Same language, different functions: A cross-genre analysis of Chinese EFL learners’ writing performance

Wenjuan Qin\textsuperscript{a,}\textsuperscript{*}, Paola Uccelli\textsuperscript{b}

\textsuperscript{a}Graduate School of Education, Harvard University, Larsen Hall 303, 14 Appian Way, Cambridge, MA 02138, United States
\textsuperscript{b}Graduate School of Education, Harvard University, Larsen Hall 320, 14 Appian Way, Cambridge, MA 02138, United States

\textbf{ARTICLE INFO}

Article history:
Received 20 May 2015
Received in revised form 24 May 2016
Accepted 1 June 2016
Available online xxx

Keywords:
Chinese EFL learners
EFL writing
School genre
Narrative writing
Argumentative writing
Sociocultural pragmatics

\textbf{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 performance in two genres, argumentative essays and narratives. Cross-genre research on native language writing has documented that, at the macro-text level, secondary-school students’ narratives are typically of higher quality than their written essays; while at the lexicosyntactic level, essays display higher complexity than narratives. To investigate cross-genre differences in EFL learners, 200 English written texts (100 essays; 100 narratives) were collected from 100 EFL Chinese secondary school learners and scored for quality, lexicosyntactic, and genrespecific discourse features. Unlike prior research on native language writing, no significant differences in quality ratings were found across the two genres. However, in line with prior research, results revealed that argumentative essays displayed a higher lexicosyntactic complexity. Regression analyses identified distinct sets of predictors of writing quality ratings for each genre. Controlling for length, lexicosyntactic complexity and diversity of organizational markers were identified as predictors of argumentative essay 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.

Copyright © 2016 Elsevier Inc. All rights reserved.
considerable EFL writing research has focused on the post-secondary years, with only minimal research devoted to understanding secondary school students’ performances.

The current study investigated secondary school Chinese EFL students’ writing performance in two different genres: narrative and argumentation. Writing performance is defined by two major constructs: 1) holistic writing quality ratings evaluated by experienced EFL teachers; 2) text-based lexical, syntactic, and discourse features. By comparing writing performance across narrative and argumentative genres, this study seeks to reveal differences in performance that result from the genre-specific language demands faced by these learners. Two main questions drove this study. First, do EFL learners’ overall writing quality ratings and/or lexi-co-syntactic features vary by genre? Second, which lexi-co-syntactic and discourse features predict overall writing quality ratings within each genre?

The ultimate goal of this line of research is to generate findings that would inform the design of pedagogical approaches that will be especially attuned to the needs of Chinese EFL students as they learn to participate in two prevalent discourse genres: narrative production and evidence-based argumentative writing.

1. Theoretical framework: a sociocultural pragmatics-based view

A sociocultural pragmatics-based view of language development understands 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., 2015). This view of language entails that being a skilled language user in some social contexts does not guarantee adequate performance in other social contexts. From a pragmatics-based developmental perspective, learners 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 as a challenging task for many colloquially fluent monolingual or bilingual students with scarce opportunities to be socialized into more academic discourse (Schleppegrell, 2004; Snow & Uccelli, 2009; Uccelli et al., 2015). In the context of his research with 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, 1991).

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 (Bereiter & Scardamalia, 2013; Flower & Hayes, 1981; Gee, 2001; Graham & Perin, 2007). The performance on 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). Certain types of texts are believed to be linguistically and cognitively more demanding than others and the frequency of exposure to different communicative contexts differs remarkably across individuals. Consequently, we anticipate that EFL learners might be able to excel at writing in one genre but not in another.

The present study was conducted to advance our current understanding of EFL writing by addressing two research gaps. First, this study focuses on secondary school EFL learners, an understudied population in L2 writing research. Second, informed by a pragmatics-based view of language in which writers are understood as not necessarily equally skilled across all uses of language, this study compares EFL students’ writing performance across genres. Adolescent second language learners represent the largest and fastest growing population in international settings, yet research that documents their needs in writing development to inform instruction is scarce (Leki, Cumming, & Silva, 2008; Matsuda & De Pew, 2002; Ortmeier-Hooper & Enright, 2011). Though recent studies have investigated adolescent L2 writing development in some English-dominant contexts (Enright & Gilliland, 2011; Fránquiz & Salinas, 2011), in EFL settings, adolescents continue to be an understudied population. In China, more specifically, the majority of recent empirical studies focused on undergraduate or graduate learners (Li & Wharton, 2012; Liardet, 2013; Liu, 2013; Miao & Lei, 2008; Qin & Karabacak, 2010). Additionally, many genre-based studies using L2 corpus focused on advanced academic discourse such as academic research articles (Marco, 2000; Swales et al., 1998) or writing across academic disciplines (Hardy & Rümér, 2013). Yet, less is understood about EFL adolescent learners’ writing and how their skills might vary across argumentative and narrative genres.

2. Writing performance across two genres

Narrative and argumentative texts are two distinct discourse genres 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 (e.g., narratives); (2) factual genres (e.g., procedures); and (3) analytic genres (e.g., argumentative essays). Recent empirical data on native language writing development support this developmental sequence, showing that while written narrative structures tend to be well mastered by around age ten, argumentative writing constitutes a later developmental accomplishment (Berman & Nir-Sagiv, 2007). A number of empirical studies examining writing produced by middle and high school students further suggest that adolescents are typically able to produce higher-quality narratives than argumentative essays (Crowhurst, 1980, 1990; Englehard, Gordon, & Gabrielson, 1992; Hall-Mills & Apel, 2013; Reed, 1992; Scott & Windsor,
2000). However, all of these studies investigated native English writing. To our knowledge, to date no similar contrastive studies have been conducted for EFL adolescent learners.

This documented developmental sequence can be explained in relation to the different linguistic and cognitive demands posed by each genre. 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). Argumentative essays are, instead, topic-oriented, require the writer to impose a logical structure to interrelate ideas in a coherent manner, and to organize claims and arguments in a stepwise 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. At the syntactic level, compared to written narratives, argumentative texts produced by native language writers tend to display more complex structures (Beers & Nagy, 2009; Berman & Nir-Sagiv, 2007; Ravid and Berman, 2010).

Though few L2 writing studies focused on writing across genres, longitudinal studies have shed light on the development of lexical and syntactic complexity in L2 writing over time in a single genre. In studying a common corpus of L2 written descriptions produced by students in a semester-long ESL writing course at a U.S. university, research has shown that growth in L2 syntactic complexity occurred as a function of time spent learning English (Crossley & McNamara, 2014). Using the same corpus, Bulté and Housen (2014) also found increases in measures of syntactic complexity at all levels of structure (phrasal, clausal, and sentential), but no significant change in measures of lexical complexity (e.g., lexical diversity/richness). Interestingly, in this study, the only syntactic measure that did not demonstrate significant change over time was frequency of subordinate clauses. This finding might echo a critical evaluation on this measure done by Biber, Gray, and Poonpon (2011), showing that most clausal subordination measures are actually more effective in discriminating language proficiency in spoken rather than in writing registers.

In light of the cross-genre lexical and syntactic differences documented in research on native language writers, and a lack of knowledge in cross-genre performance among EFL writers, this study will first explore if Chinese EFL learners’ writings in two genres achieve different levels of quality, as rated by experienced EFL teachers. Additionally, guided by the inconclusive argument regarding the best-fitted lexical-syntactic measures within and across genres, various lexical and syntactic measures – commonly used in both native language and L2 writing research – will be used to examine how lexico-syntactic features vary by genre.

3. Predictors of writing quality ratings within each genre

Unpacking the predictive relationship between certain text-based features and overall writing quality rated by teachers, research on both native English speakers and EFL learners has identified lexical, syntactic, and discourse features that are significantly associated with overall writing quality ratings. However, as reviewed below, this line of research has mostly been conducted with post-secondary students, with only a few studies focused on secondary school students. Besides, most of them examine only argumentative essays with minimal attention to narratives.

3.1. Research on native English writing

Research on native English speakers, at both post-secondary and secondary level, has been inconclusive on which lexico-syntactic or discourse factors are predictive of overall quality ratings of argumentative and narrative writing. For instance, using an automated text analysis tool (Coh-Metrix), 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. However, a later study conducted by the same group of researchers on post-secondary freshmen’s writing revealed that essays rated as higher-quality were characterized by not only lexical diversity, but also more cohesive textual information (Crossley, Roscoe, & McNamara, 2011). At the secondary school level, Connor (1990) analyzed argumentative essays written by high school students and found that higher rated essays used sophisticated syntactic factors, including nominalizations, propositions, passives and specific conjuncts. On the other hand, Uccelli, Dobbs, and Scott (2013) and Dobbs (2014) identified 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 (e.g., Certainly, Possibly, It is unlikely that . . .).

Fewer studies have explored the relation between linguistic features and writing quality ratings of narrative texts, and most of them have been conducted with younger students. For instance, studies with 2nd and 4th grade students have found that length and diversity of vocabulary predicted narrative writing quality ratings (Olinghouse & Learid, 2009). Additionally, in 9th graders’ writing, frequency of cohesive indices was found to explain a significant amount of variations in narrative writing quality ratings (Cameron et al., 1995).
3.2. Research on EFL writing

The majority of studies on predictors of EFL writing quality ratings focused on argumentative writing at the undergraduate or graduate level. Using the Test of Written English (TWE) corpus, researchers found that highly rated EFL essays were usually longer, with longer average word length, and higher frequencies of certain lexical and grammatical categories (e.g., nouns, adjectives, conjuncts) (Ferris, 1994; Frase, Faletti, Ginther, & Grant, 1999; Grant & Ginther, 2000). Interestingly, findings by Jarvis, Grant, Bikowski, and Ferris (2003) and Friginal, Li, and Weigle (2014) suggest 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 (2013) operationalized “voice” as a macro level construct characterized by hedges and boosters that convey the author’s attitude and credibility, and directives or questions that establish engagement with the reader. In her study, a strong positive correlation was detected between EFL writers’ authorial voice and writing quality rating. Thus, for university-level EFL argumentative writing, raters’ perception of quality is associated with both micro-level lexical features and macro-level discourse features.

EFL narrative writing has received less attention than argumentative writing. To our knowledge, only two studies have focused on this question. Narrative writing quality ratings have been found to be related to length and syntactic complexity in Japanese post-secondary students’ writing (Ishikawa, 1995). In a second study, Hungary EFL learners, when compared to native English speakers’ higher-rated narrative writing, tended to use less diverse and sophisticated vocabulary (Kormos, 2011). Thus, the rating of university-level EFL narrative writing quality, so far, has only been found to relate to micro-level lexico-syntactic features.

3.3. Contrasting predictive relations across genres

To our knowledge, there is only one study contrasting predictive relation between linguistic features and writing quality ratings across genres, conducted by Beers and Nagy (2009). They found syntactic complexity – as measured by words per clause – to be positively correlated with writing quality ratings for argumentative essays but not for narratives. In other words, the syntactic complexity that contributed to argumentative quality was clause-internal, as demonstrated in the use of clause-lengthening prepositional phrases, attributive adjectives, gerund phrases, and infinitive phrases. On the other hand, Beers and Nagy (2009) found that clauses per T-unit was positively correlated with quality ratings for narratives, but negatively correlated with quality ratings for essays. This finding can be interpreted as aligned with the claim that increased subordination is a more effective discriminator of conversational proficiency, which is arguably closer to the narrative genre, rather than to academic writing (Biber et al., 2011). In light of the interesting contrast found in Beers and Nagy (2009), the present study will further unpack the relation between these two syntactic measures and overall writing quality ratings within and across genres.

In summary, research on both native English speakers and EFL learners has been inconclusive about which lexico-syntactic and discourse features are associated with writing quality ratings in each genre. Moreover, limited research is available to contrast such predictive relations across genres, especially in EFL adolescent learners’ writing.

4. Research questions

In order to better understand Chinese secondary school EFL learners’ writing performance across genres and unpack the relations between certain linguistic features and writing quality ratings within each genre, the present study seeks to answer the following two sets of research questions:

1. Comparing writing performance across genres:
   (a) Do overall quality ratings of Chinese secondary school EFL learners’ written personal narratives and argumentative essays differ by genre? (b) Do incidences of key lexico-syntactic features vary by genre in these Chinese secondary school EFL learners’ written texts?

2. Predicting overall writing quality ratings within each genre:
   (a) What genre-specific discourse features characterize Chinese 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 ratings in each genre?

5. Methods

5.1. Participants

The sample consisted of 100 secondary school EFL learners, whose ages ranged from 11 to 17 years (see Table 1). Students’ grade-level ranged from 6th to 11th grade. The sample was relatively balanced by gender, with 53 boys and 47 girls. All participants received comparable standard instruction in the same language institute in east China. Besides the regular English classes they took at local middle/high schools, these participants attended an extra language program outside of school in the same language institute, mostly for the purpose of test preparation (e.g., TOEFL, IELTS) in order to apply for undergraduate study in English-dominant countries, such as the U.S. and U.K. As described in the curriculum and by teachers
at the language institute, most writing practices focused on argumentative genre, which is not surprising given that this is the most commonly tested genre in high-stake examinations.

5.2. Data collection

Compositions written by Chinese 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 min (40 min for each text plus a 10-min break). Both writing prompts were on a similar topic to optimize comparison across genres (see Appendix A for the full writing prompt).

5.3. 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 ratings 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:

5.3.1. Writing quality

Writing quality was rated using two genre-specific six-point-scale holistic scoring rubrics (The National Assessment of Educational Progress, 2011). 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. In addition, scorers were made aware of the different demands expected in each genre and were asked to assess “whether the text elicited appropriate information and language style for the particular genre.” Scorers were also provided with a packet of prototypical examples, selected by an experienced native-English-speaking scorer and the authors, which represented different levels of writing quality in both genres. The writing quality ratings are comparable across genres in that, for instance, a 6-point argumentative essay and a 6-point written narrative both represent the best possible writing performance in the corresponding genre.

Two native English-speaking, experienced teachers who were blind to the research questions scored each text. All texts were double-scored and a satisfactory level of inter-rater reliability was achieved (κ = 0.83 for argumentative; κ = 0.80 for narrative). Following standard SAT scoring practices, when two scorers reached either exact or adjacent agreement, their ratings were summed to form the final scores; when the two scorers were more than one point apart, a third expert scorer intervened to settle the disagreement, and the expert’s score was doubled as the final score.

5.3.2. Length

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).

Table 1

Demographic characteristics (n = 100).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>53</td>
</tr>
<tr>
<td>Female</td>
<td>47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade level</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 6</td>
<td>23</td>
</tr>
<tr>
<td>Grade 7</td>
<td>24</td>
</tr>
<tr>
<td>Grade 8</td>
<td>27</td>
</tr>
<tr>
<td>Grade 9+</td>
<td>26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English proficiency level (Common European Framework)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1.1—Intermediate 1</td>
<td>15</td>
</tr>
<tr>
<td>B1.2—Intermediate 2</td>
<td>32</td>
</tr>
<tr>
<td>B2.1—Upper intermediate 1</td>
<td>35</td>
</tr>
<tr>
<td>B2.2—Upper Intermediate 2</td>
<td>4</td>
</tr>
<tr>
<td>Not available</td>
<td>14</td>
</tr>
</tbody>
</table>
5.3.3. Lexical measures
This set of measures captured word-level characteristics for both narrative and argumentative texts, measuring the frequency of:

a) Long words: or polysyllabic words; specifically, words with three or more syllables (e.g., perspective, transportation) (Wimmer, Köhler, Crotjahn, & Altmann, 1994).

b) Abstract nouns: using a four-level semantic abstractness scale (Berman & Nir-Sagiv, 2007; Ravid, 2006), nouns used in students’ writing were categorized into: concrete objects or proper names (e.g., bike, Mary); categorical and generic nouns (e.g., people, doctor); abstract but high-frequency nouns (e.g., idea, exam); abstract and low-frequency nouns or derivational nouns (e.g., perspective, communication).

c) Academic vocabulary: words that appear on the Academic Vocabulary List (Coxhead, 2000).

d) 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 frequency of long words, academic vocabulary, and VocD were automatically generated using the CHILDES programs (MacWhinney, 2000). Two researchers doubly coded all narrative and argumentative texts for frequency of abstract nouns, with an inter-rater reliability of $\kappa = 0.92$.

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

a) 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.

b)Clauses 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, 1965). Using multiple clauses per T-unit is one way of expressing complex relations between ideas and the number of clauses per T-unit is found to increase with students’ age (Beers & Nagy, 2009).

Two researchers double-coded 20% of transcripts. Inter-rater reliability was estimated by applying Cohen’s kappa statistics, with $\kappa = 0.89$ for T-unit coding and $\kappa = 0.93$ for clause coding.

5.3.5. Genre-specific discourse measures
While the same lexico-syntactic features can be investigated across genres, 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) evaluative stance:

**Organizational markers in argumentative texts:** Following research on metadiscourse analysis (Dobbs, 2014; Hyland, 2005; Uccelli et al., 2013), markers used to explicitly signal the organization of argumentative text structure were identified and classified into four subcategories:

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

b) Code glosses introduce an example or paraphrase (e.g., for example, in other words);

c) 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;

d) 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, Reilly, Baruch, Jisa, and Berman (2002), Hyland (2005), and Uccelli et al. (2013), stance markers in argumentative essays were identified and coded as:

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

b) 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 relation 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) relations. The coordinating conjunction and was not included in this coding.

Stance markers in narrative texts: Following Peterson and McCabe (1983) tools for identifying evaluative markers in narratives, narrative texts were coded for four types of evaluative stance markers used by the writer to offer an implicit or explicit subjective interpretation of the narrated events. Stance markers in narrative texts were subsequently coded for: 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).

Two researchers doubly coded 20% of the data for discourse markers and achieved adequate reliability (κ = 0.88 for argumentative organizational markers; κ = 0.91 for argumentative stance markers; κ = 0.92 for narrative organizational markers; κ = 0.87 for narrative stance markers).

5.4. Analytic plan

To address the first set of research questions, descriptive statistics were generated for writing quality ratings by genre. After examining the score distribution, we conducted paired sample t-tests with writing quality scores as the dependent variable. Then descriptive statistics were generated for lexical and syntactic measures by genre. After examining the distribution of variables and checking the important assumptions, paired sample t-tests were conducted to investigate whether the incidences 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. Principal Component Analysis (PCA) was performed to reduce collinearity among certain linguistic variables. Then, correlational analysis informed the construction of a series of theory-based hierarchical regression models, independently for narrative and argumentative genres.

6. Results

6.1. RQ1: writing quality and lexico-syntactic variations across genres

6.1.1. Comparing writing quality across genres

Argumentative essays and narratives were double-scored for overall writing quality (with a possible range of 2–12). Narrative mean quality score was 7.31 (SD = 2.80). Argumentative essays had a slightly higher mean quality score of 7.53 (SD = 2.11). A paired sample t-test showed that the difference between mean quality ratings of argumentative and narrative writing was not statistically significant (t(99) = 0.89, p = 0.37). Furthermore, a closer examination of the data revealed enormous individual variability within and across genres. As illustrated in Fig. 1, 80% of argumentative essays 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 a high number of narratives displaying either a high score (11–12) or a score at the lower-end (2–5) on the scale.

6.1.2. Comparing lexical-syntactic features across genres

Table 2 summarizes descriptive statistics for text length and lexico-syntactic features in sampled argumentative and narrative writing. Given that we are investigating eight measures and therefore performing eight tests on the same dataset simultaneously, we employ the Bonferroni correction to avoid spurious positives, following (Lu, 2010). This sets the alpha value for each comparison to 0.05/8, or 0.006. Paired t-tests demonstrated that argumentative and narrative genres varied

![Fig. 1. Distribution of writing quality scores by genre.](image-url)
Finally, additive markers (e.g., students used causal markers (e.g., events in a temporal order, and thus the organizational markers in narrative discourse mostly denote the transition at micro essays.

Finally, epistemic boosters (e.g., the data revealed overabundant usage of this type of stance markers in repetitive pattern, such as using type of stance markers were markers of personal beliefs (e.g., decided whether this is a form of epistemic stance or was only used as discourse-sequencer. 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 narratives.

Nouns (e.g., I think), which appeared in 85 essays. However, a closer look at the data revealed overabundant usage of this type of stance markers in repetitive pattern, such as using “I think . . . ” every time a new argument was introduced. Consistently with prior research (Reilly et al., 2002), it was not always possible to decide whether this is a form of epistemic stance or was only used as discourse-sequencer. 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 narratives.

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. As expected, temporal markers (e.g., earlier, finally) displayed the highest frequency. Comparably, 88 students used causal markers (e.g., because, the reason that . . . ) followed by adversative markers (e.g., though, however). Finally, 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 stance markers. All four types of stance markers were identified in 80 out of the 100 narratives. Evaluative qualifiers, that is, the use of adjectives (e.g., unforgettable), and adverbs (e.g., actively), and intensifiers (e.g., really, very much), were 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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Argumentative</th>
<th>Narrative</th>
<th>(t(99))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length</strong></td>
<td><strong>M (SD)</strong></td>
<td><strong>Min–Max</strong></td>
<td><strong>M (SD)</strong></td>
</tr>
<tr>
<td>Number of clauses</td>
<td>32.81 (12.07)</td>
<td>6–68</td>
<td>36.7 (18.19)</td>
</tr>
<tr>
<td>Number of words</td>
<td>197.61 (76.57)</td>
<td>38–409</td>
<td>200.23 (101.83)</td>
</tr>
<tr>
<td><strong>Lexical features</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freq of Long words</td>
<td>11.91 (7.41)</td>
<td>0–31</td>
<td>9.07 (6.90)</td>
</tr>
<tr>
<td>Freq of Abstract nouns</td>
<td>19.28 (8.81)</td>
<td>4–44</td>
<td>12.53 (9.28)</td>
</tr>
<tr>
<td>Freq of Academic words</td>
<td>1.36 (1.95)</td>
<td>0–13</td>
<td>2.76 (2.70)</td>
</tr>
<tr>
<td>Lexical diversity</td>
<td>67.26 (21.42)</td>
<td>22.12–117.34</td>
<td>57.27 (15.77)</td>
</tr>
<tr>
<td><strong>Syntactic features</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses per T-unit</td>
<td>1.77 (0.33)</td>
<td>1.17–3.1</td>
<td>1.68 (0.33)</td>
</tr>
<tr>
<td>Words per Clause</td>
<td>6.02 (0.73)</td>
<td>4.55–8.40</td>
<td>5.44 (0.51)</td>
</tr>
</tbody>
</table>

* p < 0.006, adjusted alpha value based on Bonferroni correction.

significantly at most lexico-syntactic dimensions. When length is measured by number of clauses, narratives were significantly longer than argumentative essays (p = 0.003). Despite that no significant difference is found in the total number of words across genres (p = 0.69), argumentative texts displayed considerably higher frequencies of complex lexical features, namely: a significantly greater frequency of long words that contain three or more syllables (p < 0.001), semantically abstract nouns (p < 0.001), and more diverse vocabulary (p < 0.001). Surprisingly, narratives contained, on average, more academic vocabulary than argumentative essays (p < 0.001). However, it is noteworthy that both genres displayed only limited use of academic vocabulary, ranging from an average of 1.36–2.76 academic words per text. In addition, argumentative texts also differed from narratives in syntactic features when measured by the number of words per clauses (p < 0.001), but there is no statistically significant cross-genre difference in the number of clauses per T-unit (p = 0.05). Thus, argumentative texts tended to display a higher level of lexico-syntactic sophistication than narrative texts, as indexed by most measures, with the exception of frequency of academic vocabulary and number of clauses per T-unit.

6.2. RQ2: predicting writing quality within genre

6.2.1. Genre-specific discourse features

Table 3 exhibits both frequency and diversity of discourse markers coded in these secondary school EFL learners’ argumentative and narrative texts. In argumentative essays, 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. Additionally, more than 60% of essays contained code glosses (e.g., for example, such as), 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, which 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. The most widely used type of stance markers were markers of personal beliefs (e.g., I think), which appeared in 85 essays. However, a closer look at the data revealed overabundant usage of this type of stance markers in repetitive pattern, such as using “I think . . . ” every time a new argument was introduced. Consistently with prior research (Reilly et al., 2002), it was not always possible to decide whether this is a form of epistemic stance or was only used as discourse-sequencer. 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. As expected, temporal markers (e.g., earlier, finally) displayed the highest frequency. Comparably, 88 students used causal markers (e.g., because, the reason that . . . ) followed by adversative markers (e.g., though, however). Finally, 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 stance markers. All four types of stance markers were identified in 80 out of the 100 narratives. Evaluative qualifiers, that is, the use of adjectives (e.g., unforgettable), and adverbs (e.g., actively), and intensifiers (e.g., really, very much), were 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.
6.2.2. Correlations of linguistic features with writing quality

Table 4 presents the intercorrelations among argumentative and narrative writing quality ratings and all linguistic measures. As expected, text length, as measured by number of clauses, displayed statistically significant high correlations with both argumentative and narrative quality ratings, indicating the necessity to include it as a control variable in subsequent regression analyses. All lexical measures not only showed moderate-to-high correlations across genres, but also strong correlations among themselves (see Table 4). Therefore, following previous research (Uccelli et al., 2013), we created a composite score named lexical complexity using Principal Component Analyses (Eigenvalue = 2.06, 70% of the variation was explained), and weighted the three lexical measures—frequency of long words, abstract nouns, and lexical diversity—by their factor loadings of 0.62, 0.60, and 0.51, respectively. At the syntactic level, words per clause showed a moderate and significant relation with argumentative quality, and a low-to-moderate correlation with narrative quality that approached significance. However, no statistically significant relation was detected between clauses per T-unit and writing quality ratings in either genre.

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 ratings. For argumentative essays, frequency and diversity of organizational markers displayed moderate pairwise correlations with writing quality, but only diversity of stance markers showed a weak positive (but non-significant) association with argumentative quality. As for narratives, both frequency and diversity of organizational and stance markers showed high correlations with writing quality ratings, but it needs to be tested in regression analyses whether such association is confounded by text length.

6.2.3. Distinct predictors of writing quality within genre

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 ratings of argumentative and narrative writing quality (see Tables 5 and 6, respectively). Visual inspection of the distribution of standardized residuals for each model confirms that all assumptions of Multiple Regression are met.
A4 was retained as the final model as it is the most parsimonious model containing promising predictors of writing quality scores, which explain 51.6% of the variation in writing quality ratings. All possible interactions were tested, but none was found to be statistically significant.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argumentative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. text length</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. w_long</td>
<td>0.58</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. w_abstract</td>
<td>0.58</td>
<td>0.61</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. w_diversity</td>
<td>0.48</td>
<td>0.34</td>
<td>0.42</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. sx_w/c</td>
<td>0.32</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.36</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. sx_c/t</td>
<td>-0.13</td>
<td>-0.18</td>
<td>-0.24</td>
<td>-0.17</td>
<td>-0.05</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. org_freq</td>
<td>0.47</td>
<td>0.58</td>
<td>0.41</td>
<td>0.47</td>
<td>0.09</td>
<td>0.19</td>
<td>-0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. org_divs</td>
<td>0.46</td>
<td>0.42</td>
<td>0.39</td>
<td>0.35</td>
<td>0.19</td>
<td>0.22</td>
<td>-0.09</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. sta_freq</td>
<td>0.04</td>
<td>0.29</td>
<td>0.15</td>
<td>0.08</td>
<td>0.03</td>
<td>-0.06</td>
<td>0.02</td>
<td>0.34</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>11. sta_divs</td>
<td>0.15</td>
<td>0.28</td>
<td>0.17</td>
<td>0.11</td>
<td>0.16</td>
<td>0.04</td>
<td>0.14</td>
<td>0.36</td>
<td>0.27</td>
<td>0.75</td>
</tr>
<tr>
<td>Narrative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. text length</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. w_long</td>
<td>0.62</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. w_abstract</td>
<td>0.59</td>
<td>0.70</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. w_diversity</td>
<td>0.44</td>
<td>0.45</td>
<td>0.51</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. sx_w/c</td>
<td>0.19</td>
<td>0.05</td>
<td>0.28</td>
<td>0.28</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. sx_c/t</td>
<td>0.09</td>
<td>0.25</td>
<td>0.22</td>
<td>0.23</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. org_freq</td>
<td>0.60</td>
<td>0.73</td>
<td>0.53</td>
<td>0.47</td>
<td>0.23</td>
<td>0.30</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. org_divs</td>
<td>0.55</td>
<td>0.62</td>
<td>0.44</td>
<td>0.45</td>
<td>0.20</td>
<td>0.21</td>
<td>0.10</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. sta_freq</td>
<td>0.70</td>
<td>0.69</td>
<td>0.56</td>
<td>0.50</td>
<td>0.27</td>
<td>0.21</td>
<td>0.02</td>
<td>0.58</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>11. sta_divs</td>
<td>0.84</td>
<td>0.56</td>
<td>0.34</td>
<td>0.38</td>
<td>0.19</td>
<td>0.17</td>
<td>0.53</td>
<td>0.50</td>
<td>0.53</td>
<td></td>
</tr>
</tbody>
</table>

Note: quality = writing quality ratings; text length = number of clauses; w_long = frequency of long words; w_abstract = frequency of abstract nouns; w_diversity = vocd; sx_w/c = words per clause; sx_c/t = clauses per t-unit; org_freq = frequency of organizational markers org_divs = diversity of organizational markers; sta_freq = frequency of stance markers; sta_divs = diversity of stance markers.

We first used a theory driven incremental approach of hierarchical multiple regression analyses to explore linguistic features that predict writing quality ratings of argumentative essays. First, we entered text length as a control variable, which accounted for 40% of the variation in writing quality ratings. Then, we introduced the key predictor variables one at a time, starting with lexical complexity. We found a significant main effect of lexical complexity on writing quality, controlling for text length. In Model A3, syntactic complexity explained an additional 5% of the variance in quality scores. In Model A4, 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 relation worth exploring further in a larger sample. In Model A5, we added diversity of stance markers as an additional predictor, which, however, showed no statistically significant association with writing quality, controlling for other variables. Using Akaike information criterion (AIC), Model A4 was retained as the final model as it is the most parsimonious model containing promising predictors of writing quality scores, which explain 51.6% of the variation in writing quality ratings. All possible interactions were tested, but none was found to be statistically significant.

Table 5

<table>
<thead>
<tr>
<th></th>
<th>Model A1</th>
<th>Model A2</th>
<th>Model A3</th>
<th>Model A4</th>
<th>Model A5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>0.11 (0.014)</td>
<td>0.04 (0.018)</td>
<td>0.06 (0.018)</td>
<td>0.05 (0.019)</td>
<td>0.06 (0.019)</td>
</tr>
<tr>
<td>Lexical complexity</td>
<td>0.72 (0.146)</td>
<td>0.47 (0.164)</td>
<td>0.46 (0.163)</td>
<td>0.45 (0.164)</td>
<td>0.45 (0.164)</td>
</tr>
<tr>
<td>Syntactic complexity</td>
<td>0.70 (0.233)</td>
<td>0.62 (0.236)</td>
<td>0.63 (0.237)</td>
<td>0.63 (0.237)</td>
<td>0.63 (0.237)</td>
</tr>
<tr>
<td>Organizational markers (diversity)</td>
<td>0.24 (0.157)</td>
<td>0.26 (0.161)</td>
<td>0.26 (0.161)</td>
<td>0.26 (0.161)</td>
<td>0.26 (0.161)</td>
</tr>
<tr>
<td>Stance markers (diversity)</td>
<td>3.93 (0.475)</td>
<td>5.99 (0.584)</td>
<td>1.29 (1.669)</td>
<td>1.36 (1.658)</td>
<td>1.38 (1.664)</td>
</tr>
<tr>
<td>_cons</td>
<td>3.93 (0.475)</td>
<td>5.99 (0.584)</td>
<td>1.29 (1.669)</td>
<td>1.36 (1.658)</td>
<td>1.38 (1.664)</td>
</tr>
<tr>
<td>N</td>
<td>100</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.400</td>
<td>0.456</td>
<td>0.504</td>
<td>0.516</td>
<td>0.518</td>
</tr>
<tr>
<td>AIC</td>
<td>383.680</td>
<td>347.986</td>
<td>341.094</td>
<td>340.747</td>
<td>342.33</td>
</tr>
</tbody>
</table>

\( p < 0.1 \)
\( p < 0.05 \)
\( p < 0.01 \)
\( p < 0.001 \)
In a similar process, text length was first entered into a baseline model to predict narrative writing quality (see Table 6), which accounted for 53.6% of the variance in narrative writing quality ratings. In Model N2, when introducing lexical complexity as a key predictor, we saw a decrease in the $R^2$, which indicates that lexical complexity did not contribute to explain additional variation in writing quality ratings above and beyond text length. This is expected given the high correlation between length and lexical complexity ($r = 0.72$, $p < 0.001$). Therefore, lexical complexity was dropped from subsequent models to resolve the collinearity issue. In Model N3, syntactic complexity showed a positive main effect controlling for text length, but in contrast to what we observed in argumentative models, such effect became non-significant in Models N4 and N5 when introducing the frequency of the two types of discourse markers. We were also encouraged to see that the frequency of stance markers was a significant predictor of writing quality. Moreover, the significant interaction (in Model N6) denoted that the effect of frequency of stance markers in written narratives was weaker for shorter texts than longer ones. Ultimately, using Akaike information criterion (AIC), Model N6 was retained as the final model, explaining 63.7% of the variability in narrative writing quality.

### 6.2.4. Illustrating predictors of writing quality

In the appendices, we display excerpts from two examples per genre that illustrate the higher and lower ends of the writing quality continuum for the sample of students. These examples were selected to illustrate how specific lexico-syntactic and discourse features contributed to differentiated higher-quality and lower-quality writing in each genre. It is worth noticing that all four examples show space for improvement in lexico-grammatical accuracy. Given that a large amount of research has been conducted to specifically address issues of mechanical or grammatical accuracy (Chandler, 2003; Polio, 1997), the present study focused instead on lexico-syntactic and discourse resources.

Argumentative Essay 1 received a score of 4 out of 12, whereas Argumentative essay 2 received 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, this essay represents the best writing performance produced by this sample. In comparing the two argumentative texts displayed in Appendix B, we can observe that the high-quality essay differed from the low-quality essay in lexical complexity, syntactic complexity, and diversity of organizational markers. At the lexico-syntactic level, Essay 2 demonstrated more incidences of structurally complex (e.g., extremely, talented) and semantically abstract vocabulary (e.g., achievement, courage) compared to Essay 1. Moreover, the syntactic complexity in Essay 2, as measured by average words per clause, was 6.63 compared to 5.31 in Essay 1, which indicated a higher frequency of embedded clauses used in Essay 2 (e.g., because a person who doesn’t work hard may not be successful though he is extremely talented.). At the discourse level, Essay 2 displayed the author’s strategic use of diverse organizational markers to signal argumentative structure. 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., for example) and conclusion markers (e.g., in summary), which successfully oriented the reader to the progression of arguments.

Appendix B displays excerpts of relatively low- and high-quality narrative writing, with scores of 4 and 12 respectively. As observed in the text, the two narratives do not differ remarkably in lexical and syntactic complexity, as both make use of mostly simple vocabulary (e.g., wide, big) and syntactic structure. However, we can observe a considerable difference in the use of evaluative markers between these two narratives. Narrative 2 displays abundant use of evaluative stance markers compared to Narrative 1, including description of internal states (e.g., nervous), evaluative qualifiers (e.g., scary, suddenly), exaggeration (e.g., I almost couldn’t move), and subjective 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 narrative and actively engaged readers in the story.

### 7. Discussion

In sum, in contrast to the findings for native English writers, the present study revealed no statistically significant differences in Chinese secondary school EFL learners’ writing quality in narrative and argumentative essay. However, in line...
with research on native English adolescent writers, argumentative writing demonstrated more sophisticated lexico-syntactic features, showing higher frequency of long words, abstract nouns, and words per clause. 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 was 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.

7.1. Similarities and differences with previous research

Our finding of higher-level lexico-syntactic complexity in EFL learners' argumentative writing compared to their narratives is similar to what researchers have found with native English speakers (Berman & Nir-Sagiv, 2007). This finding indicates that EFL learners not only distinguished each genre's communicative purposes—which would be expected at this age, but also adequately deployed relevant linguistic resources from their non-native language to serve these distinct purposes.

Nevertheless, our results differ from previous research on native English speakers' cross-genre writing in several aspects. First, in contrast to the replicated finding of native English secondary school writers achieving higher overall quality in narrative as compared to argumentative writing, our analysis with EFL learners revealed no differences in quality across genres.

Moreover, 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 in both genres. One possible reason that clauses per T-unit did not correlate with writing quality might be these EFL learners' redundant use of the same syntactic structure. For example, many essays started with a sentence of the form "I think . . . because . . ." and repeated this formula a number of times. The finding that clause subordination was correlated with neither narrative nor argumentative writing quality also lends support to Biber et al. (2011), showing that dense use of clause subordination may not necessarily indicate higher writing quality.

7.2. Limitations and future directions

While promising, our results must be viewed in light of some critical limitations. First, it is necessary to acknowledge that our findings reveal characteristics of these writers' performances on what is called on-demand writing, without possibly capturing their full capabilities as EFL writers under different conditions. Thus, the study did not attempt to assess students' overall writing ability; rather, it had the modest goal of assessing the features of particular writing products. Moreover, the sample contains a wide age range without large numbers of participants per grade. Thus, results only demonstrate individual variability and could by no means offer general developmental trends. Cross-sectional studies with larger sample sizes at each grade level or, ideally, longitudinal research that follows writers throughout the secondary school years would be helpful to examine developmental trajectories and individual variability in the development of EFL writing performance across genres. Thirdly, the linguistic measures adopted in the present study only assessed certain aspects of linguistic complexity (e.g., clausal complexity, clause subordination), whereas other important measures deemed to be critical in advanced academic writing (e.g., complex noun phrase constituents) were not explored. Future research could conduct a more comprehensive evaluation of linguistic complexity in this population. Lastly, students' writing performance in L1 plays an important role in their EFL writing performance, yet information on students' L1 was not available for this sample. In future research, it will be interesting to explore cross-linguistic relations between EFL learners' L1 and L2 genre knowledge and cognitive abilities. Moreover, it would be insightful to examine the relative importance of L1 genre knowledge and L2 language proficiency in predicting L2 writing performance across genres.

7.3. Instructional implications

Our study revealed several possible areas that merit special instructional attention. First, results showed that Chinese secondary-school EFL learners in the sample used a limited number of academic vocabulary in writing. This suggests that these students would benefit from instruction that was designed to expand their academic vocabulary through purposeful selection of words and optimal conditions for students to actively use academic vocabulary orally and in writing (Coxhead, 2000). EFL learners' limited use of stance markers, especially the epistemic stance markers, is worth noticing as well. These markers constitute grammatically complex and cognitively advanced forms, which usually develop around the adolescent years in native English writers. It would be interesting to explore whether the use of these markers could indeed be scaffolded through EFL instruction designed to scaffold both language and cognitive development; or whether, instead, this constitutes an area of later development in EFL writing compared to native language writing.

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. The findings are in line with a view of language learning as a repertoire of distinct context-specific types of discourses that learners acquire on the basis of opportunities to learn—learning one set of
discourse practices relevant for a particular context does not guarantee language performance in other contexts. In the case of EFL learners, we cannot assume that those students who are proficient in complex uses of language (e.g., argumentative writing) would automatically master other supposedly “easier” discourse practices or those acquired earlier by native language learners (e.g., narrative writing). Teaching EFL cannot be understood as promoting a global English proficiency expected to be equally functional across contexts. Instead, EFL researchers and practitioners should reflect on the diverse needs students would encounter in authentic communicative contexts in order to make informed decisions regarding which discourse practices should be selected for instruction.

Another important motivation of this work is to make visible to EFL practitioners and researchers a repertoire of linguistic features that are closely associated with high writing quality, at least as measured by experienced native-English-speaking teachers. The predictive relation differs by genre. Therefore, our findings highlight the need to make students aware of the different language resources expected across genres. Our findings are relevant to inform the design of pedagogical approaches attuned to different genres, in particular, narrative and argumentative writing. Future research could design interventions informed by the strengths and weaknesses identified in this study in order to explore how to effectively teach these research-based repertoires of skills. Testing research-based interventions can inform classroom practices that seek to promote EFL learners’ writing performance and especially encourage students to understand language as a functional solution to specific contexts of communication.

Appendix A. : The writing prompt

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. You will have 40 min to complete this essay. A typical effective response would be between 300 and 350 words.

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. You will have 40 min to complete this narrative story. A typical effective response would be between 300 and 350 words.

Appendix B.

Students’ argumentative writing: low-quality and high-quality examples

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

Lots of people think success is a further question, it’s hard to be. But I think successful is just around us. Everyone wants to be success. because of his hard work. It’s the thing that everybody can do.

Essay 2: High-quality writing (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.

Students’ narrative writing: low-quality and high-quality examples

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

I have a story let me tell you. It is about my success. When I was 11 years old in summer, I played computer games at home. I finished my homework and have a break. It was five o’clock. At that moment, I heard someone was shouting, so I ran out of the door and look what wrong. Suddenly, I know at six
floor was on fire. Everybody were ran outside the building. I get back to my home. There was no people in it, only me. So I pack something very important. And I ran outside the building. A few minutes later, firemen put down the fire.

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

I didn’t like swimming at all when I was young. But after that rainy night, I think that I’m interested in it. I’m afraid of water when I was young. I think there might be something scary in the deep water. Because the water is so wide and big. I can’t imagine what can I move in the scary place. But I changed my opinion after one night. That night, I was taken by my mother. She wanted me to learn how to swim. It was such bad 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 scary water quickly. My mother first dived into the water, But I just waited on the bank. My mother was a little angry and asked me to jump into the pool. I was very nervous at that moment. I didn’t want to swim at all because of the scary water. But I should obeyed my mum. I almost couldn’t move then. 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 surprised 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 very young, he was swimming difficulty but very hard. I admired him and I suddenly got a point. Such a young boy could swim, but why I couldn’t make it? I felt so shame and I thought about my mum’s words again. After a while, I suddenly stood up and rushed into the water without hesitation. When I got into the water, I felt so weird. Why should I be afraid of it? Then I became to swim slowly. My mum looked at me and gave me a smile. I was full of confidence at that time and became more and more faster. I know that I should believe myself. The things that disturb you are not so difficult, I can make it!

**Bold:** Organizational markers.

_Underline:_ Stance markers.

References


Paola Uccelli is associate professor at the Harvard Graduate School of Education. Her current research focuses on writing development of learners using English as a Foreign Language (EFL). Beyond lexico-grammatical accuracy, she is particularly interested in how EFL writers learn to flexibly and effectively deploy a variety of lexical, syntactic and discourse features that are attuned to different communicative contexts, genres and registers.