Teaching English Language, Vol. 13, No. 2, Summer & Fall 2019, pp. 269-298

Analysis of Organizational Knowledge in Writing Corpora of the General Training IELTS-Practice Materials

Sara Mansouri
Ph.D. student, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Bahram Hadian[1]
Assistant professor, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Omid Tabatabaei
Associate professor, English Department, Najafabad Branch, Islamic Azad University, Najafabad, Iran
Ehsan Rezvani
Assistant professor, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Abstract
Inspired by the concept of organizational knowledge, this research intended to analyze a corpus of 180 writing performances of two tasks of the General Training IELTS-practice scripts across three band scores, 7, 8 and 9 in grammatical (GK) and textual knowledge (TK) features. It adopted the taxonomies of Bachman and Palmer (2010) and Connor and Mbaye's (2002) to makes analysis through Coh-metrix in 22 GK and TK features. The band comparisons indicated that the highest-scored scripts tended to be longer and include structurally more diverse sentences. Lexically, 8 and 9 band scripts almost showed no differences, however, they were denser and involved more multiple-meaning and abstract words than those of band 7. Textually, band 9 writings proved to be more coherent and cohesive, although they did not differ from band 8 or even band 7 in some other discourse features. The task comparisons demonstrated task two (T2)'s superiority in length, structural diversity, density, noun phrase (NP) concentration and reading and comprehension difficulty. Lexically, task one (T1) superseded in using more frequent words but not in diverse, multiple-meaning and the abstract ones. It was also indicated that T2 needed more coherence, conceptuality, causality but not coreferentiality in cohesion than T1.

***Keywords*:** Band Score, Grammatical Knowledge, IELTS, Organizational Knowledge, Textual Knowledge, Writing

Received on February 6, 2019
Accepted on December 20, 2019

[1] Corresponding author: hadian.31961@gmail.com

1. Introduction

Over the recent years, there has been a dramatic increase in the assessment of different aspects and scheme devise of Communicative Language Ability (CLA) to provide hypothetical fundamentals for the nature of communicative competence, ultimately supposed to be a combination of various elements which illuminated the degree of learners' mastery over a second language (L2). Organizational knowledge (Bachman & Palmer, 2010) or competence is one of CLA's currently most investigated components, encompassing those abilities of language users in managing the formal construction of language to create or identify grammatically correct sentences, to understand their propositional content and to systematize them to form texts.

One of the primary goals of organizational knowledge research during the past years has been to develop a rigorous understanding of linguistic and textual features in second language performances of language tests specially the IELTS exam (Banerjee, Franceschina, & Smith, 2007; Barkaoui, 2016; Iwashita & Vasquez, 2015), because it enlightens the value of test performance through the analysis of the features of the performance itself and clarifies the degree of mastery over English as the nowadays' academic and social lingua franca.

This study details research into responses to two writing tasks of accessible General Training IELTS-practice scripts, in the hope of contributing to a greater understanding of teachers and candidates of grammatical and textual features on which should be focused to develop band scores and accomplish each task. It builds on the organizational knowledge of CLA and the destiny-determining role of IELTS exam and intends to give a numerical and quantitative account of grammatical and textual features in two writing tasks of a corpus across three band scores to determine the features variations not only among bands but also between tasks.

Although the Academic module of IELTS exam, TEOFL test and various L2 written texts have been widely investigated in their linguistic and discourse features (Kennedy & Thorp, 2007; Mayor, Hewings, North, Swann, & Coffin, 2007; Riazi & Knox, 2013), the IELTS General Training has always been neglected to the best of the researchers' knowledge.

A few studies have been found that surveyed the essential qualities of its test takers' written performance at inter/intra band level in each task and compared its tasks in GK and TK knowledge features. Therefore, the results of this study would be advantageous for the researchers interested in using computational tools in corpus analysis, writing teachers of English, IELTS teachers, and prospective IELTS candidates who want to know the linguistic characteristics that distinguish one level of performance from another, how writing abilities progress with increasing proficiency, and how the linguistic and discursive features change in different writing tasks in the General Training IELTS. However, to pave the way for the aforementioned groups, this research analyzed the accessible General Training IELTS-practice scripts because of no access to authentic IELTS writings.

2. Literature Review

2.1 Communicative Language Ability (CLA)

The term communicative competence has been exchanged for many years and used extensively in rationalizations and clarifications of communicative language teaching and testing. Recently, researchers have shown interest in models proposed to represent constructs similar to communicative competence with language assessment in mind (Bachman & Palmer, 2010).

By the mid-1980s, language testing platform had evolved considerably from the emergence of communicative language testing which no longer aimed to evaluate students' knowledge of the language, especially vocabulary and grammar, but to design communicative language testing tasks (Brown, 2010)

of which the approach was mainly an opposition of the reliability and validity aspects of language testing during the 1960s (Fulcher, 2000). Due to the communicative language teaching approach, more researches on communicative language testing have been carried out and more awareness regarding its benefits has been raised to help teachers to test their students' ability to use the language in the realistic content-specific situations and tasks (Ahmadi & Montasseri, 2019; Moradian, Miri & Qassemi, 2015).

2.2 IELTS and CLA Investigations

In the recent decades, IELTS researches have increasingly been heightened due to the exam necessity as the language proficiency measurement tool of people intending to live and work in the societies where English is used as a language of communication.

Mayor et al. (2007) examined the errors, complexity and discourse of the Academic IELTS writing T2 scripts written by high-scoring (bands 7 & 8) and low-scoring (band 5) candidates. They determined text length, formal error rate, sentence complexity, the use of the impersonal pronoun *one*, thematic structure, argument genre and interpersonal tenor, as significant surface prognosticators of T2 scores.

Banerjee et al. (2007) compared the linguistic characteristics of scripts written by Chinese and Spanish L1 candidates in response to both tasks of the Academic IELTS writing, scored at bands 3 to 8. They found a greater display of lexical variation and sophistication at increasing IELTS band levels, salient vocabulary gains at lower levels, grammatical accuracy as a good discriminator of proficiency level regardless of task type and test taker L1 and critical effects of L1 and writing tasks on some of these criteria.

McNamara, Crossley and McCarthy (2010) used the Coh-metrix tool to investigate linguistic differences between high and low-proficiency writers.

The results indicated that the three most predictive features of essay quality were found to be syntactic complexity, lexical diversity, and word frequency.

Crossley, Weston, McLain Sullivan, and McNamara (2011) also tried to quantitatively and computationally investigate the differences between the linguistic features and cohesion in essays of Grade 9 and Grade 11 students and college freshmen. They found that more sophisticated words and more complex sentence structures were made use of as grade levels improved. Conversely, fewer cohesive features were found in the scripts as a function of grade level. Therefore, the authors illustrated that linguistic development comes about in the later stages of writing development and that this development is first and foremost interrelated with producing texts that are less cohesive and more elaborate.

Crossley, Salsbury, and McNamara (2012) tried to pinpoint a range of linguistic features of L2 writing through quantitative methods of Coh-metrix online tool in a corpus of 100 writing samples of L2 leaners. They found that the strongest predictors of an individual's proficiency level were word imagability, word frequency, lexical diversity, and word familiarity.

More recently, Riazi and Knox (2013) provided an in-depth comparison of the linguistic and discourse characteristics of the Academic IELTS writing T2 scripts written by three L1 candidate groups at three different band levels. They found that 6 and 7 band score scripts were longer and comprised a higher amount of low-frequency words, greater lexical diversity, and more syntactic complexity than did the scripts of lower bands. However, high-scoring scripts were not necessarily more cohesive than low-scoring scripts. This study also found significant differences in terms of some linguistic characteristics (e.g., lexical diversity) across L1 groups.

Iwashita and Vasquez (2013) investigated the features of discourse competence in IELTS speaking part 2 and examined the relationship between

these idiosyncratic features and the IELTS speaking band descriptors. This profound analysis revealed that some features of discourse (e.g., use of a wider range of conjunctions, more accurate use of referential expressions) were more distinctively observed in the higher-level test-taker performance than the lower level test-takers, but other features (e.g., ellipsis and substitution, use of reference) were not clearly distinguished across the levels.

Barkaoui (2016) carried out a numerical study of changes in the linguistic characteristics of IELTS repeaters' responses to the Academic IELTS T2 across three bands. He revealed more inclusion of longer introductions and conclusions, linguistic accuracy, syntactic complexity, lexical density, diversity and sophistication, cohesion, fewer informal and more formal features, more hedges, and fewer self-mentions in the higher bands scripts in test occasion one. Moreover, he confirmed that the higher writing scores were devoted to the longer scripts with greater lexical diversity and lexical sophistication, greater syntactic complexity, more self-mentions, and fewer contractions.

Although all of the above-mentioned studies have provided important insights into grammatical and discourse features of the IELTS exam and well-acknowledged the effects of L1, band score and task factors on the characteristics of L2 writers' texts, a detailed study of organizational knowledge components in the General Training IELTS particularly the case of writing tasks has been somewhat neglected. To fill this gap, the current study, building on the investigations of Riazi and Knox (2013) and Barkaoui (2016), made a computational analysis of some of the totally ignored GK and TK features of this type of writing to reveal each band differences and individual tasks distinctive characteristics in a corpus of the General Training IELTS-practice materials.

3. Methodology

This study quantitatively and computationally addressed two components of organizational knowledge, three band scores and task types as variables to find both the differences among the writing scripts of the online General Training IELTS-practice materials of bands 7, 8 and 9 and the variations between the two types of writing tasks in GK and TK features.

Table 1
The Analyzed Corpus

Bands	T1	T2	Total	Sum
7	30 letters	30 essays	60	180
8	30 letters	30 essays	60	
9	30 letters	30 essays	60	

It hypothesized the greater inclusion of GK and TK features in the higher band scripts (9) than those of 8 and in those of 8 more than those of the lower one (7) and assumed no differences between the tasks in these two types of knowledge features presentation. Therefore, it reviewed a corpus of 180 Online General Training IELTS-practice writings (Table 1), numbering 30 in each of the three band scores (7, 8 and 9) and 90 in each of two tasks. The investigated scripts, which were written by different writers, varied from each other in topic and coincided on in the number of scripts written formally and informally.

The band scores of the writing performances had already been determined in task achievement, grammar, lexical resources, cohesion and coherence in the websites and other online resources. Nevertheless, three Ph.D. Iranian IELTS teachers, majored in English Language Teaching and trained informally to score IELTS writings with at least four years of experience, were asked to score a corpus of 230 writings based on the public version of the band descriptors of the writing tasks of the IELTS General Training. Ultimately, the researchers engaged those scripts upon which the teachers had the most agreement on the bands in the corpus.

Table 2
Adopted Taxonomy

	Writing features	Evaluation criteria	Coh-Metrix Index
Grammatical Knowledge (Bachman & Palmer, 2010)	Fluency	1. **Number of words per script**	3
	Syntactic complexity (Connor & Mbaye, 2002)	1. **Left-embeddedness** (number of words before the main verb of main clauses) (mean)	67
		2. **NP density** (number of modifiers (e.g., determines, adjectives) per NP) (mean)	68
		3. **Syntactic similarity** (uniformity and consistency of the syntactic constructions in the text)(Sentence syntax similarity, all combinations, across paragraphs, mean)	73
		4. **Logical operators** (variants of *or, and, not,* and *if–then*)(incidence or the number of word classes or constituents per 1,000 words)	52
	Lexical Features (Connor & Mbaye, 2002)	1. **Lexical variation MTDL** (Type-Token Ratio not varying as a function of text length)	48
		2. **Word frequency** (how frequently particular words occur in the English language) (CELEX word frequency for content words, mean)	92
		3. **Polysemy for content words** (the number of senses of a word) (mean)	100
		4. **Hypernym for noun and verb** (the number of levels in a conceptual taxonomic hierarchy that is above (i.e., superordinate to) a word)(mean)	103
		5. **Familiarity for content words:** (How frequently a word appears in print) (mean)	96
		6. **Concreteness for content words:** (How concrete or nonabstract a word is, on the basis of human ratings) (mean)	97
		7. **Imagability for content words:** (How easy it is to construct a mental image of the word in one's mind, according to human ratings)(mean)	98
		8. **Meaningfulness (** how strongly words associate with other words and how likely words are to prime or activate other words) (mean)	99
Textual Knowledge (Bachman & Palmer, 2010)	Cohesion and Coherence (Connor & Mbaye, 2002)	1. **Conceptual cohesion** (the extent to which the content of sentences or paragraphs is similar semantically or conceptually)	
		a) **LSA overlap**, adjacent sentences (similarity in meaning or conceptual relatedness between sentences) (mean)	38
		b)**LSA overlap**, adjacent paragraphs (similarity in meaning or conceptual relatedness between paragraphs) (mean)	42
		2. **Coreference cohesion** (when a noun, pronoun, or noun phrase refers to another constituent in the text)	
		a) **Argument overlap**, adjacent sentences (how often two sentences share nouns with common stems) (mean)	29
		b)**Content word overlap**, all sentences (how often sentences share content words) (mean)	36
		3. **Connectives density**: (All connectives incidence score)	50
		4. **Casual cohesion** (the extent to which sentences are related by causal cohesion relations, appropriate only when the text refers to events and actions that are related causally) (Causal verbs and causal particles incidence)	60
5. Text Easability Principle Component (PC) Scores			
		a) **Text Easability PC Narrativity** (z score)	12
		b) **Text Easability PC Word concreteness** (z score)	16
6. Readability			
		a)**Coh-Metrix L2 Readability**	106

The estimates of intra-class coefficient reliability as a measure of interrater reliability among the websites as the first rater and three IELTS teachers are available in supplemental materials.

To investigate the organizational features of the corpus, the researchers adopted the taxonomy of Bachman and Palmer (2010) of organizational

knowledge, including GK and TK, operationalized based on the four writing features of Writing Competence model of Connor and Mbaye's (2002) (Table 2).

The computer program used to make linguistic text analysis was Coh-metrix 3.0 (Graesser, McNamara, Louwerse & Cai, 2004), used in various researches to investigate written texts (McNamara et al., 2010; Riazi & Knox, 2013).

Several analyses were conducted on the corpus of two writing tasks of the Online General Training IELTS-practice scripts (Table 1) to measure the features of GK and TK, itemized in Table 2, across three bands and two tasks to address each research question of the study.

Firstly, mean ($\overline{X}$), standard deviation (SD), skewness (Sk) and kurtosis (Ku) or what is generally called descriptive statistics were necessary to gain a holistic understanding of the corpus under study in each of three bands across two tasks (supplemental materials).

Secondly, investigating the first research question concerning the differences between the band scores in GK and TK, the researchers conducted tests of normality (supplemental materials) to demonstrate the dispersion of 22 features of organizational knowledge. The non-parametric test of Independent-Samples Kruskal-Wallis (Table 3) was used for the features of nonnormal distribution across three bands. The features which distributed normally needed the Levene statistics to indicate the dis/satisfaction of homogeneity pre-condition (Table 4). Therefore, the parametric test of One-way ANOVA (Table 4) and its post hoc of Tukey-b (Table 5), exactly showing where the differences occurred, were applied to compare the elements distributed both normally and homogeneously. Table 6 and 7 show the comparison among band scores in

elements that were distributed normally but not homogeneously through the post hoc of Dunnett's T3.

Thirdly, the second research question, concerning the comparison between two tasks in GK and TK presentation, required another test of normality. Moreover, it was necessary to apply Independent-Samples T-Test for those having normal and homogeneous distributions and the nonparametric test of Mann-Whitney U for the features distributing nonnormally across two tasks.

4. Results

As the research questions necessitated, the results were reported in two parts, band score comparisons and tasks comparisons respectively.

Table 3

The Independent Samples Kruskal Wallis Test for Syntactic Similarity

a. Kruskal Wallis Test b. Grouping Variable: Band Score *. p < .05

	Band Score	N	Mean Rank	Chi-Square[a, b]	Df	Asymp. Sig.*
Syntactic similarity (Task One)	7	30	68.18	36.851	2	.000
	8	30	39.92			
	9	30	28.40			
Syntactic similarity (Task Two)	7	30	75.50	64.183	2	.000
	8	30	23.38			
	9	30	37.62			
Content word overlap (Task Two)	7	30	55.43	6.514	2	.039
	8	30	40.72			
	9	30	40.35			
Coh-Metrix L2 Readability (Task one)	7	30	43.13	1.672	2	.433
	8	30	50.53			
	9	30	42.83			
Coh-Metrix L2 Readability (Task Two)	7	30	60.03	13.934	2	.001
	8	30	38.05			
	9	30	38.42			

4.1 Band Score Comparison

The supplemental materials indicated descriptive statistics or a general representation of fluency, 12 grammatical and nine textual features in the intended corpus. They also provided the results of the normality tests, which

indicated that all of the elements of GK and TK except the features of syntactic similarity, Coh-metrix L2 readability (in both tasks) and content overlap (in T2) were distributed normally.

Table 4
ANOVA for GK and TK Feature

		T1					T2				
		Sum of Squares	df	Mean Square	F	Sig.*	Sum of Squares	df	Mean Square	F	Sig.*
Fluency	BG[a]	10416.356	2	5208.17	11.188	.000	110441.867	2	55220.933	55.656	.000
	WG[b]	40500.100	87	465.518			86319.733	87	992.181		
Left-embedded-ness	BG[a]	168.586	2	84.293	783.612	.000	53.935	2	26.967	59.695	.000
	WG[b]	9.359	87	.108			39.303	87	.452		
NP Density score	BG[a]	.530	2	.265	15.289	.000	47.293	2	23.647	1047.85	.000
	WG[b]	1.508	87	.017			1.963	87	.023		
Logical operators	BG[a]	5854.056	2	2927.02	24.784	.000	5439.971	2	2719.985	13.403	.000
	WG[b]	10275.029	87	118.104			17656.005	87	202.943		
Lexical variation	BG[a]	17219.836	2	8609.91	19.427	.000	4409.459	2	2204.729	30.778	.000
	WG[b]	38557.950	87	443.195			6232.178	87	71.634		
Word frequency	BG[a]	.055	2	.027	1.368	.260	14.347	2	7.173	371.492	.000
	WG[b]	1.734	87	.020			1.680	87	.019		
Polysemy	BG[a]	37.197	2	18.599	41.830	.000	2.895	2	1.447	8.942	.000
	WG[b]	38.683	87	.445			14.081	87	.162		
hypernym	BG[a]	16.647	2	8.323	173.684	.000	17.786	2	8.893	354.405	.000
	WG[b]	4.169	87	.048			2.183	87	.025		
Familiarity	BG[a]	30.733	2	15.367	.302	.740	712.902	2	356.451	5.560	.005
	WG[b]	4430.420	87	50.924			5577.448	87	64.109		
Concreteness	BG[a]	186.811	2	93.406	.185	.831	2551.965	2	1275.982	3.738	.028
	WG[b]	43867.297	87	504.222			29695.060	87	341.323		
Imagability	BG[a]	132.360	2	66.180	.144	.866	2423.627	2	1211.813	4.547	.013
	WG[b]	40040.250	87	460.233			23188.261	87	266.532		
Meaningfulness	BG[a]	49.023	2	24.511	.112	.894	44.109	2	22.054	.091	.913
	WG[b]	18993.382	87	218.315			20998.214	87	241.359		
LSA,adjacent sentences	BG[a]	.019	2	.010	7.158	.001	.652	2	.326	121.257	.000
	WG[b]	.116	87	.001			.234	87	.003		
LSA, adjacent paragraphs	BG[a]	.039	2	.019	18.411	.000	.285	2	.143	15.077	.000
	WG[b]	.091	87	.001			.823	87	.009		
Argument overlap	BG[a]	.407	2	.096	8.230	.001	.723	2	.911	263.439	.000
	WG[b]	2.152	87	.076			.642	87	.003		
Content word overlap	BG[a]	.005	2	.002	4.105	.020					
	WG[b]	.053	87	.001							
Connectives	BG[a]	9482.455	2	4741.22	10.002	.000	41217.087	2	20608.544	180.168	.000
	WG[b]	41242.096	87	474.047			9951.521	87	114.385		
Casual cohesion	BG[a]	7062.426	2	3531.21	22.444	.000	4425.880	2	2212.940	17.936	.000
	WG[b]	13687.794	87	157.331			10733.840	87	123.377		
Text Easability Narrativity	BG[a]	.232	2	.116	.306	.737	4.064	2	2.032	8.559	.000
	WG[b]	32.934	87	.379			20.655	87	.237		
Text Easability, Word Concreteness	BG[a]	.224	2	.112	.152	.859	1.372	2	.686	1.139	.325
	WG[b]	63.826	87	.734			52.391	87	.602		

*. p <.05 a. Between Groups b. Within Groups

According to the test of homogeneity (Levene statistic) for the normally distributed features, the non-homogeneity of variances of left-embeddedness, polysemy for content words and LSA overlap for adjacent paragraphs in both tasks, LSA overlap for adjacent sentences and casual cohesion in T1 and fluency, NP density score, word frequency, hypernym and connectives in T2

was revealed.However, logical operators, lexical variation, coreference cohesion for both tasks, fluency, NP density score, word frequency, hypernym, content word overlap and connectives in T1 and LSA overlap for adjacent sentences and casual cohesion in T2 and familiarity, concreteness, imagability, meaningfulness and indices of text easability in both tasks proved to have homogeneous variances. Therefore, ANOVA (Table 4) and a post hoc of Tukey-b test (Table 5) were employed to compare the group of elements with homogeneous variances across the bands.

Table 5
A post hoc of Tukey-b test

		T1			T2		
		1	2	3	1	2	3
Fluency	7	169.43					
	8		185.43				
	9		195.57				
Density score	7	.6611			.7353		
	8		.8234			.7683	
	9			.8723			.8807
NP density	8	.7587					
score	7	.7719					
	9		.9277				
Logical	7	37.4940			46.9037		
operators	8		54.4677			55.6453	
	9		54.7343				65.9267
Lexical	7	101.057			93.9583		
variation	8		113.1490			107.2703	
	9			134.513		109.9720	
Hypernym	7	1.6248					
	8		2.4863				
	9		2.5807				
Familiarity	7				575.0520		
	8					568.9269	
	9					569.2494	
Concreteness	7				358.5448		
	8					369.2254	
	9					370.3691	
Imagability	7				391.6037		
	8					401.9813	
	9					403.1495	
LSA, adjacent	7				.1893		
sentences	8					.2807	
	9						.3973
Argument	7	.5255			.4418		
overlap	8		.6427			.5733	
	9		.6843				.7870
Content word	7	.06617					
overlap	8		.08293				
	9	.06843					
Connective	7	66.6617					
	8	70.1150					
	9		89.9563				
Casual	7				29.5463		
Cohesion	8					44.3877	
	9					44.4567	
Text Easability	7				-.32693		
PC Narrativity	8					-.77923	
	9					-.77617	

The results revealed statistically significant differences between the bands with regard to left-embeddedness, a feature of writings of band 9 more than 8 and those of 8 more than 7 in both tasks (Table 6).

Number of modifiers per NP or NP density score, a GK feature, proved to be higher in band 9 than the other two bands although there were no differences between the others in this regard in T1. Significant differences among bands were also indicated in this feature of GK in T2 in a respective order of 9, 8, and 7 from the highest to the lowest band (Table 6).

The nonparametric test of Independent-Samples Kruskal-Wallis in Table 3 for the non-normally distributed feature of syntactic similarity indicated significantly statistical differences between the bands in both tasks. It pointed out that while there were no differences between bands 8 and 9 scripts, those of band 7 were syntactically less complex in both tasks. That is, there were other more effective features which differentiated between bands 8 and 9 than syntactic similarity.

The analysis of logical operators (Table 5) revealed that the two higher bands included more uses of logical operators than band 7 in T1. However, its application in T2 was in a step-like fashion, from the lowest, 7 then 8 then 9, with significant differences among them.

Table 6

Dunnett-T3 Test Results for GK and TK Features

Dependent Variable	(I)	(J)	Mean difference (I-J)	Std. Error	Sig.*
Fluency	7	8	-19.33333*	7.20969	.031
(T2)		9	-82.06667*	7.48554	.000
	8	7	19.33333*	7.20969	.031
		9	-62.73333*	9.50912	.000
	9	7	82.06667*	7.48554	.000
		8	62.73333*	9.50912	.000
Left-embeddedness	7	8	-.94733*	.05830	.000
(T1)		9	-3.25867*	.09788	.000
	8	7	.94733*	.05830	.000
		9	-2.31133*	.09239	.000
	9	7	3.25867*	.09788	.000
		8	2.31133*	.09239	.000
Left-embeddedness	7	8	-.77567*	.15411	.000
(T2)		9	-1.88633*	.17272	.000
	8	7	.77567*	.15411	.000
		9	-1.11067*	.19175	.000
	9	7	1.88633*	.17272	.000
		8	1.11067*	.19175	.000
NP density score	7	8	-.82100*	.03795	.000
(T2)		9	-1.77400*	.03490	.000
	8	7	.82100*	.03795	.000
		9	-.95300*	.04307	.000
	9	7	1.77400*	.03490	.000
		8	.95300*	.04307	.000
Word frequency	7	8	.83833*	.03146	.000
(T2)		9	.85533*	.03624	.000
	8	7	-.83833*	.03146	.000
		9	-.01700	.03949	.963
	9	7	-.85533*	.03624	.000
		8	.01700	.03949	.963
Polysemy	7	8	-.18460	.11021	.268
(T1)		9	-1.44667*	.19039	.000
	8	7	.18460	.11021	.268
		9	-1.26207*	.20132	.000
	9	7	1.44667*	.19039	.000
		8	1.26207*	.20132	.000
Polysemy	7	8	-.38533*	.08913	.000
(T2)		9	-.37533*	.10479	.002
	8	7	.38533*	.08913	.000
		9	.01000	.11595	1.000
	9	7	.37533*	.10479	.002
		8	-.01000	.11595	1.000
Hypernym	7	8	-.93900*	.03767	.000
(T2)		9	-.94700*	.03870	.000
	8	7	.93900*	.03767	.000
		9	-.00800	.04581	.997
	9	7	.94700*	.03870	.000
		8	.00800	.04584	.997
LSA, adjacent sentences	7	8	.01860	.00903	.131
(T1)		9	-.03573*	.00797	.000
	8	7	-.01860	.00903	.131
		9	-.01713	.01107	.332
	9	7	.03573*	.00797	.000
		8	.01713	.01107	.332

The study of lexical variation (Table 5) indicated a ladder arrangement of 7, 8 and 9 one after the another in T1. In contrast, there were no differences between 9 and 8 in this lexical feature while differing from 7 significantly in T2.

Despite of no differences among the bands in word frequency in T1 (Table 5), the corpus analysis showed significant variances between two higher bands and band 7 in T2

Table 7
Dunnett-T3 Test Results for TK Features

Dependent Variable	(I) Band Score	(J) Band Score	Mean Difference (I-J)	Sig.*
LSA overlap, adjacent paragraphs (T1)	7	8	-.00323	.982
		9	-.04543*	.000
	8	7	.00323	.982
		9	-.04220*	.000
	9	7	.04543*	.000
		8	.04220*	.000
LSA overlap, adjacent paragraphs (T2)	7	8	-.03267	.355
		9	-.13233*	.000
	8	7	.03267	.355
		9	-.09967*	.001
	9	7	.13233*	.000
		8	.09967*	.001
Connectives (T2)	7	8	-22.12533*	.000
		9	-52.21733*	.000
	8	7	22.12533*	.000
		9	-30.09200*	.000
	9	7	52.21733*	.000
		8	30.09200*	.000
Casual cohesion (T1)	7	8	-21.60577*	.000
		9	-12.53910*	.001
	8	7	21.60577*	.000
		9	9.06667*	.049
	9	7	12.53910*	.001
		8	-9.06667*	.049

$p < .05$

Table 6 shows the results of comparison between bands in polysemy for content words. The results presented differences between bands 7 and 9 and also between 9 and 8 but not between 7 and 8 in T1. That is, the level of

polysemy was higher in band 9 than 7 and 8 in T1. This table pointed to not only the surpass and similarity of bands 9 and 8 but also their differences with 7 in T2 in this feature.

According to Table 5 and 6, no differences were found between bands 8 and 9 in the lexical feature of hypernym, but band 7 differed from them with less levels in this feature in both T1 and T2.

Table 8
Group Statistics and Independent Samples T-Test Results

				Levene [a]					
	Task	Mean	SD		F	Sig.[d]	t	df	Sig. (2-tailed)*
NP density score	T1	.8194	.15133	Y[b]	159.420	.000	-12.037	178	.000
	T2	1.7827	.74394	N[c]			-12.037	96.353	.000
Logical operators	T1	48.8987	13.46201	Y	5.360	.022	-3.281	178	.001
	T2	56.1586	16.10917	N			-3.281	172.556	.001
Lexical variation	T1	103.7336	10.93476	Y	53.761	.000	4.343	178	.000
	T2	116.2400	25.03431	N			4.343	121.767	.000
Word frequency	T1	2.2854	.14176	Y	145.671	.000	8.479	178	.000
	T2	1.8856	.42435	N			8.479	108.619	.000
Polysemy	T1	3.7589	.43674	Y	27.746	.000	6.215	178	.000
	T2	4.4281	.92335	N			6.215	126.924	.000
LSA overlap, adjacent para	T1	.1725	.03817	Y	52.123	.000	-26.568	178	.000
	T2	.5027	.11156	N			-26.568	109.551	.000
Coreference cohesion	T1	.6175	.16956	Y	.271	.604	.694	178	.488
	T2	.6007	.15442	N			.694	176.465	.488
Connectives	T1	75.5777	23.87339	Y	.100	.752	-10.112	178	.000
	T2	111.6436	23.97766	N			-10.112	177.997	.000
Casual cohesion	T1	39.4636	13.05120	Y	2.315	.130	4.602	178	.000
	T2	49.2072	15.26920	N			4.602	173.788	.000

a. Levene's Test for equality of variances
b. Equal variances assumed.
c. Equal variances not assumed.
d. $p \leq 0.05$ is nonhomogeneous and has no equal variances.

Despite the lexical feature of meaningfulness which proved not to make differentiation between bands in both tasks, the other three elements of familiarity, concreteness and imagability for content words proved to significantly differ across the bands of T2 but not T1(Table 4). That is, the index of word familiarity was significantly greater for scripts of band 7 than

those of 8 and 9. However, writings of band 8 and 9 significantly overdid those of 7 in concreteness and imagability for content words.

With regard to the first feature of conceptual cohesion, developed in Coh-metrix as an evaluation criterion of semantic cohesion and coherence, LSA overlap for adjacent sentences in T1, a significant difference was designated only between 9 and 7 and the override of the former. However, T2 showed this type of overlap in scripts of 9 more than those of 8, exceeding those of 7 (Table 6).

Table 9
The Results of Mann-Whitney U Test

	Task Type	Mann-Whitney U	Asymp. Sig. (2-tailed)*
Fluency	T1	19.500	.000
	T2		
Left-embeddedness	T1	263.000	.000
	T2		
Density score	T1	3975.000	.830
	T2		
Syntactic similarity	T1	3388.000	.058
	T2		
hypernym for noun and verb	T1	2979.500	.002
	T2		
Familiarity for content words	T1	2569.000	.000
	T2		
Concreteness for content words	T1	3409.000	.067
	T2		
Imagability for content words	T1	3554.500	.156
	T2		
Meaningfulness	T1	2982.000	.002
	T2		
LSA overlap, adjacent sentences	T1	343.000	.000
	T2		
Content word overlap	T1	4011.500	.912
	T2		
Text Easability PC Narrativity	T1	264.000	.000
	T2		
Text Easability PC Word concreteness	T1	3739.500	.374
	T2		
Coh-Metrix L2 Readability	T1	3004.000	.003
	T2		

* $p < 0.05$

The text analysis for the second feature of conceptual cohesion, LSA overlap for adjacent paragraphs, revealed a significant difference just between bands 9 and 7 in T1 (Table 7). That is, the scripts of bands 9 and 8 indicated these features of TK equally but more than those of band 7. T2' scripts indicated the equality between 7 and 8, but their difference with and exceedance of band 9 writing performances.

The results of coreference cohesion analysis (Table 5) of argument overlap demonstrated both the equality of bands 8 and 9 and their outdoing band 7 in T1. This great care of cohesion was more visible in band 9 scripts than 8 and in 8 more than 7 in T2. Concerning next index of this type of cohesion (Table 5), the results indicated significantly greater content word overlap for band 8 scripts of T1 and 8 and 9 in T2 (Table 3).

In regard with connectives, Table 5 illustrated significant differences between two lower bands and band 9 in T1. However, significant differences were found among all bands in this feature in T2.

The last feature of textual knowledge, casual cohesion, displayed fluctuations with band 8 surpassing the other two, although 7 and 9 varied from each other considerably in T1 (Table 7). This feature showed no differences between scripts of band 8 and 9 while they were casually more cohesive than those of band 7 (Table 5).

Although the results indicated no differences between the scripts of different bands in text easability (word concreteness) in both tasks (Table 4) and in text narrativity and Coh-metrix L2 readability in T1 (Table 4), they demonstrated greater text narrativity for scripts of 8 and 9 (Table 5) and higher Coh-metrix L2 readability for those of 7 in T2 (Table 3).

4.2 Task comparison

Answering the second research question, focusing on tasks comparisons in GK and TK features, firstly required the tests of normality (supplemental

materials). They indicated that NP density, logical operators, lexical variation, word frequency, polysemy, LSA overlap for adjacent sentences, coreference cohesion, connectives and casual cohesion were distributed normally. Therefore, an Independent-Samples T-test helps to make comparisons across tasks in these features (Table 8).

It proved the differences between two tasks in all of these element except coreference cohesion. According to Table 8, T2 had more NP density, logical operators, word frequency, LSA overlap, connectives and casual cohesion but less lexical variation and polysemy levels. To compare the tasks in the elements distributed non-normally (supplemental materials), the researchers utilized the nonparametric Two Independent-Samples Test of Mann-Whitney U (Table 9). It illustrated significant differences between tasks and the surpass of T2 in fluency, left-embeddedness, meaningfulness, hypernym and LSA overlap for adjacent sentences although it showed no differences in syntactic similarity, concreteness and imagability, content word overlap and text Easability (word Concreteness). This table also indicated the surpass of T1 in indices of familiarity and Coh-metrix L2 Readability.

5. Discussion

The findings of the current corpus analysis pointed out that writing scripts of the online General Training IELTS-practice resources of bands 7, 8 and 9 indicate significant differences in some but not all of the grammatical, textual, readability and easability features. They also confirmed significant differences between the two writing tasks of the online General Training IELTS-practice resources in some of these features but not all. In other words, the application of GK and TK could be influenced by both task type and band levels, in line with Banerjee et al. (2007).

The interpretation of results starts with the first element of Table 2, fluency. It was apparent from post hoc results of Table 6 and 7 for both tasks that the higher the band score, the more fluent were the writers and the more were amounts of the words produced. This finding is in line with Mayor et al. (2007), Crossley and McNamara (2012) and Riazi and Knox (2013) who considered text length one of the influential factors that characterizes the L2 high-quality writings and results into higher test scores. On the other part of the survey, task comparisons revealed an evident task effect on fluency because of at least 100 more words that T2 requires and the penalty being received for too short essays (Table 9)

The findings for the next feature of GK, left embeddedness, were not in agreement with Riazi and Knox (2013) who believed in no differences between bands 5, 6 and 7 in terms of this feature.

The post hoc table (Table 6) indicated that the high scoring writings included more words before the main verb of main clauses, were structurally denser, had more embedded constituents and put heavier loads on working memory of the readers, thus applied more complex sentences in both tasks. That is, it was indicated that the more the number of the words before the main verb, the higher the band score or proficiency level. In the second part of this survey, the semi-formal/neutral discourse of essay of T2 (Table 9) also proved its surpass in this feature because of the discussion that should be given for a point of view in this task.

Based on the mean number of modifiers such as determiners and adjectives used per noun phrase or what is called NP density in Table 2, the results indicated that the writing performances with higher NP density score obtained the highest band score, in line with Barkaoui (2016), in T1. However, this feature succeeded in differentiating among bands in T2 more than T1 because the writers using longer NPs gained higher scores, namely, 9, those using long

ones got 8 and small ones were scored 7 in T2, consistent with Crossley and McNamara (2012) finding texts written by more advanced writers denser in NPs. Furthermore, the task comparison of Table 8 revealed that T2 required longer NPs because of justifications and evaluations that should be present in an essay.

The analysis of syntactic similarity index (Table 3) illustrated that 8 and 9 scored scripts tended to be structurally more diverse and complex because of lower scores in similar sentence structures in both tasks, despite of Barkaoui (2016) who found no band score differences in terms of this feature. This summation is based on Crossley et al. (2011) who believed that low syntactic similarity indices signposted more complex syntax.

However, the task comparisons results showed no differences between the tasks in structural variety and difficulty.

The results of logical operators' analysis indicated the two higher band score writings to be analytically denser and more working memory demanding than those of band 7. It also specified that levels of density were completely band differentiating because the scripts of 9 involved more logical operators than those of 8 and 8 more than 7 in T2. In addition, the results confirmed the greater use of logical operators in T2 to follow English discursive writing conventions in factual information or an outline presentation in an essay.

Generally, the results of writings analysis for syntactic complexity is in line with Mayor et al. (2007) considering it a feature of high-band score writings. T2's superiority in syntactic complexity reflected the findings of Beers and Nagy (2011) who believed the argumentative essays lean towards complex sentences because writers need to establish close causal links between facts and their opinions, and a syntactically complex sentence allows them to make the connections clear. They intensified that logical reasoning is used to support the arguments, the quality of which can be increased by using more

sophisticated sentence structures. Therefore, it is expected that variability in writing quality might be explained particularly well in T2 by the use of syntactically complex sentences.

The analysis of the first lexical feature, lexical variation, indicated that scripts of high bands had greater lexical diversity, consistent with Cumming et al. (2006), Crossley, Louwerse, McCarthy, and McNamara (2007), and Crossley and McNamara (2012) who found a greater variety of words in the essays of more advanced writers. It is also in agreement with Banerjee et al. (2007) and Riazi and Knox (2013) who found lexical diversity in type-token ratio as one of the indicators of the highest band scores. However, the results of band comparisons in T2 both confirmed their finding because of more vocabulary varieties in bands 8 and 9 than 7 and contradicted them due to no differences in this regard between 8 and 9 bands (Table 5). On the other hand, consistent with findings by Banerjee et al. (2007), this study indicated more lexical variation in T2 scripts, including a greater variety of words, than T1 (Table 8).

Despite of Barkaoui (2016), the findings indicated that word frequency or lexical sophistication could not differentiate between the bands in T1 (Table 5). That is, the bands investigated in this study, were the same in the required processing, reading and understanding time in that task. In contrast, in line with them, it was found that more proficient L2 writers used less frequent words in T2. That is, the scripts of band 7 could be read normally more quickly, understood better and needed less word processing time than those of 8 and 9 because of the application of more frequent words consistent with Meara and Bell, (2001) (Table 6). It was also in line with Riazi and Knox (2013) who found more use of less frequent words in band 7 than in 5 and 6 in the writings of T2 of the Academic IELTS. Besides, task comparison results indicated the use of more frequent words in T1 of the Online General Training IELTS-

practice scripts because of common everyday situations that should be written about and informal or semiformal styles being used. However, this is contrary to the study conducted by Banerjee et al. (2007) who found that test-takers (particularly at lower IELTS band levels) were more likely to use high-frequency words in T2 of the Academic IELTS than in its T1.

Inferential statistical analysis for polysemy as a way of measuring the ambiguity of words evaluated in this study for content words pointed to the greater involvement of words with multiple meanings such as spring, which is a season, a cured metal that will return to its previous shape after being pressed, and a place where water comes up naturally from the ground, in band 9 writings, not being considered before.

It means that they included more words with numerous senses which make text ambiguous and slow to process in both tasks, not in line with Crossley and McNamara (2012) with the notion of less ambiguous words in the L2 essays of advanced writers. In addition, this multiplicity was found with the words of T2 more than T1 for the evidence, ideas evaluations and justifications that should be provided in this task.

The evaluation of hypernym in both tasks, ignored in the literature, proved that the writings with assigned bands 8 and 9 included more abstract words, having limited superordinate levels and few distinctive features and attributes that can be pictured in the mind. That is, they encompassed mainly the words with less superordinate levels in a conceptual taxonomic hierarchy, therefore, the words were more abstract than concrete. Furthermore, it is apparent form Table 9 that T2's words have less superordinate levels because of the more abstract and complex ideas that should be communicated in this task.

Regarding the word information measure of familiarity for content words or word exposure, the corpus analysis indicated the easiness of the writings of the

scripts of band 7 in T2 because of the use of more familiar words. It shows that these scripts have some daily conversational words which are not corresponding to the style of writing. However, writers of 8 and 9 of T2 facilitated their writings through evoking mental and sensory images and using words more meaningful to the reader because of the higher amounts of concreteness and imagability of content words, in a good agreement with Crossley and McNamara (2012) who found that advanced writers utilized more concrete in their essays. The comparisons between tasks also confirmed that T1 more activated the use familiar words although the tasks were the same in arousing mental and sensory images. Conversely, T2 indicated surpass with the measure of how strongly words associate with other words, and how likely words are to prime or activate other words because of the higher lexical meaningfulness index. In other words, T1 was illustrated to engage the words weakly associated with each other despite of no differences among the corpus bands in this regard, not in agreement with Crossley and McNamara (2012) who found the use of less meaningful words as a feature of high-quality L2 essays.

The investigation of two indices of LSA for adjacent sentences and paragraphs or semantic cohesion showed the equality of two lower bands but the preponderance of band 9 in similarity of meaning or conceptual relatedness. That is, bands 7 and 8 had the same level of cohesion and similarity between sentences and paragraphs, however, band 9 showed to be more cohesive because of higher similarity of sentences and paragraphs to their adjacent ones, in agreement with Crossley et al. (2007) and Landauer, McNamara, Dennis, and Kintsch (2013). In line with these studies, this research indicated T2's bands greater differences from each other in this aspect, decreasing as the bands lessens (Table 6). Therefore, people would feel difficult to read scripts of lower bands because of their lower LSA cosine score.

Likewise, task comparisons illustrated the dominance of T2 in these two aspects of conceptual cohesion through using the words that are close in meaning in a piece of text (Table 8 & 9).

The next far too little attention-paid to feature of TK is coreference cohesion of which predominance in band 9 in both tasks was indicated in Table 5 in T1. Argument overlap for adjacent sentences index from among several types of coreferentiality indices was measured in this study to see the extent to which two adjacent sentences share common arguments such as nouns, pronouns and noun phrases, consequently; the readers tend to feel easy to read sentences with argument overlaps. The results indicated it as another distinctive textual feature of higher bands only in T1. However, the texts of band 8 in T1 and 7 in T2 used content word overlap to indicate of text cohesion, to construct larger units of meaning in a text and to determine paragraph boundaries. On the other hand, the tasks themselves also equalized in reading and comprehension easiness (Table 8) in regard with these two measures of coreference cohesion.

The analysis of all connectives (additives, casual, temporal and clarification) as text comprehension facilitators (Murray, 1997) across bands helped to assume that the writing performances of band 9 were more coherent and organized while bands 7 and 8 did not vary significantly (Table 5), not consistent with Iwashita and Vasquez (2015) who found no coherence/cohesion variations across bands and also not in agreement with Crossley and McNamara (2012) finding skilled writers creating less cohesive essays. The findings of T2 analysis (Table 7) more contradicted with them because the higher the bands, the more cohesive they were, not also in agreement with Kennedy and Thorp (2002) who believed in the less use of overt cohesive devices in the higher bands of IELTS. However, in line with McCutchen (1986) (who found more local coherence between sentences of eighth grade essays than those of sixth grade), the findings confirmed the

existence of great amounts of cohesive devices in the scripts of more proficient L2 writers than those of less proficient ones. In task comparisons, this study found greater amount of connectives in T2 because of its need to organize and link information coherently and cohesively and to develop a continuous theme as well as connect ideas with topics.

The texts of bands 8 and 9 only in T2 proved higher in formality construct and easier for readers to detect, process and understand because of the higher text narrativity closely associated with everyday oral conversation although the other index of easability, text easability (word concreteness), could make no differentiations neither among the bands nor between the tasks.

The two higher bands scripts of T2 proved to be more difficult to read and comprehend that those band 7 because of the lower Coh-Metrix L2 Readability. From among the two tasks, T1 was easier for reading and comprehension because of the higher amount of this index.

The analysis of writing scripts of T1 indicated that band 8 more involved causal relationship between clauses that refer to events and actions, although this feature did not cause to get the highest band score. It means that band 8 writers' answers were more like stories with an action plot in T1. Besides, the writings of bands 8 and 9 were more casually cohesive in T2 (Table 5) despite of Riazi and Knox (2013) who found T2 writings of the Academic IELTS of higher bands not necessarily more cohesive. It was also visible in Table 8 that casual cohesion markers were more frequent in T2 to enable the writers to establish their ideas clearly and support surely their argument with relevant examples or evidences.

6. Conclusions

The study set out to assess the writing samples of varying band scores of the Online General Training IELTS-practice scripts based on the components of organizational knowledge. It has shown that although there were no differences

between bands 9 and 8 in five features in T1 and T2, between 8 and 7 in five other ones in T1 and one feature in T2 or even between 7 and 9 in two elements in T1, high scored writing performances showed more competency organizationally, grammatically and textually. Therefore, to increase the bands in both tasks, the IELTS candidates and their teacher ought to focus on the syntactic complexity measures such as using complex sentences, varied in structure, with high amounts of options of or, and, not, and if– then to connect either noun and verb phrases, nevertheless, the NPs should become longer via determiners and adjectives.

Lexically, to improve the band significantly from 7 to 9, they should use a more varied and less familiar but more sensory images motivating vocabulary range, including synonyms, collocation, semantically related words and words having different meanings, instead of repetition of a set of specific words. Although the study did not show significant textual differences between successive bands, it did substantiate that enough care should be taken of cohesion and coherence criteria such as causality, argument and concept unity, formality construction, and coherence markers to increase the bands from 7 to 9.

Regarding the other purpose of the present research in comparison of two tasks of the Online General Training IELTS-practice writings, it has been revealed that although T2 writings were longer, more involved complex and logically related sentences, and more evoked lexical meaningfulness, T1 overdid in word exposure, narrativity and readability features. In addition, it proved that the tasks equally transferred a sense of syntactic consistency and similarity in the level of comprehension, lexical concreteness, mental sense motivation and text easability (word concreteness). This research contributed to more understanding of the General Training IELTS exam, however, a number of limitations need to be considered. For instance, it did not analyze

the real IELTS writings, written under the pressure of time and exam condition, moreover, the researchers had to revise, rewrite and ignore lots of mistakes that occurred in samples of 8 in order to compensate for the lack of scripts with band 9. Besides them, the corpus analysis findings would become more reliable if the future researchers manually recheck the corpus to verify the software outputs for the final results in order to deal with the probable mistakes that a software would make. Further contrastive researches, investigating not only the three higher bands differences, but also the lower ones such as 4, 5 and 6 distinctions in more indices of GK and TK features, seem so interesting and additive to further understandings of IELTS as an international-destiny affecting test.

References

Ahmadi, A., & Montasseri, Z. (2019). Interactional Competence in Paired vs. Group Oral Tests. *Teaching English Language*, 13(1), 1-26.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

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*.

Barkaoui, K. (2016). What changes and what doesn't? An examination of changes in the linguistic characteristics of IELTS repeaters ' Writing Task 2 scripts. *IELTS Research Reports*, *3*, 1–55.

Barkaoui, K. and Knouzi, I. (2012). Combining score and text analyses to examine task equivalence in L2 writing assessment. In E. Van Steendam, M. Tillema, G. Rijlaarsdam and H. van den Bergh (Eds), *Measuring writing: Recent insights into theory, methodology and practices,* vol 23 of *Studies in Writing*. Amsterdam: Elsevier

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing: An Interdisciplinary Journal*, *24*(2), 183–202.

Brown, D. H. (2010). *Language Assessment.: Principles and Classroom Practices*. London: Longman.

Connor, U., & Mbaye, A. (2002). Discourse approaches to writing assessment. *Annual Review of Applied Linguistics, 22*, 263-278.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading, 35*(2), 115-135.

Crossley, S. A., Louwerse, M., McCarthy, P. M. and McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, *91*, 15–30.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing, 29*(2), 243-263.

Crossley, S., Weston, J., McLain Sullivan, S., & McNamara, D. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28*, 282–311

Fulcher, G. (2000). The 'communicative'legacy in language testing. *System*, *28*(4), 483–497.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. doi: 10.3758/BF03195564

Iwashita, N., & Vasquez, C. (2015). An examination of discourse competence at different proficiency levels in IELTS Speaking Part 2 proposals. *IELTS Research Report Series*, *5*, 1–44. Retrieved from www.ielts.org/researchers

Kennedy, C., & Thorp, D. (2007). A corpus-based investigation of linguistic responses to an IELTS academic writing task. *Studies in Language Testing*, *19*, 316–377.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2013). *Handbook of latent semantic analysis*. Psychology Press.

Mayor, B., Hewings, A., North, S., Swann, J., & Coffin, C. (2007). A linguistic analysis of Chinese and Greek L1 scripts for IELTS academic writing task 2. In L. T. and P. Falvey (Eds.), *IELTS collected papers: Research in speaking and writing assessment* (pp. 250–313). Cambridge: Cambridge University Press.

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language, 25*(4), 431-444.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86.

Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect, 16,* 5–24.

Moradian, M., Miri, M., & Qassemi, Z. (2015). Dynamic Assessment and Microgenetic Development of EFL Teachers' Classroom Interactional Competence. *Teaching English Language*, *9*(2), 1-36.

Murray, J. (1997). Connective and narrative text: The role of continuity. *Memory & Cognition*, *25*(2), 227–236.

Rezaei, O., Barati, H., & Youhanaee, M. (2016). Evaluation of IELTS Preparatory Courses in Iran: Teaching Practices and Strategies in Focus. *Teaching English Language*, *10*(2), 47-71.

Riazi, A. M., & Knox, J. S. (2013). An investigation of the relations between test-takers' first language and the discourse of written performance on the IELTS Academic Writing Test, task 2. *IELTS Research Reports, 2,* 1–87.