DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Same Language, Different Functions: A Cross-Genre Analysis of Chinese EFL Learners' Writing Proficiency

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Citation	Qin, Wenjuan. 2017. Same Language, Different Functions: A Cross-Genre Analysis of Chinese EFL Learners' Writing Proficiency. Doctoral dissertation, Harvard Graduate School of Education.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797222
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Same Language, Different Functions:

A Cross-genre Analysis of Chinese EFL Learners' Writing Proficiency

Qualifying Paper

Submitted by

Wenjuan Qin

March 2015

Acknowledgements

I would like to express deep gratitude to Dr. Paola Uccelli, Dr. Catherine Snow and Dr. James Kim for their support and guidance in this study. Many thanks to everyone who provided feedback, suggestions, and insight on the many drafts of this paper, including Shireen Al-Adeimi, Emily Phillips Galloway, Christina Dobbs and Dr. Eileen McGowan. I want to also thank Steve Crooks, Yerrie Kim and Christopher McCormick from EF Educational First for their support in data collection across the research sites. Finally, I would thank Diwen Shi, Sarah Surrain and Kathleen Zheng from the Harvard Graduate School of Education for their help in collecting, transcribing, coding and scoring the data.

Table of Contents

Abstract

Secondary school learners of English as a foreign language (EFL) in China constitute a rapidly growing yet understudied population. This study examined Chinese secondary school EFL learners' writing proficiency in two genres, argumentative essays and narratives. Research on adolescents' native language writing has documented consistent higher quality in narrative than in essay writing; and, aligned with the higher linguistic demands of essay writing, more complex lexico-syntactic features in adolescents' essays. To investigate cross-genre differences in EFL learners, a total of 200 English written compositions (100 essays and100 narratives) produced by 100 EFL Chinese secondary school learners were scored for quality, lexico-syntactic, and genre-specific discourse features. Unlike prior research on native language writing, no significant differences in quality across the two genres were found. However, in line with prior research, Analysis of Variance results revealed that argumentative essays displayed a higher lexico-syntactic complexity. Regression analysis identified a distinct set of predictors of writing quality for each genre. Controlling for length, lexico-syntactic complexity and diversity of discourse organizational markers were identified as predictors of argumentative quality. Conversely, controlling for length, narrative quality was only predicted by the frequency of stance markers. Results are discussed in relation to pedagogical implications and directions for future research.

Introduction

English is spoken by approximately 1.75 billion people worldwide, accounting for a quarter of the world's population (Yang, 2006). Remarkably, the majority of these language users are now non-native English speakers, including approximately 350 million English users and learners in China (Kirkpatrick, 2007). China, like many other non-native English speaking countries, is now gradually coming under the influence of English, through its use as a dominant second language in academia, business and commerce, media, science and technology (Kachru, Kachru, & Nelson, 2009).

Among many pedagogical initiatives that have been undertaken in China to improve English as a foreign language (EFL) proficiency, the domain of writing development has only recently come to the forefront of discussions (Wang, 2014; Yang & Gao, 2013). Writing proficiency in English has been recognized as decisive for students' success in academic and professional endeavors (Grabe & Kaplan, 1996). Despite emerging recognition of the important role of writing and considerable research focused on the college level, there is still a great deal to learn about characteristics of Chinese EFL learners' writing, especially at the secondary school level, and about factors related to EFL learners' overall writing quality. Seeking to address this research gap, the current study investigated secondary school Chinese EFL students' writing proficiency in two different genres, namely narrative and argumentation. By comparing writing performance across narrative and argumentative genres, this study seeks to reveal specific

language demands faced by these learners. Two main questions drove this study: (1) do differences in overall writing quality and in the incidence of lexico-syntactic features vary by genre?; and (2) does the incidence of lexico-syntactic and discourse features predict overall writing quality within each genre? The ultimate goal is to generate findings that will inform a pedagogical approach that will be specially attuned to the needs of Chinese EFL students in both narrative production and evidence-based argumentative writing.

In the next section, we briefly introduce the sociocultural pragmatics-based view of language as our guiding conceptual framework and then offer a review of relevant developmental linguistics findings that have guided this study. In the subsequent sections, we explain the study design and present results from an exploration into cross-genre differences in EFL learners' writing performance as well as a regression analysis with linguistic features predicting writing quality. Finally, results are interpreted in relation to previous studies on native language writing, and pedagogical implications for EFL writing instruction as well as questions for further exploration are discussed.

Literature Review

Writing Proficiency across Two Genres

A sociocultural pragmatics-based view of language development understands oral and written language learning as the result of individuals'

socialization and enculturation histories (Halliday, Matthiessen, & Matthiessen, 2014; Ninio & Snow, 1996; Ochs, 1993; Snow & Uccelli, 2009; Uccelli et al., 2014). This view of language entails that being a skilled language user in some social contexts does not guarantee adequate language proficiency in other social contexts. Thus, it would not be unexpected to find writers who can excel at writing in one genre but not in another. Writing is a highly complex task that integrates cognitive processing, deployment of linguistic knowledge, and awareness of the social context in which the written communication takes place (Gee, 2001). The quality of writing is expected to be influenced by the writer's ability to flexibly use a variety of language forms and functions that are attuned to different communicative contexts, specific audiences, and purposes (Hyland, 2009; Ravid & Tolchinsky, 2002; Schleppegrell, 2002).

Informed by this pragmatics-based view of language, the present study was conducted to examine secondary school EFL learners' writing proficiency across two genres. Despite of the rapidly growing population of adolescent EFL learners around the world, and particularly in China, this is still a relatively understudied group. The majority of recent empirical studies on EFL writing in China have been conducted at the undergraduate or graduate level (Li & Wharton, 2012; Liardet, 2013; Liu, 2013; Miao & Lei, 2008; Ong, 2011; Qin & Karabacak, 2010), mostly overlooking adolescence despite it being a period of critical growth in writing proficiency (Berman, 2008). Additionally, to my knowledge, the majority of these studies exclusively focused on argumentative writing proficiency, with scarce

research contrasting EFL learners' writing across genres. Instead of looking at the mechanics of writing or spelling, this study focuses on lexico-syntactic and discourse features essential to the production of two types of written communication, namely, personal narratives and evidence-based argumentation. Given the scarcity of research on EFL adolescent learners' writing development, in this study I draw heavily from prior discourse research conducted with native speakers of English (e.g. Berman & Nir-Sagiv, 2007; Uccelli, Dobbs & Scott, 2013), as well as from research on heritage language learners who grow up speaking a language other than English at home in English-dominant societies (e.g. Cummins, 1979, 1981). First, developmental linguistics studies that examine native speakers' proficiency in different genres have shed light on the differences in complexity involved in acquiring the discourse forms expected for writing narrative versus argumentative genre. Second, research on English learners growing up in English dominant societies – guided in particular by Cummins' theories – has focused on differences across contexts of language use and emphasized the impact of the social environment on learners' diverse constellations of discourse proficiencies.

Sequence of Genre Acquisition. In this study, genre is defined as "a distinctive type or category of literacy composition" following Swales (1990, p. 33). Narrative and argumentative texts are two distinct genres of discourse defined by different communicative functions (Berman, 2008; Grabe, 2002; Paltridge, 2001). The progress in mastering new genres in one's native language has been

characterized by Martin (1989) and Schleppegrell (2004) as moving progressively across three categories: (1) personal genres, such as narratives, and recounts; (2) factual genres, such as procedures and reports; and (3) analytic genres, those focused on analysis and argumentation (e.g. argumentative essays). Recent empirical data on native language writing development support this conclusion, showing that while written narrative structures tend to be well mastered by age ten, argumentative writing constitutes a later developmental accomplishment (Berman & Nir-Sagiv, 2007). A number of empirical studies particularly examining writing produced by middle and high school students further suggest that adolescents are typically able to produce a higher level of writing quality for narratives than for argumentative essays (Crowhurst, 1980, 1990; Engelhard Jr, Gordon, & Gabrielson, 1992; Hall-Mills & Apel, 2013; Reed, 1992; Scott & Windsor, 2000). However, all of these studies investigated native language writing. To our knowledge, to date no similar contrastive studies have been conducted for the EFL learners.

These documented developmental sequences make sense because writing across these two genres requires the writer to activate distinct cognitive and linguistic processes. Narratives are agent-oriented, that is, they focus on people, their actions, and the unfolding of events in a temporal order that tends to mimic real world events (Berman & Slobin, 2013; Hickmann, 2003). Argumentative essays are topic-oriented, requiring the writer to impose a logical structure to interrelate ideas in a coherent manner, and to discuss claims and arguments, often

in a hierarchical format (Grabe, 2002). Apart from the different macro-level organization, the two genres also vary in their micro-level linguistic features. At the lexical level, research on monolingual students has reported that argumentative texts contain a higher proportion of structurally complex, semantically abstract and low-frequency vocabulary items than narratives (Berman, Nayditz, & Ravid, 2011; Berman & Nir-Sagiv, 2007). At the syntactic level, compared to written narratives, monolingual writers' argumentative texts tend to display more complex structures, including larger proportion of embedded clauses and higher level of information packing in single clauses (Beers & Nagy, 2009; Berman & Nir-Sagiv, 2007; Ravid & Berman, 2010). Developmental linguistics research on native language writing has identified adolescence as a period of critical growth in the development of these lexical and syntactic skills (Berman & Nir-Sagiv, 2007).

The Effect of Contexts. From a pragmatics-based developmental perspective, speakers are first enculturated at home into the language of face-to-face interaction, which typically prepares them for colloquial conversations in their respective communities (Heath, 1983, 2012; Ochs, 1993). However, being able to successfully participate in academic discourses has been documented to be a challenging task for many colloquially fluent monolingual or bilingual students with scarce opportunities to be socialized into more academic ways of using the language (Schleppegrell, 2004; Snow & Uccelli, 2009; Uccelli et al., 2014). For heritage language speakers growing up in English-dominant countries, Cummins proposed the well-known distinction between Basic Interpersonal Communicative

Skill (BICS) and Cognitive Academic Language Proficiency (CALP). This distinction has triggered substantive research that documents both the more challenging nature of CALP and also the often unsupportive instructional conditions in which it is expected to be mastered (Cummins, 1980; Cummins, 1981).

EFL learners' Writing Performance across Genres. Taking the theoretical and empirical perspectives above as a point of departure, this study will first explore if Chinese EFL learners' writings in two genres achieve different levels of overall quality. On the one hand, similar to the research on native language writing, we might expect that these students' EFL written narratives will be more advanced in quality than their EFL argumentative essays. Moreover, following Cummins' BICS vs. CALP distinction and the documented challenges heritage language learners face in learning academic English (Snow & Uccelli, 2009). On the basis of the degree of complexity across genres, one might also expect narrative writing to display a higher quality given that this genre has been characterized as less cognitively and linguistically demanding than argumentative writing.

On the other hand, given that EFL instruction at secondary school in China has mostly focused on test preparation – including preparation for the conventional argumentative essay writing prompts of tests like the TOEFL or IELTS – students can be presumed to have had more opportunities to learn and practice academic argumentation than narrative production. Thus, an alternative

scenario might be that, in contrast to native language writers, Chinese EFL writers would demonstrate a higher level of proficiency in writing argumentative essays than narratives. In light of these two possible alternatives, the first goal of the present study was to investigate Chinese secondary school EFL learners' overall narrative and argumentative writing quality and to examine potential differences in the frequency or diversity of key lexical and syntactic features by genre.

Predictors of Writing Quality within each Genre

In traditional educational settings, trained, professional readers (e.g. teachers) typically assess writing quality. These evaluations have important consequences for the writer because these judgments provide a source of feedback and determine passing or failing grades. During the past two decades, with the development of computer-assisted language analysis, a number of empirical studies have been conducted to explore more objective and efficient measures of writing quality by investigating the predictive relationship between linguistic features, usually coded or tagged in computer programs, and the holistic quality scores given by human raters.

Research on Native English Speakers' Writing. Research has been conducted, at both college and secondary level, to explore predictors of native-English speakers' writing quality. For instance, Witte and Faigley (1981) provided a descriptive analysis demonstrating how cohesive linguistic ties (e.g. *reference, substitution, ellipsis, conjunction and lexical reiteration*) enhanced writing quality

of college students' argumentative writing. More recently, McNamara and her colleagues adopted quantitative methods to explore predictors of argumentative writing quality in U.S. college students using Coh-Metrix, an automated text analysis tool that provides a large array of linguistic indices for the analysis of cohesion and coherence (Graesser, McNamara, Louwerse, & Cai, 2004). McNamara, Crossley, and McCarthy (2009) found that the three most predictive indices of essay quality were syntactic complexity, lexical diversity and word frequency, but none of the cohesion indices correlated with essay ratings. A later study conducted by the same group of researchers on college freshmen's writing revealed that essays scored as better by human raters were characterized by more different words (types), few personal pronouns, a strong link to previously given information and conclusion paraphrases (Crossley, Roscoe, & McNamara, 2011). At the secondary school level, Connor (1990) looked at 150 argumentative essays written by 16-year-old high school students from the United States, England and New Zealand, and found higher rated essays contained more advanced argumentative strategies, such as using data and warrant to support argumentation, incorporating writer's personal experience, knowledge of the subject and awareness of the audience's values; as well as using sophisticated syntactic factors (featuring nominalizations, propositions, passives and specific conjuncts). More recently, Uccelli, Dobbs, and Scott (2013) extended prior research by identifying additional discourse-level components that are predictive of writing quality, including organizational markers that signal argumentative structure and epistemic

stance markers which entail degree of possibility, certainty or acknowledgment of the writer's beliefs about the truth of certain assertions or state of affairs.

Fewer studies have explored the relationship between linguistic features and writing quality of narrative texts, and most of them have been conducted with younger students. For instance, Cameron et al. (1995) studied 9-year-old English-speaking children's narrative writing, and concluded that the writing quality was positively related to frequency of cohesive indices, number of longer words and sentences, as well as higher vocabulary variety. In addition, Olinghouse and Leaird (2009) found that, in second and fourth grade students' written narratives, diversity, length and sophistication of vocabulary explained unique variance in writing quality.

Research on EFL Writing. Many studies in the field of EFL research have also focused on identifying linguistic features that are predictive of overall writing quality. The majority of these studies focused on argumentative writing, especially those at the undergraduate or graduate level. For instance, taking advantage of the large corpus of Test of Written English (TWE) essays, researchers (Ferris, 1994; Frase, Faletti, Ginther, & Grant, 1999; Grant & Ginther, 2000) found that highly rated EFL essays were usually longer, with longer average word length and with higher frequencies of certain lexical and grammatical categories (e.g. nouns, hedges, conjuncts). Moreover, Jarvis, Grant, Bikowski, and Ferris (2003) and Friginal, Li, and Weigle (2014) emphasized that the quality of a written text may depend less on the use of individual linguistic features than on the underlying

patterns of co-occurrence of these features. Apart from these lower-level linguistic measures, Zhao (2010) developed and validated an analytic rubric that measured authorial voice strength, that is, "how the writer made linguistic and discourse choices to negotiate representation of ourselves and take on the discourse of our communities" (p. 9), in L2 argumentative writing. EFL writers' authorial voice was found to have strong positive correlation with overall writing quality. At the secondary level, we have found only one study looking at linguistic features and writing quality of graduating Hong Kong high school students (who were not typical EFL learners because of the special language status of English in Hong Kong[1]). In this study, Crossley and McNamara (2012) found that L2 writers categorized as highly proficient did not produce essays that are more cohesive but instead produced texts that are more lexically sophisticated, featuring greater lexical diversity and lower-frequency vocabulary.

EFL narrative writing has received less attention than argumentative writing. Only three studies are considered relevant. Using a sample of Japanese college students, Ishikawa (1995) found the predictors of English narrative writing quality was syntactic complexity (as measured by total words in error-free clauses) and composition length (as measured by total number of error-free clauses). Kormos (2011) did not directly model the relationship between linguistic features

---

[1] In Hong Kong there has been a pursuit of a "trilingual, biliterate" language policy that recognizes Cantonese, Mandarin Chinese and English as spoken languages, and written Chinese and English as written languages (Bolton, 2002). By 2001, the census results indicated that, overall, 43 percent of the population now claims to speak English.

and writing quality, but instead compared linguistic features in narrative writing of secondary school EFL learners with that of native English speakers, showing that the major differences between L1 and foreign language narrative writers were lexical variety, sophistication and range of vocabulary. Additionally, Kang (2005) investigated how Korean EFL learners used discourse strategies for establishing textual cohesion and marking written register, showing that Koreans' narrative writing in English was highly influenced by linguistic strategies of their native language (e.g. highly frequent use of demonstrative references and repetition).

Contrasting the Relationship across Genres in EFL learners. To my knowledge, there are only two studies contrasting predictive relationships between linguistic features and writing quality across genres in adolescents. Studying genre differences in native English writing, Beers and Nagy (2009, 2011) examined the relationship between syntactic complexity and writing quality across genres and using two types of syntactic measures. They found syntactic complexity as measured by words per clause was positively correlated with quality for argumentative essays but not for narratives. Clauses per T-unit was positively correlated with quality for narratives, but negatively correlated with quality for essays. Such findings intrigued us to further explore other linguistic features and investigate how their relationship with writing quality may vary across genres.

Research Questions

In order to better understand secondary school EFL learners' writing proficiency across genres and unpack the relationship between certain linguistic features and writing quality within each genre, the present study seeks to answer the following research questions:

1. Comparing writing proficiency across genres:

   (a) Does the overall quality of Chinese secondary school EFL learners' written personal narratives and argumentative essays differ?; (b) Does the frequency or diversity of key lexico-syntactic features vary by genre in Chinese secondary school EFL learners' written personal narratives and argumentative essays?

2. Predicting overall writing quality within each genre:

   (a) What genre-specific discourse features characterize these secondary school EFL learners' writing? (b) Controlling for essay length and participants' grade level, what lexical, syntactic, and discourse features are predictive of overall writing quality in each genre?

Methods

Participants

As displayed in Table 1, the sample consisted of 100 secondary school EFL learners, whose ages ranged from 11 to 17 years. Students' grade-level ranged

from 6th to 11th grade. The sample was relatively balanced by gender, with 53 boys and 47 girls. The majority of participants came from middle to upper middle class families as indicated by their parents' high educational level. All participants received comparable standard instruction in the same language institute in east China and, according to the language school records, had achieved intermediate or upper-intermediate language proficiency in English.

Data Collection

Compositions written by secondary school EFL learners were collected using a digital platform as part of students' regular classroom activities in the language institute. During the computer-based writing assessment, each student was asked to respond in writing to a narrative prompt and an argumentative prompt in 90 minutes (40 minutes for each text plus a 10-minute break). Both writing prompts were on a similar topic to optimize comparison across genres (see Appendix A).

Data Analysis

Data were transcribed, segmented into clauses, coded and analyzed using the transcription conventions and automated language analysis tools of the CHILDES program (MacWhinney, 2000). To homogenize the formatting of the data and avoid any subjective impressions of writing quality due to mechanical mistakes, all unconventional spellings, capitalizations and punctuations were removed from the texts, and were recorded on a separate coding tier. After data

were transcribed and verified by a second researcher, the following measures were

generated (see Appendix B for detailed explanation and examples for each

measure):

Writing Quality and Length Measures. Writing quality was estimated

using two genre-specific six-point-scale holistic scoring rubrics (NAEP, 2011)[2].

These rubrics offered the advantage of providing genre-specific yet comparable

scores across genres along similar dimensions, including content, organization, use

of details, voice and effective use of language. Two native English-speaking,

experienced teachers who were blind to the research questions, scored each text.

They were first trained to score 20% of the data. After agreement was reached on

all scoring criteria for that first batch of data, they double-scored the remaining 80%

of the data and achieved satisfactory inter-rater reliability, with $\kappa = .83$ for

argumentative writing and $\kappa = .80$ for narrative writing. Following prior standard

procedures used in standardized writing assessments (Uccelli et al., 2013), when

two scores reached either exact or adjacent agreement, they were summed to form

the final scores (with a range of possible scores from 2 to 12). When the two

scorers were more than one point apart, a third expert scorer intervened to settle

the disagreement, and this score was doubled as the final score.

---

[2] The NAEP rubric was originally scaled within grade, making it unrealistic to compare scores cross grade level. However, for the purpose of the present study, scorers are blind to students' demographic information and are instructed to use a "single ruler" to measure performance across the whole sample. Therefore, the same scale is used for students at all grade levels.

In addition, text length was measured by the total number of clauses. A clause is defined as "a unit that contains a unified predicate, … [i.e.,] a predicate that expresses a single situation (activity, event, state). Predicates include finite and nonfinite verbs, as well as predicate adjectives" (Berman & Slobin, 2013, p. 660).

Lexical Measures. This set of measures captured word-level characteristics for both narrative and argumentative texts:

1) *Word Length*: measures the raw frequency of polysyllabic words; specifically, words with three or more syllables (e.g., *perspective, transportation*) (Wimmer, Köhler, Grotjahn, & Altmann, 1994).

2) *Nominal Abstractness*: measures the raw frequency of abstract nouns used in students' writing, using a four-level semantic abstractness scale developed in previous studies (Berman & Nir-Sagiv, 2007; Ravid, 2006), that categorizes nouns from most concrete (e.g., *bike, Mary*) to most abstract (e.g., *perspective, authority*).

3) *Academic Vocabulary:* measures raw frequency of words that appear on the Academic Vocabulary List (Coxhead, 2000).

4) *Lexical Diversity:* was captured through the widely used vocD measure, which reduces the impact of length in estimating the variety of words used in a text (McKee, Malvern, & Richards, 2000).

The measures for academic vocabulary and lexical diversity were automatically generated using CHILDES (MacWhinney, 2000). Two researchers

doubly coded all narrative and argumentative texts for word length and nominal abstractness. Inter-rater reliability was assessed using Cohen's kappa, with $\kappa = .96$ for word length, and $\kappa = .92$ for nominal abstractness.

Syntactic measures. Two measures based on Beers and Nagy (2009) were generated to assess syntactic complexity in both narrative and argumentative texts.

1) *Words per Clause*: A higher ratio of words per clause is associated with the literate or academic register, which indicates the writers' skill to convey information in a more concise manner (Snow & Uccelli, 2009).

2) *Clause per T-unit:* T-units are defined as thematic units of complete and autonomous meaning, corresponding to a main clause plus all the subordinate clauses embedded in it (Hunt, 1983). A higher ratio of clauses per T-unit reflects more frequent usage of subordinated and embedded sentence structures, which are characteristic of academic writing. To establish inter-rater reliability, two researchers double-coded 20% of transcripts. Inter-rater reliability was estimated by applying Cohen's kappa statistics, with $\kappa = .89$ for T-unit coding and $r = .93$ for clause coding.

Genre-specific Discourse measures. While same the lexico-syntactic features can be investigated across genres, argumentative and narrative texts require attention to distinct components at the discourse level. Well-formed text construction in each genre requires the writer to apply different discourse markers to generate the text structure and express a personal evaluative stance. Using

research-based genre-specific features, two discourse dimensions were coded for each genre: (1) discourse organization and (2) writer's evaluative stance:

Organizational Markers in argumentative texts: Following research on metadiscourse analysis (Hyland, 2005; Uccelli et al., 2013), markers used to explicitly signal the organization of argumentative text structure were identified and coded with four subcategories:

1) *Frame markers* signal the sequence of arguments or counter-arguments (e.g., *first of all, on the other hand*);

2) *Code Glosses* introduce an example or paraphrase (e.g., *for example, in other words*);

3) *Transition markers* signal additive, adversative or causal relations between clauses and paragraphs (e.g. *moreover, even though, because*). Temporal markers and the coordinating conjunction "*and*" were not coded;

4) *Conclusion markers* explicitly state the writer's summary or conclusion of the essay (*e.g. in conclusion; all in all*).

Stance Markers in argumentative texts: Based on Berman (2004), Reilly et al. (2002) and Uccelli et al. (2013), stance markers in argumentative essays were identified and coded for type of marker:

1) *Deontic markers,* indicate a writer's absolute or categorical stance or viewpoint towards an assertion (e.g. *everybody should do…, it is wrong to…*);

2) *Epistemic markers* display the writer's stance towards the truth of an assertion. Three subtypes of epistemic markers were identified: a) *Epistemic hedges* which express degree of uncertainty, signaling a writer's cautiousness when making assertions (e.g., *it might be true…, it is possible that…*); b) *Epistemic boosters* which emphasize the writers' commitment to the truth of an assertion (e.g., *it is absolutely true…*); c) *Personal beliefs* signal that assertions are the result of one's or others' personal beliefs (e.g. *I think, people assume that…*).

Organizational Markers in narrative texts: Following a classic framework for cohesion analysis (Halliday & Hasan, 2014), narrative texts were coded for transitional connectives that denote temporal and logical relations at the inter-clausal level. These markers were identified and subsequently coded for type of relationship signaled including 1) additive (e.g. *furthermore, that is), 2)* adversative (e.g. *but, although), 3)* causal (e.g. *because, therefore)* and 4) temporal/aspectual (e.g. *first, last, finally)* relationships. The coordinating conjunction "*and*" as well as repetitive use of "*then*" were not included in this coding.

Stance Markers in narrative texts: Borrowing tools from the evaluative elements analysis of personal narrative discourse developed by Peterson and McCabe (1983), narrative texts were coded for four types of stance – or evaluative – markers used by the writer to evaluate or color the

narrative, offering an implicit or explicit subjective interpretation of the

narrated events. Stance markers in narrative texts were subsequently coded

for type of marker: a) internal states (e.g. *markers that express emotions,*

thoughts), b) rhetorical moves (e.g., *similes and metaphors; exaggeration*),

c) objective judgments (e.g. *a means by which the narrator uses other*

people to evaluate the narrated event) d) evaluative qualifiers (e.g.

adjective, adverbs, intensifier).

To estimate inter-rater reliability, 20% of the data was doubly coded by two

researchers, with high reliability. Cohen's kappa statistics was $\kappa = .88$ for

argumentative organizational markers, $\kappa = .91$ for argumentative stance markers, κ

$= .92$ for narrative transitional markers and $\kappa = .87$ for narrative evaluative

markers. Then researchers independently coded the remaining 80% of the data.

Analytic Plan

To address the first set of research questions, descriptive statistics were

generated for writing quality scores by genre and by participants' grade level.

After examining the score distribution, a one-way repeated measures (narrative vs.

argumentative) analysis of variance (ANOVA) with writing quality scores as the

dependent variable, was conducted. Then descriptive statistics were generated for

lexical, and syntactic measures by genre. After examining the distribution of

variables, Multivariate Analysis of Variance (MANOVA) was conducted to

investigate whether the frequency or diversity of lexical and syntactic features

varied by genre. To address the second set of research questions, a descriptive analysis was first conducted to examine genre-specific discourse features. Then, correlational analysis informed the construction of a series of hierarchical regression models, independently for narrative and argumentative genres. Principal Component Analysis (PCA) was performed to reduce collinearity among certain linguistic variables.

Results

Comparing Writing Quality across Genres (RQ1a)

Argumentative essays and narratives were double scored for overall writing quality (with range of possible writing quality scores of 2 to12). For narratives, the mean quality score was 7.31 with a standard deviation of 2.80, and an observed range from 2 to 12 points. Argumentative essays displayed the same range, but with a slightly higher mean quality score of 7.53 and a smaller standard deviation of 2.11. Though the difference between mean quality scores by genre was not statistically significant, $F(1, 99)=0.80$, $p=0.37$, a closer look at the data revealed an interesting statistically significant discrepancy across genres. The F test for the homogeneity of variances revealed that EFL learners' narrative performances displayed greater variability in writing quality scores than argumentative essays, (*F=1.78, p=.004*). As illustrated in Figure 1, the majority of essays (80 out of 100) displayed a score that fell in the upper half of the scale (6 - 10). In comparison, the

distribution of narrative scores was more spread-out across all levels of writing

quality scale, with an equal high number of narratives (41) displaying either a high

score (11 - 12) or a score at the lower-end (2 - 5) on the scale.
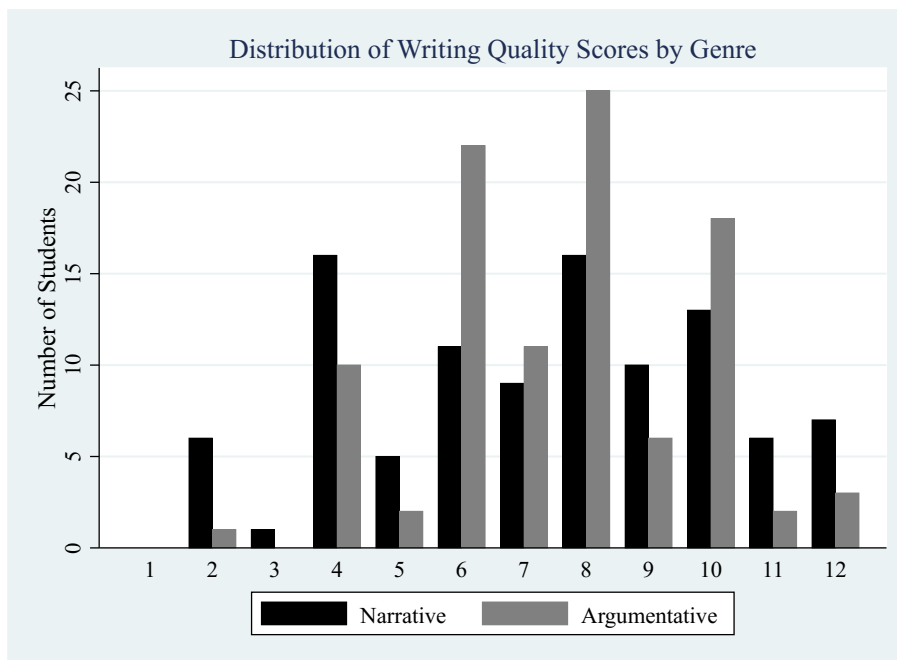
Figure 1. Distribution of writing quality scores by genre

Another intriguing finding is that the within-grade difference in scores

tends to be larger for argumentative than narrative writing. As shown in Table 2,

the high school argumentative score is about .68 standard deviations higher, on

average, than the middle school score, whereas the difference is about .31 standard

deviations in narrative writing. When scores were disaggregated further by

students' grade level, we found distinct patterns of cross-grade difference between

argumentative and narrative writing quality. For argumentative essays, there was a

steady increment of the average argumentative score from Grade 6 (mean = 6.39,

SD=1.97) to Grade 11 (mean = 8.71, SD=1.60), and a narrowing variability within

each grade. In contrast, narrative quality scores did not follow a typical

developmental pattern, with no association between the scores and participants'

grade level. For instance, the mean narrative score of Grade-9 students (mean =

7.46, SD = 2.67) was lower than that of Grade-7 (mean = 7.54, SD=2.87) and

Grade-8 (mean = 8.33, SD = 2.63) participants. Although none of the cross-genre

variance was shown to be statistically significant at grade level and the descriptive

statistics should be cautiously interpreted because of the small sample size, this

interesting pattern indicated the need to include students' grade level as a control

variable[3] in later regression analysis.

Cross-genre Variance of Lexico-syntatic Features (RQ1b)

Table 3 summarizes descriptive statistics for text length and lexico-

syntactic features in secondary school EFL learners' argumentative and narrative

writing. Multivariate Analysis of Variance (MANOVA) supports that cross-genre

difference does exist at the lexical level, $F(4, 185) = 23.60$, $p<.001$; Wilk's $\Lambda =$

0.66, and syntactic level, $F(2, 197) = 26.15$, $p<.001$; Wilk's $\Lambda = 0.79$. Follow-up

separate univariate analyses demonstrate that argumentative and narrative genres

vary significantly at each lexico-syntactic dimension. Argumentative essays

displayed a mean length of 32.81 clauses, whereas narratives were four clauses

longer on average, $F(1, 99) = 9.51$, $p = .003$. Despite being shorter on average,

argumentative texts displayed significantly higher frequencies of complex lexical

---

[3] Students' grade level was used as a control variable rather than a substantive predictor variable is due to the limited sample size at each grade level. We feel hesitate to over-interpret the cross-sectional differences, which should be further explored in future research.

features: a significantly greater frequency of polysyllabic words, $F(1, 99) = 18.66$, $p < .001$, semantically abstract nouns, $F(1, 99) = 63.85$, $p < .001$, and more diverse deployment of vocabulary, $F(1, 99) = 26.75$, $p < .001$. Surprisingly, narratives contained, on average, more academic vocabulary than argumentative essays, $F(1, 99) = 21.90$, $p < .001$. However, it is noteworthy that both genres display only limited use of academic vocabulary (Coxhead, 2000), ranging from an average of 1.36 to 2.76 academic words per text, including both genres[4]. In addition, argumentative texts also differed from narratives in syntactic features. Argumentative texts not only displayed significantly higher levels of sentence subordination, indicated by the higher ratio of clauses per T-unit, $F(1, 99) = 4.11$, $p = .05$, they also demonstrated denser information packing in single clauses, indicated by the higher ratio of words per clause on average, $F(1, 99) = 58.32$, $p < .001$. Thus, argumentative texts, despite being significantly shorter than narratives, tended to display a higher level of lexico-syntactic sophistication, as indexed by most measures, with only one surprising exception, i.e., frequency of academic vocabulary.

Genre-specific Discourse Features (RQ2a)

Table 4 exhibits both frequency and diversity of discourse markers coded in these secondary school EFL learners' argumentative and narrative texts. In

---

[4] Since the Academic Vocabulary variable captured limited variability across the sample and had minimum correlation with other lexical measures, it was excluded from the lexical complexity composite in later analysis. However, the limited presence of academic vocabulary in EFL learners' writing could merit special instructional attention.

argumentative essays, students used between 0 and 17 organizational markers, with an average of 7.07 per essay. Among the four types of organizational markers coded, inter-clausal transitional markers displayed the highest frequency within each essay and appeared in 97 out of the 100 essays. In addition, more than 60% of essays contained code glosses (*e.g. for example, such as*) to explicitly introduce specific examples or paraphrases to support argumentation, and nearly half of the sample used frame markers to explicitly signal sequence and organization of arguments (*e.g. first, second; on the other hand*). The least frequently used markers within a single essay were markers of conclusion. Whereas the frequency is expected to be low due to the functional use of one conclusion marker per essay, conclusion markers were only present in 37 out of the 100 essays in the sample.

Stance markers in argumentative essays were used less frequently than organizational markers, with an average of 3.25 instances per essay. The two most widely used types of stance markers were markers of personal beliefs (*e.g. I think*) and deontic markers of absolute stance (*e.g. you should, people must*), which appeared in 85 and 52 essays respectively. However, a closer look at the data revealed overabundant usage of these two types of stance markers in repetitive pattern, such as using "*I think…*" every time a new argument was introduced, or using "*you should…*" several times in a row. Such patterns might indicate the writer's colloquial use of language to sequence ideas rather than cautious deployment of these linguistic devices to indicate personal stance on the issue. On the other hand, only a third of the essays included epistemic hedges (*e.g. it might*

be, it is possible), with an average of less than one instance per essay. Finally,
epistemic boosters (*e.g. it is absolutely true*) were the least frequently used stance
marker both within and across essays.

Narrative discourse organization differs from that of argumentative
discourse, in that it usually follows a sequence of events in a temporal order, and
thus the organizational markers in narrative discourse mostly denote the transition
at micro inter-clausal level rather than at the macro discourse level. As expected,
temporal markers (*e.g. earlier, finally*) displayed the highest frequency, with an
average of 3.54 instances, and appeared in 87 out of the 100 narratives.
Comparably, 88 students used causal markers (*e.g. because, the reason that…*)
(mean = 3.05 per narrative) followed by adversative markers (*e.g. though,
however*) (mean = 1.91 per narrative). Finally, it is not surprising to notice that
additive markers (*e.g. also, too*) had a lower frequency, given that the most
frequently used colloquial additive marker "*and*" was excluded from the coding.

Finally, written narratives demonstrated a variety of evaluative markers
with an average frequency of 17.18 in total. All four types of evaluative markers
were identified in 80 out of the 100 narratives. Evaluative lexicon, that is, the use
of adjectives (*e.g. unforgettable*), and adverbs (*e.g. actively*), and intensifiers (*e.g.
really, very much*), was the most frequently used type of evaluation. Internal states
(e.g. emotion, hypothesis) and rhetorical moves (e.g. metaphor or exaggeration)
showed comparable frequencies, with more than three instances per text, on

average. Finally, 80 students also used other people's perspectives to evaluate the narrated events.

Correlations of Linguistic Features with Overall Writing Quality: Varied by Genre

Lexical Measures and Writing Quality. The sample demonstrated distinct patterns of correlations between lexical measures and writing quality in each of the two genres. As shown in Table 5, all three measures of lexical complexity (i.e., word length, lexical abstractness, and lexical diversity) displayed moderate-to-high correlations with argumentative writing quality, a relationship that remained statistically significant when controlling for text length, with the partial correlation coefficients ranging from .27 to .37 ($p < .01$). In comparison, whereas the pairwise correlations between lexical measures and writing quality of narratives seemed to be moderately strong, this relationship became non-significant when controlling for text length (see Table 6). After adjusting for length, only lexical diversity displayed a significant but low-to-moderate correlation with narrative quality (partial $r = .20$, $p = .02$). In addition, given that the three lexical measures were significantly associated with each other (r ranges from 0.40 to 0.82 in both genres). Thus Principal Components Analysis was performed to reduce collinearity among these variables before including them as independent variables in regression analyses (see Table 7).

Syntactic Measures and Writing Quality. Words per clause was the syntactic measure that best predicted quality, with a moderate and significant relationship with argumentative quality ($r = .32, p = .001$), and a low-to-moderate correlation with narrative quality that approached significance ($r = .19, p = .06$). However, no statistically significant relationship was detected between clauses per T-unit and writing quality in either genre. It is also interesting to note that there is limited association between these measures of syntactic complexity and measures of lexical complexity. Such lexical-syntactic disassociation will be discussed later in comparison to previous findings on native language writers.

Discourse Measures and Writing Quality. Finally, we were encouraged to see that most discourse markers coded in both argumentative and narrative writing captured individual variability relevant to predicting the variability in overall writing quality. For argumentative essays, both frequency and diversity of organizational markers displayed moderate pairwise correlations with writing quality, and when controlling for text length the relationship between diversity and quality of organizational markers remained significant (partial $r = .27, p = .006$). Among the four types of stance markers coded, only epistemic hedges demonstrated a weak positive (but non-significant) association with writing quality. As for narratives, the diversity of organizational markers showed potential association with writing quality after controlling for length (partial $r = .19, p = .08$). Finally, a significant relationship was found between narrative quality and frequency as well as diversity of evaluative markers.

Predicting Writing Quality within Genre (RQ2b)

Informed by the correlation analysis, a series of hierarchical regression models was built to explore the predictive power of lexical, syntactic and discourse features in explaining quality of argumentative essays and written narratives, respectively (see Table 8 and Table 9).

We used a theory driven incremental approach of hierarchical multiple regression analyses to explore linguistic features that predict writing quality of argumentative essays. First, we entered participants' grade level as a control variable[5], which accounted for 13% of the variance in the writing quality score. Then, we entered text length, which, as expected, was a significant predictor, accounted for another 25% of the variance in quality scores. Starting from Model A3, we introduced the key predictor variables one at a time, starting with lexical complexity, a composite generated from three lexical measures. We found a significant main effect of lexical complexity on writing quality, with a one-point difference in the lexical complexity score associated with 0.63-point difference in essay quality, controlling for text length and grade level. Interestingly, the effect of grade on writing quality became non-significant after introducing lexical complexity into the model, which suggests that lexical complexity helps explain the cross-grade variation in writing quality. In Model A4, syntactic complexity

---

[5] We have also explored gender as a potential control variable, but no significant difference was found between the writing quality of male and female students. Moreover, the English proficiency level as rated by the Common European Framework was missing for 14 participants. Therefore, grade was entered into the model as the only demographic control variable.

explained an additional 4% of the variance in quality scores above and beyond the effect of lexical complexity. In Model A5, the effect of the diversity of organizational markers on writing quality approached significance. Though adding this predictor only increased the R^2 by 1%, it is indicative of a potential relationship worth exploring further in a larger sample. Model A5 was retained as the final model, explaining 52% of the variance in writing quality. All possible interactions were tested, but none was found to be statistically significant.

In a similar process, grade and text length were first entered into a baseline model to predict narrative writing quality. Participants' grade level accounted for 10% of the total variance in narrative writing quality and text length contributed to explain an additional 40% of the quality variance. It is noteworthy that the cross-grade difference in narrative writing quality became non-significant when adjusting for length. After Model N2 was established as the baseline model, we started exploring the additional impact contributed by the hypothesized predictors. In Model N3, lexical complexity explained an additional 1% of the variance in writing quality, but the effect of this variable on writing quality only approached significance. Similarly, adding syntactic complexity failed to contribute much to the prediction of writing quality, in contrast to what we observed in predictive models of argumentative writing quality. Diversity of organizational markers was entered as a predictor in Model N5, and showed an effect that approached significance. In Model N6, we were encouraged to see that the frequency of stance markers turned out to be a significant predictor of writing quality above and

beyond the effect of length. Moreover, the significant interaction between frequency of stance markers and length denoted that the effect of frequency of stance markers in written narratives was weaker for shorter texts than longer ones. However, it is also worth noticing that with the addition of the interaction term, the effect of organizational markers became non-significant. Ultimately, Model N7 was retained as the final model, explaining 61% of the variability in narrative writing quality.

In sum, the present study revealed no statistically significant differences in secondary school EFL learners' writing proficiency in narrative and argumentative essay as measured by holistic scoring of overall quality. However, argumentative writing demonstrated more sophisticated lexico-syntactic features, showing higher frequency of polysyllabic words, abstract nouns, words per clause and clauses per T-unit. Finally, distinct lexical, syntactic and discourse features were identified as predictors of overall writing quality for each genre. Controlling for text length, lexical complexity, syntactic complexity and diversity of discourse organizational markers were found to significantly and independently contribute to explain the variability in argumentative writing. For narrative writing quality only frequency of stance markers was found to be predictive after controlling for length.

Illustrating Domains of Individual Variability in Cross-genre Writing: A Few

Examples

This section displays two examples per genre that illustrate the higher and lower ends of the writing quality continuum for the sample of students examined in this study. It is worth noticing that all four examples show space for improvement in lexico-grammatical accuracy (e.g. subject-verb agreement, tense, problematic speech parts, etc.). A large amount of research has been conducted to specifically address these issues (Chandler, 2003; Polio, 1997), so the present study does not focus on mechanical or grammatical accuracy, but on lexico-syntactic and discourse resources.

Argumentative Essay 1 received a score of 4 out of 12, whereas Argumentative essay 2 received a score of 12, the highest score possible on the writing quality scale. The holistic rubric used in this study was calibrated to capture the variability within the sample; thus, despite the notable opportunities for improvement in Essay 2, it represented the best writing performance produced by this group of students. As shown in Essay 1 (see Figure 2), the student used a limited number of organizational markers (highlighted in bold). The few that appeared were mostly used to indicate inter-clausal relationships, rather than signaling overall text structure at the discourse level. In contrast, Essay 2 demonstrated the author's strategic use of organizational markers to construct argumentative structure at both local and global level. In addition to several transitional markers adopted to explicitly indicate the logical relations across

sentences (e.g. *because*, *that is*), the author also used a variety of frame markers (e.g. *on the one hand, the final thing*), code glosses (e.g. *let me show you an example, for example*) and conclusion markers (e.g. *in summary*), which successfully oriented the reader to the progression of arguments. As for stance markers, both essays displayed limited frequency and diversity of stance markers (underlined in the examples). Both essays contained several instances of markers of personal beliefs (e.g. *I think*), yet it is not clear that their function was to temper the argument or just served as discourse sequencers. In addition, it is worth noticing that Essay 2 demonstrated three instances of epistemic hedges, which might indicate the writer's reflective and cautious stance about the degree of certainty of the advanced assertions. In sum, unlike the salient discrepancy in organizational marker usage, the two essays did not differ considerably in the usage of stance markers, which further suggested this might be a more advanced skill in academic writing worth more explicit instruction.

Essay 1: Low-quality Writing (7ᵗʰ grade, male student)

<u>Lots of people think</u> success is a further question, it's hard to be. ***But*** <u>I think</u>, successful is just around us. Everyone wants to be success, ***but*** a few people are really hardworking for it. <u>I think</u> be success is very easy. A very warm wish, a hot drink in winter all can be little success. Doing good in the Final exam, get much money by the first month to work. They all are success. ***So***, do well in everything you will be successful.

Essay 2: High writing quality (9th grade, female student)

I think success comes from three things, hard work, being careful with details and courage. Hard work is the easiest, the most basic and the most important thing of the three. You can get abilities to succeed. It doesn't need talent. ***Because*** a person who doesn't work hard may not be successful ***though*** he is extremely talented. ***On the contrary***, a not so talented man will succeed ***because of*** his hard work. It's the thing that everybody can do. ***Next*** is being careful with details. ***On the one hand***, it means to do everything carefully. ***Let me show you an example***. In China, students need to take exams before they go to high schools or universities. This kind of exam is very important to students. ***But*** the exam is not very difficult. Students just need to be careful and make mistakes as few as possible. ***On the other hand***, being careful with details can bring you some chances. ***That is***, details in our life can show us the way to success. ***So***, we should pay attention to them and then get the chance. ***The final thing*** is courage. You need courage to do many things. ***For example***, when you face your chance, you need courage to decide to take it. ***Another example***, when you meet some difficulties, you need some courage to move on, to fight against it, to continue to walk on your way to success. ***In summary***, courage can help you never give up. Maybe, we are not as successful as those celebrities. And maybe, we don't get many achievements in our life. ***But*** I think it doesn't matter. ***Because*** I think everybody is the biggest success of his generation.

Figure 2. Students' argumentative writing: Low-quality and high-quality

examples

Figure 3 displayed examples of relatively low- and high-quality narrative

writing, with scores of 4 and 12 respectively. As can be observed, Narrative 1 used

several temporal markers (highlighted in bold) to explicitly mark the sequence of

events, a typical feature in most participants' written narratives. However, in

addition to temporal markers, Narrative 2 contained more diverse organizational

markers that indicated causal (e.g. *because*) and adversative (e.g. *but*)

relationships, which not only served as sequencers between events, but also added

an evaluative stance to the story. In addition to organizational markers, Narrative 2

also displayed abundant use of evaluative stance markers compared to Narrative 1,

including description of internal states (e.g. *afraid, nervous*), evaluative qualifiers

(e.g. *scary, suddenly*), exaggeration (e.g. *I almost couldn't move*), and objective

judgments (e.g. *my mum looked at me and gave me a smile*). In doing so, the

author vividly incorporated her evaluative stance to the narration and actively

engaged readers in the story.

Narrative 1: Low writing quality (7ᵗʰ grade, female student)

When I was grade five, I just study at school every week, just like other students.

One day, my class teacher told me: "There's a test about going to school choir, I

think you can go for the test, four p.m. in music room." I went to the room, and I

saw <u>a lot of</u> students in the room. Maybe I won't pass the test. When teacher called

my name. I stood up, and she took me to another room. She asked me <u>a lot of</u> questions, and let me sing a song. I remember that song called "mama who bore me". *After about thirty minutes*, we wrote down our phone number, and back to our classroom. *The next day morning*, my mum told me that she got a message. I pass the test! I was <u>excited</u>. *After school* the music teacher gave us <u>a lot of</u> paper of songs. I got a <u>good</u> chance and I chose it, do it until the end. *That's why* I achieved success.

Narrative 2: High writing quality (8th grade, male student)

I didn't like swimming <u>at all</u> when I was young. *But after that rainy night*, I think that I'm <u>interested</u> in it. I'm <u>afraid of</u> water when I was young. I think there might be something <u>scary</u> in the <u>deep</u> water. *Because* the water is <u>so</u> <u>wide</u> and <u>big</u>, I can't <u>imagine</u> what can I move in the <u>scary</u> place. *But* I change my opinion *after one night*. *That night*, I was taken by my mother. She wanted me to learn how to swim. It was <u>such bad</u> information for me. *But* I couldn't be against my mother, I followed her to the swimming pool near my home. We wore swimming suits and walked to the <u>scary</u> water <u>quickly</u>. My mother first dived into the water. *But* I just waited on the bank. My mother was a little <u>angry</u> and asked me to jump into the pool. I was <u>very</u> <u>nervous</u> *at that moment*. I didn't <u>want</u> to swim <u>at all</u> *because of* the <u>scary</u> water. *But* I <u>should</u> obeyed my mum. <u>I almost couldn't move then</u>. My mum saw the situation of me and said 'You just can't swim forever, you will never be great because of your heart.' I was <u>surprised</u> when I heard that. My heart?

What's wrong with my heart? I was thinking about the question and sitting on the bank. ***Then*** I saw a boy who is <u>very</u> <u>young</u>, he was swimming <u>difficultly</u> ***but*** <u>very hard</u>. I <u>admired</u> him and I <u>suddenly</u> got a point. Such a young boy could swim, ***but*** why I couldn't make it? I <u>felt so shame</u> and I thought about my mum's words again. ***After a while***, I <u>suddenly</u> stood up and rushed into the water <u>without hesitation</u>. When I got into the water, I felt <u>so weird</u>. Why should I afraid of it? ***Then*** I became to swim <u>slowly</u>. <u>My mum looked at me and gave me a smile</u>. I was <u>full of confidence</u> at that time and became more and more <u>faster</u>. I know that I should believe myself. The things that disturb you are not <u>so difficult</u>. I can make it!

Figure 3. Students' narrative writing: Low-quality and high-quality examples

Discussion

Chinese EFL Learners Writing two Genres: Similarities and Differences with Previous Research

Our results built on previous findings on native English speakers' cross-genre writing (Berman & Nir-Sagiv, 2007), showing that EFL learners' writing in two genres can also be captured through rigorous measures of lexical, syntactic and discourse features previously applied to native English speakers. This finding is encouraging in that it confirms EFL learners' cognitive awareness of distinct communicative purposes required by different genres and their ability to deploy

relevant linguistic knowledge available – distinctions between word classes (abstract or concrete), simple or complex sentences, impersonal or involved stance – that specifically serves such purposes.

Nevertheless, our results differ from previous research on native English speakers' cross-genre writing in several aspects. First, in contrast to previous research showing that native English speaking secondary school students wrote better narratives than arguments (Crowhurst, 1980, 1990; Engelhard Jr et al., 1992; Hall-Mills & Apel, 2013; Reed, 1992; Scott & Windsor, 2000), our results revealed no evidence of higher quality in narratives written by our a sample of EFL learners. The prevalent narrative-argumentative developmental model does not fully apply to the EFL sample in this study.

Moreover, a potential lexico-syntactic disassociation was found in EFL learners' writing in our sample. Previous research on native English speakers' language analysis commonly used a lexico-syntactic composite variable to capture the complexity of students' lexical and syntactic skills because of their closely related development (Berman, 2008; Uccelli et al., 2013). However, in our sample, we only found either weak or even negative correlation between the lexical and the two syntactic measures. Such dissociation might be aligned with Pearson, Fernandez, and Oller (1993), who found children's lexical skills correlated closely with morphosyntax score for native English speakers, but not for the bilingual learners.

Finally, different from Beers and Nagy (2009) who found that the quality of argumentative and narrative writing was correlated with different measures of syntactic complexity, our results revealed that words per clause had positive and stronger correlations with quality of both genres. One possible reason that clause per T-unit did not correlate with writing quality of EFL learners might be students' problematic use of subordinated clauses or grammatical inaccuracy. For example, many essays started with a sentence of the form "*I think … because …*" and repeated this formula a number of times. For narrative writing, a large number of students used repetitive sentences of the form "*and … and … and*" to connect a series of verbs without subjects. Such language style might account for why high number of clauses per T-unit did not yield high rating for quality.

Comparing results of the present study with those on native English speakers, this study also aims to make methodological contribution to the field of EFL writing research. We borrowed research tools from development linguistics – mostly from studies conducted with native English speakers – as complementary to other tools already used in the field of EFL writing. These measures of lexical, syntactical and discourse-features measures have shown to be appropriate and sensitive also to EFL writers' individual variability in this developmental period.

Pedagogical Implications and Questions for Future Research

Our study revealed several possible areas merit special instructional attention because our measures captured generally poor performance. First, results

showed that only limited number of academic vocabulary words (e.g. *substitute,*

underlie, inherent) (Coxhead, 2000) were used by most EFL learners. This might

indicate the necessity to incorporate more vocabulary of higher academic register

in EFL teaching materials and provide optimal conditions for students to meet and

learn academic vocabulary. Another area of improvement concerns EFL learners'

limited use of stance markers, especially the epistemic stance markers. This type

of stance marker, expressing the writer's caution about degree of possibility,

certainty or evidence for the individual's belief and the truth of a given state of

affairs, is considered both grammatically more complex and cognitively more

advanced, and therefore worth special instructional attention from EFL writing

instructors.

That EFL learners in the sample did not perform better on narrative writing

seemed to counter the data-driven conventional belief about the narrative –

argumentation developmental trajectory. We will surely need a longitudinal study

to confirm the developmental trajectory for EFL learners, but the present study

provides some preliminary understanding of EFL learners' writing performance

across genres. We could cautiously interpret this phenomenon as deriving from a

combination of linguistic and pragmatic accessibility. Given that a sociocultural

pragmatics-based view of language development understands language learning as

the result of individuals' socialization and enculturation histories (Halliday et al.,

2014; Ninio & Snow, 1996; Ochs, 1993; Snow & Uccelli, 2009; Uccelli et al.,

2014), the lower proficiency in written narratives might not be surprising in the

context of EFL instruction in China. Secondary EFL learners in China tend to

receive structured instruction and practice on argumentative writing in order to

prepare for high-stakes examinations (Xu & Wu, 2012; You, 2004a, 2004b).

Especially as students enter high school (9th grade and beyond), they have already

acquired solid understanding of the discourse expectations of this particular genre,

which enabled the majority of participants in the present sample to achieve a

medium score (the minimum score for 9th grade argumentative writing was 6 out

of 12). Among the entire sample, there were only two students who obtained a

score below 4 in overall argumentative quality. On the other hand, there is

minimum instruction and practice on EFL narrative writing in Chinese secondary

schools and limited opportunities are available for EFL learners to practice this

relatively informal genre inside or outside of the school due to the lack of natural

language environments that would demand such skills. This small sample of

students attended one of the most well-regarded English language institutes and

come from privileged and highly educated home environments, representing some

of the students with the best EFL training in the country. Thus, the relatively

underdeveloped narrative writing skill of this group of students suggests that these

students might benefit from more diverse instructional opportunities that address

multiple genres required to flexibly navigating different contexts, from rigorous

argumentative writing training to relatively informal personal story writing.

However, future research should be conducted to better understand the current

writing curriculum and instructional approaches in EFL writing classrooms in

China, so as to explore how students' learning experience might influence their performance on different genres.

Moreover, the identification of measures that capture relevant individual variability to predict writing quality enables the unpacking of holistic subjective impressions of quality in order to inform the design of more analytic writing tools that could guide assessment and instruction. The regression results shed light on particularly relevant areas to consider in pedagogical practices to teach writing in different genres, but future research needs to be conducted to examine whether explicit teaching of these linguistic features might yield higher writing quality. It is worth mentioning that this study did not pretend to advocate a prescriptive instructional design, solely focusing on introducing a list of linguistic forms to be memorized. An important goal is to make visible to EFL practitioners and researchers a repertoire of linguistic features that are closely associated with high writing proficiency as rated by experienced native-English-speaking teachers. Future research could explore how to best operationalize this research-based repertoire in classroom practices, so as to promote EFL learners' writing proficiency, especially encourage students to understand language as a functional solution to specific contexts of communication.

Limitations

While promising, our results must be viewed in light of some critical limitations. First, it is impossible to generalize our results beyond our sample of

EFL learners in secondary schools in China, especially considering this sample's relatively high social class background. Future research involving a larger number of students from more diverse social backgrounds would be necessary to confirm the generality of these findings. Moreover, the full capabilities of these participants in writing might not be displayed in these sample texts in part because they were written in a compressed time frame. Thus, the study did not attempt to assess students' overall writing ability; rather, it had the more modest goal of assessing the features of particular writing products. Longitudinal research that follows writers throughout secondary schools and collects their writing samples in multiple assessment contexts would be helpful to portray developmental trajectories and individual variability in acquiring writing proficiency across genres. Last but not the least, students' writing proficiency in their native language (L1) plays an important role in their EFL writing proficiency yet information on students' L1 was not available for this sample. Future research might consider collecting students' writing sample in L1 on similar topics and genres, and recruiting participants from diverse L1 background in order to investigate cross-linguistic relations in writing development across genres.

References

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing, 22*(2), 185-200.

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, 24*(2), 183-202.

Berman, R., Nayditz, R., & Ravid, D. (2011). Linguistic diagnostics of written texts in two school-age populations. *Written Language & Literacy, 14*(2), 161-187.

Berman, R. A. (2008). The psycholinguistics of developing text construction. *Journal of Child Language, 35*(04), 735-771.

Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes, 43*(2), 79-120.

Berman, R. A., & Slobin, D. I. (2013). *Relating events in narrative: A crosslinguistic developmental study*: Psychology Press.

Bolton, K. (2002). *Hong Kong English: autonomy and creativity* (Vol. 1): Hong Kong University Press.

Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics, 16*(03), 257-269.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*(3), 267-296.

Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English*, 67-87.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115-135.

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2011). *Predicting human scores of essay quality using computational indices of linguistic and textual features*. Paper presented at the Artificial Intelligence in Education.

Crowhurst, M. (1980). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education/Revue canadienne de l'éducation*, 6-13.

Crowhurst, M. (1990). The development of persuasive/argumentative writing. *Developing discourse practices in adolescence and adulthood*, 200-223.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 175-187.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. *Schooling and language minority students: A theoretical framework*, 3-49.

Engelhard Jr, G., Gordon, B., & Gabrielson, S. (1992). The influences of mode of discourse, experiential demand, and gender on the quality of student writing. *Research in the Teaching of English*, 315-336.

Ferris, D. R. (1994). Lexical and Syntactic Features of ESL Writing by Students at Different Levels of L2 Proficiency. *TESOL Quarterly, 28*(2), 414-420.

Frase, L. T., Faletti, J., Ginther, A., & Grant, L. (1999). *Computer analysis of the TOEFL Test of Written English*: Educational Testing Service.

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing, 23*(0), 1-16. doi: http://dx.doi.org/10.1016/j.jslw.2013.10.001

Gee, J. P. (2001). Reading as situated language: A sociocognitive perspective. *Journal of adolescent & adult Literacy*, 714-725.

Grabe, W. (2002). Narrative and expository macro-genres. *Genre in the classroom: Multiple perspectives*, 249-267.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*: Longman New York.

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.

Grant, L., & Ginther, A. (2000). Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences. *Journal of Second Language Writing, 9*(2), 123-145. doi: http://dx.doi.org/10.1016/S1060-3743(00)00019-9

Hall-Mills, S., & Apel, K. (2013). Narrative and Expository Writing of Adolescents With Language-Learning Disabilities A Pilot Study. *Communication Disorders Quarterly, 34*(3), 135-143.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). *An introduction to functional grammar*: Routledge.

Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in english*: Routledge.

Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*: cambridge university Press.

Heath, S. B. (2012). *Words at work and play: Three decades in family and community life*: Cambridge University Press.

Hickmann, M. (2003). *Children's discourse: person, space and time across languages* (Vol. 98): Cambridge University Press.

Hunt, K. W. (1983). Sentence combining and the teaching of writing. *The psychology of written language*, 99-125.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*: Bloomsbury Publishing.

Hyland, K. (2009). *Teaching and researching writing*: Pearson Education.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51-69.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing, 12*(4), 377-403. doi: http://dx.doi.org/10.1016/j.jslw.2003.09.001

Kachru, B., Kachru, Y., & Nelson, C. (2009). *The handbook of world Englishes* (Vol. 48): John Wiley & Sons.

Kang, J. Y. (2005). Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing, 14*(4), 259-279.

Kirkpatrick, A. (2007). *World Englishes Paperback with Audio CD: Implications for International Communication and English Language Teaching*: Cambridge University Press.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*(2), 148-161.

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes, 11*(4), 345-356.

Liardet, C. L. (2013). An exploration of Chinese EFL learner's deployment of grammatical metaphor: Learning to make academically valued meanings. *Journal of Second Language Writing, 22*(2), 161-178.

Liu, X. (2013). Evaluation in Chinese University EFL Students' English Argumentative Writing: An APPRAISAL Study. *Electronic Journal of Foreign Language Teaching, 10*(1), 40-53.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database. *Computational Linguistics, 26*(4), 657-657.

Martin, J. R. (1989). *Factual writing: Exploring and challenging social reality*: Oxford University Press, USA.

McKee, G., Malvern, D., & Richards, B. (2000). VOCD: Software for measuring vocabulary diversity through mathematical modeling. Version II for PC and Macintosh. Pittsburgh: Carnegie Mellon University, Child Language Data Exchange System (CHILDES) at h ttp://childes. psy. cmu. edu.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication*.

Miao, R., & Lei, X. (2008). Discourse Features of Argumentative Essays Written by Chinese EFL Students. *ITL International Journal of Applied Linguistics*(156), 179-200.

Ninio, A., & Snow, C. E. (1996). *Pragmatic development*: Westview Press.

Ochs, E. (1993). Constructing social identity: A language socialization perspective. *Research on language and Social Interaction, 26*(3), 287-306.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second-and fourth-grade students. *Reading and Writing, 22*(5), 545-565.

Ong, J. (2011). Investigating the use of cohesive devices by Chinese EFL learners. *The Asian EFL Journal Quarterly September 2011 Volume 13 Issue3, 13*, 42.

Paltridge, B. (2001). *Genre and the language learning classroom*: University of Michigan Press Ann Arbor.

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93-120.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*: Plenum Press New York.

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*(1), 101-143.

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System, 38*(3), 444-456.

Ravid, D. (2006). Semantic development in textual contexts during the school years: Noun Scale analyses. *Journal of Child Language, 33*(04), 791-821.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: a text-embedded cross-linguistic analysis. *First Language, 30*(1), 3-26.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language, 29*(02), 417-447.

Reed, W. M. (1992). The effects of computer-based writing tasks and mode of discourse on the performance and attitudes of writers of varying abilities. *Computers in human behavior, 8*(1), 97-119.

Schleppegrell, M. J. (2002). Linguistic features of the language of schooling. *Linguistics and Education, 12*(4), 431-459.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*: Routledge.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. *The Cambridge handbook of literacy*, 112-133.

Swales, J. (1990). *Genre analysis: English in academic and research settings*: Cambridge University Press.

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sánchez, E. (2014). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 1-33.

Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering Academic Language Organization and Stance in the Persuasive Writing of High School Students. *Written Communication, 30*(1), 36-62.

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment*: Cambridge University Press.

Wang, Z.-Q. (2014). The Application of Process Writing in Chinese EFL Classrooms in Higher Education. *International Journal of English Linguistics, 4*(3), p88.

Wimmer, G., Köhler, R., Grotjahn, R., & Altmann, G. (1994). Towards a theory of word length distribution*. *Journal of Quantitative Linguistics, 1*(1), 98-106.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College composition and communication*, 189-204.

Xu, Y., & Wu, Z. (2012). Test-taking strategies for a high-stakes writing test: An exploratory study of 12 Chinese EFL learners. *Assessing Writing, 17*(3), 174-190.

Yang, J. (2006). Learners and users of English in China. *English Today, 22*(02), 3-10.

Yang, L., & Gao, S. (2013). Beliefs and practices of Chinese university teachers in EFL writing instruction. *Language, Culture and Curriculum, 26*(2), 128-145.

You, X. (2004a). "The choice made from no choice": English writing instruction in a Chinese University. *Journal of Second Language Writing, 13*(2), 97-110.

You, X. (2004b). New directions in EFL writing: A report from China. *Journal of Second Language Writing, 13*(4), 253-256.

Zhao, C. G. (2010). *The role of voice in high-stakes second language writing assessment*. NEW YORK UNIVERSITY.

Tables and Figures

Table 1. Demographic characteristics of the sample (N = 100)

	Frequency
Gender	
Male	53
Female	47
Grade level	
Grade 6	23
Grade 7	24
Grade 8	27
Grade 9	13
Grade 10	6
Grade 11	7
English proficiency level (Common European Framework) [6]	
B1.1 - Intermediate 1	15
B1.2 - Intermediate 2	32
B2.1 – Upper intermediate 1	35
B2.2 – Upper Intermediate 2	4
Other	14
Mother's education level	
Graduate degree	33
College degree	48
No college degree	19
Father's education level [7]	
Graduate degree	39
College degree	39
No college degree	10

---

[6] According to the Common European Framework of Reference for Languages(Verhelst, Van Avermaet, Takala, Figueras, & North, 2009), intermediate language users (B1.1/B1.2) can understand the main points of clear standard input on familiar matters, deal with most situations while traveling, produce simple connected text on familiar topics, and describe experiences and events or briefly give reasons and explanations for opinions. Upper intermediate language users (B2.1/B2.2) can understand the main ideas of complex text on both concrete and abstract topics, interact with a degree of fluency and spontaneity, and produce clear, detailed text on a wide range of subjects and explain a view point on a topical issue.

[7] A number of participants did not report their parents' educational level, so their information was not presented in the table.

Table 2. Distribution of quality scores by grade level

	Argumentative			Narrative		
	Mean (SD)	SD	Min-Max	Mean (SD)	SD	Min-Max
Grade 6 (N = 23)	6.39	1.97	2 - 10	5.17	2.12	2 – 9
Grade 7 (N = 24)	6.79	2.06	4 - 10	7.54	2.87	2 – 12
Grade 8 (N = 27)	8.22	1.91	4 - 12	8.33	2.63	2 – 12
Middle School (N = 74)	7.18	2.11	2 - 12	7.09	2.87	2 – 12
Grade 9 (N = 13)	8.38	1.94	6 - 12	7.46	2.67	4 – 12
Grade 10 (N = 6)	8.50	1.76	6 - 11	8.67	3.08	4 – 12
Grade 11 (N = 7)	8.71	1.60	7 - 12	8.14	1.95	5 – 10
High School (N = 26)	8.50	1.75	6 - 12	7.92	2.54	4 – 12

Table 3. Descriptive statistics for text length, lexical and syntactic features varied by genre (N = 100)

Variable	Argumentative		Narrative		F(1, 99)	p
	Mean (SD)	Min-Max	Mean (SD)	Min-Max		
Length						
Number of clauses	32.81 (12.07)	6 - 68	36.7 (18.19)	6 - 82	9.51	0.003
Lexical features						
Word length	11.91 (7.41)	0 - 31	9.07 (6.90)	0 - 27	18.66	<0.001
Noun abstractness	19.28 (8.81)	4 - 44	12.53 (9.28)	0 - 39	63.85	<0.001
Academic words	1.36 (1.95)	0 - 13	2.76 (2.70)	0 - 11	21.90	<0.001
VocD[8]	67.26 (21.42)	22.12 – 117.34	57.27 (15.77)	21.06 – 98.41	26.75	<0.001
Syntactic features						
Clauses per T-unit	1.77 (0.33)	1.17 – 3.1	1.68 (0.33)	1 - 3	4.11	0.05
Words per Clause	6.02 (0.73)	4.55 – 8.40	5.44 (0.51)	4.13 – 7.23	58.32	<0.001

---

[8] Since CHILDES requires a minimum of 50 words token to calculate *vocd*, this measure was generated for 94 narratives and 96 argumentative essays.

Table 4. Descriptive statistics for raw frequency of genre-specific discourse markers in argumentative and narrative writing (N=100)

Variables	Mean	SD	Min - Max	# of Essays
Argumentative				
Organizational markers				
Frequency of organizational markers	7.07	4.10		
Diversity of organizational markers	2.42	1.02		
Frequency by type				
Frame markers	0.85	1.19	0-5	46
Transition markers	4.85	3.12	0-14	97
Code Glosses	0.99	0.98	0-4	62
Conclusion markers	0.38	0.51	0-2	37
Stance markers				
Frequency of stance markers	3.26	2.32		
Diversity of stance markers	1.72	0.85		
Frequency by type				
Deontic markers	1.15	1.64	0-9	52
Epistemic hedges	0.51	1.00	0-4	29
Personal beliefs	1.52	1.11	0-6	85
Epistemic boosters	0.08	0.34	0-2	6
Narrative				
Organizational markers				
Frequency of organizational markers	7.07	9.10		
Diversity of organizational markers	3.06	1.04		
Frequency by type				
Additive	1.00	1.16	0-5	55
Adversative	1.91	1.67	0-8	76
Causal	3.05	2.76	0-21	88
Temporal	3.54	2.70	0-12	87
Stance markers				
Frequency of stance markers	17.18	10.12		
Diversity of stance markers	3.54	0.81		
Frequency by type				
Lexical evaluation	8.56	5.74	0-25	98
Internal state	3.14	2.82	0-15	87
Rhetorical moves	3.48	2.63	0-13	89
Objective judgments	2.00	1.62	0-6	80

Table 5. Pairwise correlation of lexical, syntactic and discourse features with writing quality and partial correlation controlling for length: Argumentative writing

	1	2	3	4	5	6	7	8	9	10	11	12	Partial r with writing quality
1.quality	-												-
2.length	.63***	-											-
Lexical features													
3.word length	.58***	.66***	-										.27**
4.noun abstractness	.58***	.61***	.60***	-									.31**
5.lexical diversity	.48***	.34***	.42***	.40*	-								.37***
Syntactic features													
6.words per clause	.32**	-.01	.27**	.36***	.25*	-							-
7.clauses per T-unit	-.13	-.18~	-.24*	-.17~	-.05	-.28	-						-
Discourse features													
8.organizational marker frequency	.47***	.58***	.41***	.47***	.09	.19~	-.19	-					.17~
9.organizational marker diversity	.46***	.42***	.39***	.35***	.19	.22*	-.10	.63***	-				.27**
10.deontic markers	.05	.19~	.10	.01	-.02	-.01	-.05	.22*	.24*	-			-.10
11.epistemic hedges	.16~	.21*	.13	.18~	.05	.04	.19	.12	.05	.23*	-		.04
12.personal beliefs	-.11	.12	.01	-.03	.03	-.12	-.08	.28**	.17	-.03	-.10	-	-.24
13.epistemic boosters	-.03	.08	.13	.04	.04	-.10	.07	.02	.08	-.01	-.13	.16	-.11

Table 6. Pairwise correlation of lexical, syntactic and discourse features with writing quality and partial correlation controlling for length: Narrative Writing

	1	2	3	4	5	6	7	8	9	10	Partial r with writing quality
1.quality	-										-
2.length	.73***	-									-
Lexical features											
3.word length	.62***	.74***	-								.17
4.noun abstractness	.59***	.70***	.82***	-							.16
5.lexical diversity	.44***	.45***	.51***	48***	-						.20*
Syntactic features											
6.words per clause	.19~	.05	.28**	.28**	.10	-					-
7.clauses per T-unit	.09	.25*	.22*	.23*	.27**	-.08	-				-
Discourse features											
8.organizational marker frequency	.60***	.73***	.53***	.47***	.23*	.30**	.12	-			.13
9.organizational marker diversity	.55***	.62***	.44***	.45***	.20~	.21*	.09	.67***	-		.19~
10.stance marker frequency	.70***	.70***	.56***	.50***	.27*	.21*	.02	.58***	.52***	-	.39***
11.stance marker diversity	.48***	.56***	.34***	.38***	.19~	.04	.17~	.53***	.50***	.53***	.24*

Table 7: Results from principal component analysis for lexical complexity

Variable	Eigenvalue	Explained Variation	Loading
Lexical complexity intricacy	2.06	.0.69	
Word length			.62
Noun abstractness			.60
Lexical diversity			.51

Table 8. Regression models testing the effect of lexical, syntactic and discourse features on argumentative writing quality: Controlling for length, β (SE)

	Model A1	Model A2	Model A3	Model A4	Model A5
Parameter estimate					
Grade	0.56***	0.37***	0.16	0.11	0.13
	(0.13)	(0.11)	(0.12)	(0.11)	(0.11)
Length		0.10***	0.04*	0.06***	0.05**
		(0.01)	(0.02)	(0.02)	(0.02)
Lexical complexity			0.63***	0.42*	0.39*
			(0.16)	(0.17)	(0.17)
Syntactic complexity				0.66**	0.57*
				(0.24)	(0.24)
Organizational markers					0.26~
(Diversity)					(0.16)
Goodness of fit: R^2	0.13	0.38	0.47	0.51	0.52
Change of R^2		0.25***	0.09***	0.04**	0.01~

~p < 0.1; *p<.05; ** p<.01; ***p<.001

Table 9. Regression models testing the effect of lexical, syntactic and discourse features on narrative writing quality: Controlling for length, β (SE)

	Model N1	Model N2	Model N3	Model N4	Model N5	Model N6	Model N7
Parameter estimate							
Grade	0.61***	0.21	0.07	0.07	0.08	0.08	0.05
	(0.18)	(0.14)	(0.15)	(0.15)	(0.15)	(0.14)	(0.14)
Length		0.11***	0.08***	0.09***	0.07***	0.04*	0.09**
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Lexical complexity			0.43~	0.30	0.35	0.37~	0.41~
			(0.23)	(0.24)	(0.24)	(0.22)	(0.22)
Syntactic complexity				0.61	0.41	0.10	0.10
				(0.41)	(0.42)	(0.41)	(0.40)
Organizational markers (Diversity)					0.46~	0.38~	0.20
					(0.27)	(0.25)	(0.26)
Stance markers (Frequency)						0.09***	0.20***
						(0.03)	(0.05)
Interaction between evaluative markers and length							-0.01*
							(0.01)
Goodness of fit: R^2	0.10	0.50	0.51	0.52	0.54	0.60	0.61
Change of R^2		0.41***	0.01~	0.01	0.02~	0.06***	0.02*

~$p < 0.1$; *$p<.05$; **$p<.01$; ***$p<.001$

Appendices

Appendix A: Writing Prompts

Argumentative Writing Prompt:

Some people believe that success in life comes from risks or chances. Others believe that success results from careful planning. In your opinion, what does success come from? Use specific reasons and examples to explain your position.

Narrative Writing Prompt:

Write a personal story about a time when you achieved success. Please include detailed memories about that experience, including the context, your actions, feelings, etc.

Appendix B: Research Measures

Measure	Description	Example
Cross-genre Length measure		
Number of clauses	A clause is defined as "a unit that contains a unified predicate, … [i.e.,] a predicate that expresses a single situation (activity, event, state). Predicates include finite and nonfinite verbs, as well as predicative adjectives (Berman & Slobin, 1994:660)"	*I remember the day [c] when we had a soccer game with a team [c] who has won the championship for several times [c].*
Cross-genre Lexical measures		
Word Length	This measure has been validated in corpus-linguistics research (e.g., Wimmer, Kohler, Grotjahn, & Altmann, 1994) to assess the complexity of words' orthography. In English, polysyllabic words are considered less frequent and structurally more complex (e.g. *perspective*, *derivational*) than words with fewer syllables.	*perspective, derivational*
Noun Abstractness	Adapted from Ravid (2006) and Berman & Nir-Sagiv (2007), nouns were categorized into a four-place ranking: 1) concrete objects or proper names; 2) categorical and generic nouns; 3) abstract but high-frequency nouns; 4) abstract and low-frequency nouns or derivational nouns	*1) bike, Mary* *2) doctor, people* *3) answer, exam* *4) authority, communication*
Academic Vocabulary	Words that appear on the Academic Vocabulary List (Coxhead, 2000), which was compiled from a corpus of 3.5 million running words of written academic text by examining the range and frequency of words outside the first 2,000 most frequently occurring words of English.	*register, accumulate*
Lexical Diversity	*vocd*, a measure generated by CHILDES, to assess the diversity of active vocabulary deployed by writers	

	(MacWhinney, 2000; Malvern & Richards, 2000)	
Cross-genre Syntactic measures		
Words per clause	Measures writers' skill to convey information in a more concise manner by combining information from multiple clauses into a single clause.	The sentence "*His decision of resignation surprised his colleagues [c].*" has 7 words per clause.
Clauses per T-unit	Measures frequency of embedded sentence structure, such as subordinate and relative clauses. *T-units are defined as "thematic units of complete and autonomous meaning, corresponding to a main clause plus all the subordinate clauses embedded in it" (Hunt, 1983)*	The sentence "*I believe that [c] success comes from hard work [c] because a person who does not work hard [c] may not be successful [c] even if he is very talented* [c]" has five clauses per T-unit.
Genre-specific Discourse measures **- Narrative**		
Organizational markers	Following Halliday and Hasan (1994), there are four types of connectives that denote inter-clausal relationships: 1) additive; 2) adversative; 3) causal; 4) temporal *coordinative conjunction "and" will not be coded*	*1) in addition, moreover* *2) however, although* *3) because, as a result* *4) earlier, after*
Stance markers	Adapted from Peterson and McCabe's (1983) 1. Repetition; 2. Compulsion words; 3. Similes and metaphors; 4. Gratuitous terms, such as very, just, really, as intensifiers 5. Attention-getter; 6. Evaluative adjectives: fun, ugly, funny, excited, surprising, important, etc. 7. Evaluative adverbs: finally, accidentally; 8. Negatives 9. Expressions of intentions, purposes, desires or hopes	1. "I was *very very* happy". 2. "We *had to* come in then". 3. "His eyes got *as big as tomatoes*". 4. "The test was *really* hard". 5. "*You know what*?" 6. "That was an very *important* experience". 7. "*Finally*, we won". 8. "He *didn't* hit me". 9. "I *wanted* to ride the horse then". 10. "*Mom thought* I had the chicken

	10. Articulations of hypotheses, guesses, inferences, and predictions 11. Causal explanations 12. Objective judgment: the narrator uses other people to evaluate the narrated event 13. Descriptions of internal emotional states of either the narrator or some other participants in the event	pox". 11. "I won the contest *because of* my hard work". 12. "*My brother liked* my snowman much better than he liked my sisters". 13. "I was really mad at her".
Genre-specific Discourse measures ***- Argumentative***		
Organizational markers	Markers that explicitly signal the organization of argumentative text structure (Hyland, 2005) 1) *Frame markers* that indicates the sequence of arguments or counter-arguments; 2) *Code Glosses* that introduce an example or paraphrase; 3) *Transition markers* signaling additive, adversative and causal relations between clauses and paragraphs. 4) *Conclusion markers* that explicitly state the author's summary or conclusion of the essay.	1. Frame: *first of all, on the other hand*) 2. Code glosses: *for example, in other words* 3. Transition: *moreover, even though, because* 4. Conclusion: *In conclusion, all in all*
Stance markers	Adapted from Berman (2004) and Reilly et al. (2002): 1. Deontic markers: indicating the writer's absolute stance on a viewpoint 2. Epistemic markers: 1) Epistemic hedges: expressing degree of uncertainty; 2) Epistemic boosters: emphasizing the writers' commitment to the truth of an assertion; 3) Personal beliefs: acknowledging that assertions are the result of one's or others' personal beliefs	1. Deontic: *you should do…; it is wrong…* 2. 1) Epistemic hedges: *it might be…; it is possible* 2) Epistemic boosters: *it is absolutely true;* 3) Personal beliefs: *I believe; some people think…*