balance of Hedges and Boosters is an attribute of a good quality text as it indicates the extent to which the writer is willing to entertain alternatives and further opinions from the readers.

According to Hyland (2005), Intaraprawat and Steffensen (1995), and Cheng and Steffensen (1996), a judicious use of metadiscourse markers is measured as a facet of good quality texts, and the findings of this study are consistent with the predictions made by related literature in the sense that better quality essays in the sample had a higher frequency, a wider range and a judicious selection of different types of metadiscourse markers.

For the criteria of the **quality of argument structure**, each essay was rated in terms of the presence and quality of the following traits, separately: claims, data, warrants, oppositions, and responses to the oppositions. The ratings for the five traits were summed to obtain a total score out of 15 for each essay. For all the essays, the ratings for the quality of argument structure ranged from 0 to 11, whereas the mean score the essays ranged from 5.1 out of 15, indicating that the quality of the argument of the sample essays were below the average. However, *more proficient test-takers*, as indicated by their YELT scores, *had good quality arguments and an effective argument structure*, which also suggests that *less proficient writers are less skillful in generating meaningful and effective arguments* (McCann, 1989).

Table 3 illustrates interesting findings regarding how the different traits of argument structure varied across proficiency levels of the examinees. For the trait of claims, significant differences in both the presence and the quality of claims across proficiency levels were found. Although almost all the essays in proficiency Level 1 had claims, most of them either lacked relevance and/ or clarity; on the other hand, almost all the essays from Levels 2 and 3 explicitly stated a claim in some form or the other and most of these claims were relevant and clear. While only a handful of essays from the lower proficiency levels had complete and relevant claims, a large number of them from Level 4 had clear, relevant and complete claims. For the presence and quality of data, similarly, significant differences were observed with respect to the proficiency levels. At lower proficiency levels (e.g., Levels 1 and 2) there were a significant number of essays that either included no data or included data that did not address the claims or lack clarity. As the essays got better, the relevance and clarity of the data for backing up the claims increased. While majority of the Level 4 essays had relevant and clear data, a number of them also had warrants for justifying or linking the claims with the data. In fact, none of the essays from Level 1 and many from Levels 2 and 3 lacked warrant; amongst the essays that had warrants, most of them lacked clarity, were incomplete or did not fully connect

<sup>29</sup> About

<sup>30</sup> Essential

the data with the claim. Out of the entire sample, only one essay – from Level 4 – had a complete and a clear warrant that properly linked the claim with the data. The absence or warrants in most of the essays, and the failure of almost all the existing ones in relating the claims with the data are all consistent with the findings from studies by McCann (1989) and Crammond (1998).

Extremely rare in almost all the essays were the presence of counterarguments and their corresponding rebuttals. Even many of the better quality essays either failed to provide or provided weak counterarguments and rebuttals; in fact, only 6 out of the 50 essays had counterarguments or recognized opposition and only 2 out of them (one from Level 3 and another from Level 4) included a statement of opposition that was clear, complete and relevant. For the trait of responses to opposition or rebuttals, a similar trend was noticed; that is, only 3 essays from the sample had rebuttals or refutation that was relevant, whereas, none from Levels 1 and 2 had any discernible refutation. In general, *better essays seemed to acknowledge counterarguments and provided logical and pertinent refutation for them.* The findings related to the overall minimal use of counter arguments, response to opposition, and good quality and relevant warrants in the essays relative to claims and data are consistent with previous research (e.g., McCann, 1989; Cooper et al., 1984).

On the whole, the assumption that a proper argument structure requires warrants to link the claim and data (Toulmin, 1958) is somewhat incompatible with the findings in the current study, because, a number of Level 3 and 4 essays, despite having weaker warrants, still managed to have a sound argument structure and managed to get better YELT scores; furthermore, in conformity with Toulmin's (1958) model and findings it is also not always mandatory to explicitly state warrants in writing. For this study, it can be assumed that many test-takers may not have explicitly stated warrants to link their claims and data because they may were inclined to produce texts that are more *writer-based* or texts which make presumptions that minimalist use of warrants should not always be considered as the sole indicator of a comprehensive and effective argument structure. Nonetheless, the same cannot be said for the traits of recognition of opposition and responses to opposition because the very basic aim of any persuasive writing is to consider the views of the audience/ opposition, justify one's own claims, and, in that process, convince the reader that the writer's argument may also be logical and justified. The fact that most of the essays, particularly the ones in the lower proficiency levels, lacked these traits suggested that the test-takers did not possess adequate knowledge about argumentative writing and may continue writing ineffective arguments and writer-based texts unless trained to do otherwise. Besides, acknowledging an opposition and providing a logical and relevant response to it is important, and is a hallmark of effective text, because this not only nullifies the opposition but also validates and reinforces the claims made by the writer.

## VI. CONCLUSION

The findings of this study, with respect of the research question of this study, are all predominantly consistent with related and relevant literature that have established that more proficient writers tend to produce longer and more grammatically and linguistically correct texts with better quality argument structure and that are more effective in conveying the meaning and message to the reader (e.g., Banerjee et al., 2007; Cumming et al., 2006; Grant & Ginther, 2000; Intaraprawat & Steffensen, 1995; Mahmoud, 1982; Silva, 1990). Most importantly, *with reference to the validity argument, we can say that the findings of the study do support the validity argument for the interpretation and use of YELT test scores.* Firstly, the YELT test scores are interpreted as indicators of the ability of the test taker to use and understand English as it is used in target settings; therefore, since the proposed use of the YELT test scores is to distinguish between test-takers of different proficiency levels so as to facilitate admissions and placement decisions at York University, it is mandatory for them to provide valid and clear interpretations. Overall, the linguistic and discourse features for which the sample essays were analyzed indeed show differences in use and attributes across proficiency levels which were determined by variance in the scores that were assigned to the essays; that is, essays with sophisticated, accurate and more appropriate lexical, syntactic, pragmatic, and rhetorical features received higher scores, which consequently implies that more proficient writers wrote them. This validates the fact that different scores were assigned to different essays based on the quality of the texts.

### A. Implications

By documenting and comparing the linguistic and discourse features of the YELT essays, the current study attempted to augment our understanding of some of the major features of L2 writing and their variation across proficiency levels; This can be the starting point of many large-scale discourse and linguistic analysis studies for understanding L2 texts better. Analysis of student essays can also have implications for the development of writing tasks for future tests, which can in turn impact the quality of assessment of writing skills. From this study, not only were variations of some of the defining features of written discourse across the various proficiency levels revealed, but the YELT scoring rubric that has been used to differentiate and assess the language performance of test-takers was also verified; that is, the YELT scoring rubric that is used to evaluate essays and eventually categorize them into different proficiency levels appears to successfully do so. The better our understanding of the different features of writing, their manifestation at different linguistic proficiency levels, and their sensitivity to performance factors such as task effects are, the better we will be able to understand the L2 writing construct (Weigle, 2002; Hawkey & Barker, 2004) and the more effective our assessment criteria and scales will be (Banerjee et. al, 2007).

The reliability of a writing test is dependent on its scoring procedures, which encompass the rubric and the success with which raters use the rubric to rate the tasks. Scoring criteria in writing tests is one of the paramount concerns because it is the scoring process of a test and its tasks that tells us whether a test is reliable or not and also whether valid decisions and inferences about test takers' language proficiency and abilities can be made. The findings of this study *support and validate the ratings of the reading-to-write task of the YELT writing component as a measure of writing ability in English*. Several of the features examined in this study seem to explain variance in the scores assigned to the YELT essays which suggests that the YELT raters attend to these features when rating essays for the YELT test. The results of text analysis also suggest that essays written in response to the reading-to-write YELT task brought out language performance that varied significantly with the language abilities of the test-takers.

Additionally, this study *adds to the existing body of literature that recognizes integrated writing tasks as an important alternative to independent writing tasks* or as "an additional measure of writing ability that can be scored reliably," and can purposefully be used to relate language comprehension to the writing ability (Cumming et al., 2006, p. 46). In contrast to independent writing tasks, integrated writing tasks are more authentic and representative of the task types encountered in academic settings, and put demand on linguistic and cognitive abilities to comprehend and produce texts that amalgamate relevant information from the prompt as well as from personal experience. Moreover, integrated writing tasks provide the basis for background knowledge for ESL test takers (Weigle, 2004) who, being from another culture, may not possess the same cultural understanding of the topic and the task as their NES counterparts. It is assumed that the presence of a source text, especially reading, during the composition process would provide the test-takers with "a common platform" (Gebril, 2009, p. 508) thus ensuring equity and fairness.

Furthermore, the shift to using integrated writing tasks instead of or in addition to independent writing tasks in high-stakes assessment of L2 proficiency also calls for more research in order to understand how test takers address such tasks, and how such tasks influence how test takers of different proficiency levels use linguistic and discourse features in their writing. Although the YELT writing component uses only reading-to-write<sup>31</sup> tasks that, according to this study, seem to successfully differentiate between learners of varying language proficiency levels, further research can be done to examine: how such writing tasks affect L2 test-takers with different L1s, and how test takers interpret the task and their composition processes, their responses to the task, and the scores they receive. Studies can also be conducted to compare the linguistic and discourse features in the written discourse produced by test-takers in responses to integrated writing tasks that differ in terms of their topics. For

scoring the test or test score interpretations, studies can be conducted to examine how essays and task characteristics affect the scoring process. Most importantly, studies on the consequences of YELT score interpretations and use are needed, because, it is through such research that the quality of YELT can be evaluated and assured, and improvements in test design and implementation can be made.

Despite being the first of its kind to examine the YELT writing test, this study has important implications for this test. While designing writing tests and making inferences from their scores the YELT test designers need to be well aware that L2 writing is far from being a simple and predictable process; the complex interplays of all the socio-cultural and affective factors, and different educational backgrounds of ESL students and their effects on their language proficiency and their language performance need to be taken into account when assessing L2 performance. As Hamp-Lyons and Kroll (1997) have pointed out, a test is undoubtedly supposed to assess and discriminate among test-takers; however, it is equally important to make sure that the differences are actually due to the language proficiency and ability of the test takers rather than to any other background factors. Since YELT is a high-stake test, care should be taken to avoid any sort of biases in designing the test and the rubric, and in the rating procedures. However, in order to understand how the aforementioned background factors can impact the writing performance of L2 test-takers across task and topic types and proficiency levels, future research needs to study the impact of test takers' cultural, educational and linguistic background on their performance on test tasks.

### B. Limitations

The study has some limitations that need to be taken into account when interpreting its findings. Firstly this was a small-scale study with a very small sample size; additionally, the sample of test-takers, whose compositions were selected, was limited in number and may not be representative of the full range of test-takers who actually take YELT. Secondly, only a limited range of linguistic and discourse features were examined. Even though the selected features spanned the lexical, syntactic, rhetorical and pragmatic aspects of language use, there are yet many other features for which the sample compositions could have been tagged and assessed. For instance, since the test-takers had a reading prompt to facilitate their writing, the essays could be analyzed for the feature of orientation towards the source text or for the use of cohesive devices. Another limitation was that background information about the test takers, especially their L1, was not available; information about test-takers' L1 could have been helpful in interpreting and explaining the findings of the study, mainly whether the variation in the frequency and range of the features examined is due to variation in proficiency level or due to the interplay of the L1 of the test-takers with their L2 proficiency. In fact, a number of studies (e.g., Hinkel, 2002, Reid, 1990) that

<sup>31</sup> Integrated

examined the influence of test taker's L1 on L2 writing performance found that L1 influence is significant. Besides the impact L1, according to Barkaoui (2007), L2 test takers bring into their respective language performances a number of cognitive, affective, linguistic and socio cultural factors that influence how they choose, read, interpret, and respond to writing tasks and the quality of their L2 texts, writing processes and test scores. All these factors point to the necessity of more research.

#### AUTHOR'S PROFILE

Nadia Tarique Haque has MA in Applied Linguistics from York University, Toronto, Canada. She is currently working as a Lecturer at the Department of English and Modern Languages at North South University, Dhaka, Bangladesh. Her research interests include alternative assessment, language testing, second language writing, motivation and anxiety, socio-cultural theory, and second language instruction.

*The shorter version of this paper was presented at the 5<sup>th</sup> Annual Conference on Language, Literature, and Linguistics (L3 2016) in May 2016 in Singapore.*

#### REFERENCES

- [1] Banerjee, J., Franceschina, F., & Smith, A. M. (2007). Documenting features of written language production typical at different IELTS band score levels (IELTS Research Reports Volume 7). Australia & UK: IELTS Australia and British Council.
- [2] Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12 (2), 86–107.
- [3] Benson, B., Deming, M., Denzer, D., & Valeri-Gold, M. (1992). A combined basic writing/ English as a second language learners. *Dissertation Abstracts International*, 41(5), 1984A.
- [4] Brown, R. (1973). *A First Language: The early stages*. Cambridge, MA: Harvard University Press.
- [5] Brown, D., & Yule, G. (1983). *Teaching the spoken language*. Cambridge, UK: Cambridge University Press.
- [6] Campbell, C. (1987a). Writing with others' words: Native and non-native university students use of information from a background reading text in academic compositions. (ERIC Document Reproduction Service No. ED287315)
- [7] Campbell, C. (1987b). Writing with others' words: The use of information from a background reading text in the writing of native and non-native university composition students. *Dissertation Abstracts International*, 48 (7), 1679A.
- [8] Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 211-230). New York: Cambridge University Press.
- [9] Carlson, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 227-260). Newbury Park, CA: Sage.
- [10] Chapelle, C.A., Enright, M.K., & Jamieson, J.M. (Eds.). (2008). *Building a Validity Argument for the Test of English as a Foreign Language*. USA: Routledge Taylor & Francis Group.
- [11] Chaudron, C. (1987). Analysis of products and instructional approaches in writing: Two articles on the state of art, *TESOL Quarterly*, 21, 673-675.
- [12] Cheng, X., & Steffensen, M. (1996). Metadiscourse: A technique of improving student writing. *Research in the Teaching of English*, 30 (2), 149-181.
- [11] Choi, Y. (1988). Text structure of Korean speakers' argumentative essays in English. *World Englishes*, 7 (2), 129-142.
- [12] Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Research on Language & Social Interaction*, 17 (3), 301-316.
- [13] Connor, U. (1990). Linguistic/ Rhetorical Measures for International Persuasive Student Writing. *Research in Teaching of English*, 24 (1), 67-87.
- [14] Connor, U. (1991). Linguistic/rhetorical measures for international persuasive student writing. *Research in Teaching of English*, 24, 67-87.
- [15] Connor, U., & Kaplan, R.B. (Eds.). (1987). *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley.
- [16] Cooper, C., Cherry, R., Copley, B., Fleischer, S., Pollard, B., & Sartisky, M. (1984). Studying the writing abilities of a university freshman class: Strategies from a case study. In R. Beach & L. Bridwell (Eds.), *New directions in composition research* (pp. 19-52). New York: Guilford.
- [17] Corrigan, A. and Upshur, J. 1982: Test method and linguistic factors in foreign language tests. *International Review of Applied Linguistics*, 20, 313–321.
- [18] Crammond, J. (1998). The uses and complexity of argument structures in student persuasive writing. *Written Communication*, 15, 230-268.
- [19] Crismore, A., & Farnsworth, R. (1990). Metadiscourse in popular and professional science discourse. In W. Nash, (Ed.), *The Writing Scholar. Studies in Academic Discourse* (pp. 119-136). Newbury Park, CA: Sage.
- [20] Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). TOEFL 2000 writing framework: A working paper (TOEFL Monograph No. MS-18, ETSRM-00-05). Princeton, NJ: Educational Testing Service.
- [21] Cumming, A., Kantor, R., & Powers, D. (2001). Scoring TOEFL essays and TOEFL 2000 prototype tasks: An investigation into raters' decision making and development of a preliminary analytic framework (TOEFL Monograph No. MS-22, ETS RM-01-02). Princeton, NJ: Educational Testing Service.
- [22] Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). Analysis of Discourse Features and Verification of Scoring Levels for Independent and Integrated Prototype Written Tasks for the New TOEFL (TOEFL Monograph Series No. 30). Princeton, NJ: Educational Testing Service.
- [23] Cummins, J. P. (1983). Language proficiency and academic achievement. In J. Oller (Ed.), *Issues in language testing research* (pp. 108-126). Rowley, MA: Newbury House.
- [24] Cummings, V. (1990). Speech and writing: An analysis of expository texts compared by native and non-native speakers of English at The City University of New York. *Dissertation Abstracts International*, 51 (7), 2296A.
- [25] Cumming, A. & Mellow, D. (1996). An investigation into the validity of written indicators of second language proficiency. In Cumming, A. and Berwick, R., (Eds.), *Validation in language testing*. (pp.72–93). Clevedon: Multilingual Matters.
- [26] de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24, 81–112.
- [27] Dennette, J. (1985). Writing technical English: A comparison of the process of native English and native Japanese speakers. *Dissertation Abstracts International*, 46(11), 3275A.
- [28] Dennette, J. (1990). ESL technical writing: Process and rhetorical differences (ERIC Document Reproduction Service No. ED 322713).
- [29] Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16, 409–435.
- [30] Ellis, R., Tanaka, Y., & Yamakazi, A. (1994). Classroom interaction, comprehension, and L2 vocabulary acquisition. *Language Learning*, 44, 449–491.
- [31] Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139-155.
- [32] Enos, R. L., & Lauer, J.M. (1992). The meaning of heuristic In Aristotle's Rhetoric and its implications for contemporary rhetorical theory. In S.P. Witte, N. Nakadate, & R.D. Cherry (Eds.), *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy* (pp.79-87). Carbondale: Southern Illinois Press.
- [33] Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28, 414-420.

- [34] Frase, L.T., Faletti, J., Ginther, A., and Grant, L. (1999). Computer Analysis of the TOEFL Test of Written English (TOEFL Research Report 64). Princeton, NJ: Educational Testing Service.
- [35] Goldschneider, J.M., & DeKeyser, R.M. (2001). Explaining The 'Natural Order of L2 Morpheme Acquisition' in English: A meta-analysis of multiple determinants. *Language Learning*, 51, 1-50.
- [36] Grabe, W. (2001). Notes toward a theory of second language writing. In T. Silva & P. K. Matsuda (Eds.), *On second language writing*. Mahwah, NJ: Lawrence Erlbaum.
- [37] Gebril, A. (2009). *Score generalizability in writing assessment: The interface between applied linguistics and psychometric research*. Saarbrücken, Germany: VDM Verlag.
- [38] Gates, R. (1978). An analysis of grammatical structures in compositions written by four groups of students at Southern Technical Institute: Indo-Iranian, Latin Romance, Sino-Tibetan, and American. *Dissertation Abstracts International*, 39(4), 2026A.
- [39] Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9, 123-145.
- [40] Hall, C. (1990). Managing the complexity of revising across languages. *TESOL Quarterly*, 24(1), 43-60.
- [41] Hamp-Lyons, L. (2001). Fourth generation writing assessment. In T. Silva & P. K. Matsuda (Eds.), *On second language writing*. Mahwah, NJ: Lawrence Erlbaum.
- [42] Hamp-Lyons, L., and 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.
- [43] Hamp-Lyons, L., and Kroll, B. (1997). TOEFL 2000. Writing: Composition, community, and assessment (*TOEFL Monograph Series No.5*). Princeton, NJ: Educational Testing Service.
- [44] Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, 9, 122-159.
- [45] Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37, 275-301.
- [46] Hirano, K. (1991). The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students. *Annual review of English language education in Japan*, 2, 21-30.
- [47] Homburg, T.J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly*, 18, 87-107.
- [48] Hu, Z., Brown, D., & Brown, L. (1982). Some linguistic Differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1(1), 39-49.
- [49] Hunt, K. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: National Council of Teachers of English.
- [50] Hyland, K. (1998a). Persuasion and context: The pragmatics Of academic metadiscourse, *Journal of Pragmatics*, 30, 437-455.
- [51] Hyland, K. (2000). Hedges, Boosters and lexical invisibility: noticing modifiers in academic texts. *Language Awareness*, 9(4), 179-197.
- [52] Hyland, K., & Tse, P. (2004). Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics*, 25 (2), 156 – 177.
- [53] Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- [54] Intaraprawat, P. & Steffensen, M.S. (1995). The Use of Metadiscourse in Good and Poor ESL Essays. *Journal of Second Language Writing*, 4(3), 253-272.
- [55] Ishikawa, S. (1995). Objective measurement of low proficiency EFL narrative writing. *Journal of Second Language Writing*, 4, 51-70.
- [56] Kamel, G. (1989). Argumentative writing by Arab learners of English as a foreign and second language: An empirical investigation of contrastive rhetoric. *Dissertation Abstracts International*, 50(3), 677A.
- [57] Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319-342.
- [58] Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 21(1), 31-35.
- [59] Kaplan, R.B., & Grabe, W. (2002). A Modern History of Discourse Analysis. *Journal of Second Language Writing*, 11, 191-223.
- [60] Kawata, K. (1992). Evaluation of free English composition. *CASELE Research Bulletin*, 22, 49-53.
- [61] Knudson, R. (1992). Analysis of argumentative writing at two grade levels. *Journal of Educational Research*, 85, 169-179.
- [62] Krashen, S. D. (1984). *Writing: Research, theory and applications*. Oxford: Pergamon Institute of English.
- [63] Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448.
- [64] Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H.W. Seliger & M. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp. 287-304). Rowley, MA: Newbury House.
- [65] Lee, W., Kantor, R., & Mollaun, P. (2002). *Score reliability as an essential prerequisite for validating new writing and speaking tasks for TOEFL*. Paper presented at the annual TESOL Convention, Salt Lake City, UT.
- [66] Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish*. Lund, Sweden: Liber Forlag Malmö.
- [67] Lux, P. (1991). Discourse styles of Anglo and Latin American college students writers. *Dissertation Abstracts International*, 52, 06A, 2128.
- [68] Mahmoud, A. (1982). A functional analysis of written compositions of Egyptian students of English and the implications of the notional functional syllabus for the teaching of writing. *Dissertation Abstracts International*, 44(5), 1439A.
- [69] McCann, T. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English*, 23, 62-76.
- [70] McNamara, T. 1990: Item response theory and the validation of an ESP test for health professionals. *Language Testing*, 7, 52-75.
- [71] McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2010). Linguistic Features of Writing Quality, *Written Communication*, 27(1), 57-86.
- [72] Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- [73] Moragne e Silva, M. (1991). Cognitive, affective, social, and Cultural aspects of composing in a first and second language: A case study of one adult writer. *Dissertation Abstracts International*, 52(12), 4249A.
- [74] National Assessment of Educational Progress. (1980). *Writing achievement, 1969-1979: Results from the third assessment*, Vols. I, II, III. Denver, Colorado: National Assessment of Educational Progress.
- [75] North, S., & Pillay, H. (2002). Homework: Re-examining the routine. *ELT Journal*, 56, 137-145.
- [76] Nunan, D. (1989). *Designing tasks for the classroom*. Cambridge, UK: Cambridge University Press.
- [77] Oi, K. (1984). Cross-cultural differences in rhetorical patterning: A study of Japanese and English. *Dissertation Abstracts International*, 45 (8), 2511A.
- [78] Ouauicha, D. (1986). Contrastive rhetoric and the structure Of learner-produced argumentative texts in Arabic and English. *Dissertation Abstracts International*, 47 (9), 3339A.
- [79] Park, Y. (1988). Academic and ethnic background as factors affecting writing performance. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 261-272). Newbury Park, CA: Sage.
- [80] Perelman, C. (1982). *The realm of rhetoric* (W. Kluback, Trans.). South Bend, IN: University of Notre Dame Press.
- [81] Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- [82] Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19, 229-258.

- [83] Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, 9, 109–121.
- [84] Reid, J. (1986). Using the Writer's Workbench in composition teaching and testing. In C.W.Stansfield (Ed.), *Technology and Language Testing* (pp. 167-188). Washington, DC: TESOL.
- [85] Reid, J. (1988). Quantitative differences in English prose written by Arabic, Chinese, Spanish, and English student. *Dissertation Abstracts International*, 50 (3), 672A.
- [86] Rimmer, W. (2006). Measuring Grammatical Complexity: The Gordian knot. *Language Testing*, 23 (4) 497–519.
- [87] Rosenfeld, M., Leung, S., & Oltman, P. K. (2001). The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels (TOEFL Monograph No. MS-21, ETS RM-01-03). Princeton, NJ: ETS.
- [88] Silva, T. (1990). A comparative study of the composing of the selected ESL and native English speaking freshman writers. *Dissertation Abstracts International*, 51(10), 3379A.
- [89] Silva, T. (1993). Towards an Understanding of the Distinct Nature of L2 Writing: The ESL Research and Its Implications. *TESOL Quarterly*, 27(4), 657-677
- [90] Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- [91] Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning* (2<sup>nd</sup> ed.). New York: Macmillan.
- [92] Wolfe-Quintero, K., Inagaki, S., Kim, H. (1998). *Second Language Development in writing: Measures of fluency, accuracy & complexity*. University of Hawai: Second Language Teaching & Curriculum Centre.
- [93] Weigle, S. C. (2002). *Assessing Writing*, Cambridge, UK: Cambridge University Press.
- [94] Weigle, S. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9, 27–55.
- [95] Witte, S.P. & Faigley, L. (1981). Coherence, Cohesion and Writing Quality. *College composition and Communications*, 32(2), 189-204
- [96] Xu, G. (1990). An ex post facto study of differences in the Structure of the standard expository paragraphs between written compositions by native and nonnative speakers of English at the college level. *Dissertation Abstracts International*, 51(6), 1942A.
- [97] Yang, Y. (2001). *Chinese interference in English writing: Cultural and linguistic differences*. ERIC Document Reproduction Service no. ED 461 992.
- [98] Yeh, S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in Teaching of English*, 33, 49-83.
- [99] Yu, V., & Atkinson, P. (1988). An investigation of the language difficulties experienced by Hong Kong secondary school students in English Medium schools: Pt. 1. The problems. *Journal of Multilingual and Multicultural Development*, 9(3), 267-284.
- [100] Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19, 79-102.
- [101] Zare-ee, A., & Farvardin, M. T. (2009). Comparison of University Level EFL learners' Linguistic and Rhetorical Patterns as Reflected in their L1 and L2 Writing. *Novitas Royal: Research on Youth and Language*, 3(2), 143-155.
- [102] Zobl, H., & Liceras, J.M. (1994). Review Article: Functional categories and acquisition orders. *Language Learning*, 44, 159-180.