

Article

Metadiscourse, Cohesion, and Engagement in L2 Written Discourse

Jianling Liao

School of International Letters and Cultures, Arizona State University, Tempe, AZ 85287-0202, USA;
jianling.liao@asu.edu

Received: 27 February 2020; Accepted: 21 May 2020; Published: 5 June 2020

Abstract: The current study examines how L2 Chinese writers at different proficiencies employed various metadiscourse devices to shape their written descriptive discourse and also whether various metadiscourse features may distinguish levels of writing proficiency. The study also looks at how L2 learners' use of metadiscourse devices is related to their linguistic performances in descriptive writing. The findings revealed differential metadiscourse use by learners at different proficiencies on local, global, and textual organizational dimensions. For instance, compared to low-proficiency writers, more proficient writers used significantly more conditional/hypothetical markers, frame markers, and engagement markers. Multiple metadiscourse features also demonstrated significant positive and negative correlations with each other, suggesting patterns of decreases and increases in the use of particular organizational features. Several metadiscourse features characteristic of more advanced writers also displayed positive relationships with linguistic features.

Keywords: L2 Chinese; descriptive writing; metadiscourse; cohesion; local; global; text; interactive; interactional

1. Introduction

Writing in an L2 involves not only an effort to monitor linguistic quality, such as linguistic accuracy or complexity, but also an effort to make metadiscourse choices that will result in cohesive written discourse. An examination of L2 writers' metadiscourse performances will allow for a fuller understanding of L2 writing skills, in terms of how learners allocate their cognitive resources to different areas of writing and how successful they may be in each specific area.

Currently, studies have examined organizational quality in L2 texts mainly by analyzing cohesion and coherence (e.g., Chiang 2003; Crossley et al. 2016a; Ferris 1994; Guo et al. 2013; Harman 2013; Jafarpur 1991; Kormos 2011; Liu and Braine 2005; Yang and Sun 2012). It has been argued that using more logical operators and cohesive devices, such as metadiscourse markers, semantic repetitions, and co-referentiality, will contribute to a more cohesive text (Bardovi-Harlig 1990; Chen and Baker 2016; Connor 1990; Crossley et al. 2016a; Ferris 1994; Guo et al. 2013; Halliday and Hasan 1976; Reid 1992; Yang and Sun 2012). Studies have also investigated the interpersonal dimensions of L2 writing by examining authorial identity and engagement with the reader. It is claimed that the presence of devices that express authorial voice and involve readers will enhance the effectiveness of a text (e.g., Hyland 2005; Lee and Deakin 2016; Zhao 2013).

Despite increased empirical understanding of L2 textual organizational performances, overall knowledge of cohesion and other aspects of textual organization in L2 texts is still limited (Crossley et al. 2016a). For instance, it remains unclear what types of metadiscourse devices L2 writers at different proficiencies may apply to shape their writing on local, global, and text levels. How different types of metadiscourse devices may work together to affect the organizational quality of an L2 written text is also understudied. Furthermore, L2 written organizational features have often been

researched in isolation from other aspects of writing. Their relationship to such areas such as linguistic accuracy or complexity still needs to be explored to allow a more complete picture of L2 writing performance and development. Additionally, current studies have mainly examined writing in English as a second or foreign language (ESL/EFL). We still need to understand textual organizations in other L2s. The current study attempts to address these research gaps by investigating how L2 Chinese writers at different proficiencies deploy metadiscourse devices to form text dynamics and how textual organizational features interconnect with linguistic features in descriptive writing.

2. Literature Review

The literature is surveyed in three areas to provide relevant background on (a) how L2 written organizational quality is currently theorized; (b) how L2 written organizational performances are operationalized; and (c) how L2 learners' textual organizational skills develop.

2.1. Organizational Quality in L2 Texts

L2 writing researchers have investigated textual organizational features in two different yet related dimensions: text structure and interpersonal engagement (with the reader). Text structure is often characterized by two frequently researched textual organizational constructs: cohesion and coherence (e.g., Chiang 2003; Crossley et al. 2016a; Ferris 1994; Guo et al. 2013; Harman 2013; Jafarpur 1991; Kormos 2011; Liu and Braine 2005; Yang and Sun 2012). Although definitions may vary, cohesion in general refers to making connections between ideas for the creation of a coherent and comprehensible discourse (Halliday and Hasan 1976). The cohesiveness or coherence of a text concerns not only whether the preceding and incoming discourses are appropriately linked to advance meaning, but also whether the presented meaning representation may be effectively understood by the reader. Researchers have argued that higher quality writing displays stronger textual cohesion and coherence (Connor 1990; Crossley et al. 2016a; Ferris 1994; Yang and Sun 2012).

To a certain degree, cohesion and coherence appear to be a pair of related traits of textual organizational quality from the writers' and the readers' perspectives, respectively (Crossley et al. 2016a, 2016b; McNamara et al. 1996). From the writer's perspective, cohesion involves the writer's intention to create a text that flows logically; from the reader's side, coherence concerns whether a text is perceived as flowing effectively. Thus, the interpretation of either cohesion or coherence may involve a certain level of subjectivity, whether from the writer or from the reader. An evaluation of cohesion or coherence in an L2 text may also involve an additional level of addressing the possible transfer effects from the writer's L1. A piece of text deemed cohesive in the learner's L1 may be incoherent in the L2, due to possibly distinct rhetorical norms observed in the two languages.

The cohesive ties in a text can be explicit or implicit. Explicit cohesion markers often refer to logical connectives, such as conjunctions, adverbs, or lexical bundles (Chen and Baker 2016; Crossley and McNamara 2012; Guo et al. 2013; Yang 2013). Logical connectives can serve a useful role in terms of creating explicit links and relations between the ideas in a text, which have also been classified into different logical categories by researchers, further explained in the next section. Less explicit cohesive devices include global cohesion features such as lexical, argument, or semantic overlap and co-referentiality (Halliday and Hasan 1976). Nowadays, global cohesive features are often analyzed using computerized programs, especially in ESL/EFL studies (e.g., Crossley et al. 2016a; Crossley and McNamara 2012; Crossley et al. 2011; Guo et al. 2013; Kormos 2011). For instance, through computational tools, latent semantic analysis computes sentence-to-sentence conceptual similarity in a text, by examining meaning overlap between explicit words or words that are implicitly related in meaning (Graesser et al. 2004; Guo et al. 2013; Mazgutova and Kormos 2015).

Hyland (2005), however, argued that there are limitations for observing metadiscourse usages without considering the interaction between the text and the reader. He proposed an interpersonal framework of metadiscourse and posited that one essential purpose for the writer to employ metadiscourse devices is to guide the reader's understanding of the text towards his or her preferred interpretations. Hyland further categorized metadiscourse devices into two taxonomies: interactive and interactional. Interactive metadiscourse realizes functions similar to cohesion conceptualization,

but with a stronger focus on the consequential interpretations that may be made available to the reader. Second, Hyland argued for the need to examine interactional metadiscourse devices in a text, through which the writer brings in authorial voice and engages the reader. How metadiscourse features are specifically operationalized is elaborated next.

2.2. Textual Organizational Devices in L2 Texts

Researchers have proposed various frameworks to operationalize written organizational performances. Crossley et al. (2016a, 2016b) categorized cohesive indices into local, global, and text levels, to allow for a more fine-grained understanding of text cohesion. Local cohesive devices refer to the connectives within/between clauses/sentences. The quantity of preposition usages has also been evaluated to understand intra-clausal cohesion (Crossley et al. 2016a; Reid 1992; Smith and Frawley 1983). As mentioned earlier, global and text cohesion devices tend to be more implicit. Global cohesive devices include connectives between paragraphs or larger chunks of texts, as well as lexical and semantic overlap between the paragraphs in a text (Guo et al. 2013; Halliday and Hasan 1976; Li 2014). Text cohesion concerns cohesiveness across the text and is often assessed through features such as proportion of given/new information (e.g., pronoun/noun ratio, pronoun density), lexical repetitions, or lexical diversity (Crossley et al. 2016a; Kyle and Crossley 2017; Reid 1992).

As discussed earlier, Hyland's (2005) interpersonal analysis framework divides metadiscourse into interactive and interactional devices. Interactive devices build links between ideas in line with the writer's intended interpretation for the reader. According to Hyland, such connectors may include the use of transitional markers, which indicate relations between clauses (e.g., addition, adversative); frame markers, which signal discourse acts, such as sequencers, stage labels, announcements of goals, and topic shifters; endophoric markers, which provide references to information in other parts of the text; evidentials, which provide citations within a community-based literature; and code glosses, which provide reformulations and exemplifications. In contrast, interactional metadiscourse markers serve interpersonal functions, which include hedges used to withhold authorial commitment and open dialogue, boosters used to emphasize writer's certainty, attitude markers that express the writer's attitude to proposition, self-mentions explicitly referencing to authors, and engagement markers that involve the reader in the discourse.

Other researchers have proposed situational models of cohesion, which identify various situational dimensions of cohesion, such as causation, time, space, intentionality, or protagonists, expressed through particles, nouns, prepositions, verbs, or word inflection features (Kintsch 1998; Kormos 2011; Van Dijk and Kintsch 1983; Zwaan and Radvansky 1998). For example, causal cohesion evaluates the extent to which causal links between sentences are expressed; temporal cohesion reflects the extent to which tense and aspect assist in the formation of cohesion; and spatial cohesion looks at how different contents are linked by spatial particles or relations, such as the incidences of location nouns, prepositions, and motion verbs. Additionally, researchers have classified coherence relations into positive relations, i.e., extending the information provided in the text; and negative relations, i.e., restricting or ceasing to elaborate information (Louwerse 2002; Sanders et al. 1992).

Studies on L1 and L2 English writing have reported various kinds of relationships between the use of specific textual organizational features and essay quality. Studies on L1 English writers have found that global cohesion (e.g., semantic links between paragraphs) positively relates to human judgments of writing quality (McNamara et al. 2013; Neuner 1987). Local and text cohesions, however, are not strong indicators of human judgments of writing quality (Crossley et al. 2016b; Evola et al. 1980; McNamara et al. 2010). Guo et al. (2013) examined how features, including lexical sophistication, syntactic complexity, cohesion, and text length, may predict human judgments of quality of TOEFL iBT integrated essays (i.e., reading-listening to summary writing) and independent essays (i.e., argumentative writing). They found that lexical sophistication, text length, and use of past participle verbs significantly predicted essay scores for both types of essays. Nevertheless, cohesion features including semantic similarity, noun overlap, and tense repetition predicted only writing quality for integrated essays. The number of conditional connectives, content-word overlap, and aspect repetition negatively predicted or correlated with the writing quality of independent

essays. Guo et al. argued that the two writing tasks may be assessed with similar and distinct criteria. Zhao (2013) investigated authorial voice in EFL argumentative writing. He found a positive correlation between authorial voice and ratings of writing quality.

Thus, textual organizational features in L2 writing have been operationalized at multiple discourse levels (e.g., clause, sentence, paragraph, text), in intra-text or writer–reader interpersonal dimensions, as well as in various logical categories. Taken together, these perspectives improve our understanding of how L2 writers shape written discourse at various textual levels and how they communicate their intended meaning to potential readers. To obtain a fuller picture of metadiscourse measures that learners take to form their writing, the current study incorporates relevant perspectives from the theoretical frameworks discussed above, to examine how L2 Chinese learners shape written descriptive discourse. For instance, self-mentions and engagement markers in Hyland’s (2005) interpersonal metadiscourse framework were included in the current analysis because they are applicable to the current writing prompt (i.e., introducing one’s institution to friends), further explained in Section 3.2.

2.3. Development of L2 Textual Organizational Skills

Studies have investigated whether and, if so, how, L2 learners at different proficiencies may demonstrate differential patterning in using textual organizational features in their writing. Researchers have argued that low-proficiency writers may need to allocate significant attentional resources to low-level processing, such as spelling or linguistic encoding due to limited language skills (Kormos 2011). Consequently, they may not be able to devote sufficient cognitive resources to more global aspects of writing, such as textual organization (McCutchen 1996). In contrast, more proficient writers may attend to multiple writing areas more successfully, use more effective metadiscourse devices, and produce more cohesive texts (Bardovi-Harlig 1990; Chen and Baker 2016; Connor 1990; Crossley et al. 2016a; Lee and Deakin 2016; Yang and Sun 2012). More advanced writers may also have a greater assortment of lexical and referential devices at their disposal to promote textual cohesion (Halliday and Hasan 1976).

A number of studies have provided empirical evidence to show that compared to low-level writers, more proficient writers deploy more sophisticated and a greater range of metadiscourse devices, use cohesive devices more accurately, and present authorial voice and engage the reader more effectively. Yang and Sun (2012) discovered that L1-Chinese fourth-year college English learners used a greater number of cohesive devices in their argumentative writing and used them more accurately than second-year learners. Similarly, Ferris (1994) reported that higher-proficiency English learners used more cohesive devices that showed pragmatic appropriateness. Over a semester-long upper-level English for Academic Purposes course, Crossley et al. (2016a) found that students increased their use of local, global, and text cohesive devices in their writing. The usages of cohesive features at the local, global, and text levels predicted with a 71% accuracy rate whether an essay was written at the beginning or at the end of the semester. The cohesion features also explained 42% of the variance in the judgments of writing quality. Chen and Baker (2016) examined lexical bundles in argumentative and expository English writing. They discovered that the lexical bundles in lower-proficiency writing shared more similarity with conversational language, whereas more proficient essays were characterized by more formal lexical bundles that were closer to the register of academic prose. Reid (1992) reported that English learners, regardless of their L1s, used a lower percentage of prepositions than native writers in their essays of two topic types (comparison/contrast; chart/graph). Adopting Hyland’s (2005) interpersonal metadiscourse framework, Lee and Deakin (2016) compared the usages of stance and engagement resources among three corpora of college English learners’ argumentative essays: successful and less-successful essays produced by L1-Chinese learners; and successful native English essays. Their analyses revealed that successful essays by both native and L2 writers contained significantly greater instances of hedges than less-successful essays. Compared to native writers, both groups of L2 writers were overwhelmingly resistant to establishing an authorial identity in their essays. A comparative study of English and Chinese academic writing, Hu and Cao (2011) examined the use of hedges and boosters in academic article

abstracts published in applied linguistics English-medium (by both native and non-native writers) and Chinese-medium journals. They found that the abstracts published in Chinese-medium journals featured hedges markedly less frequently than those in English-medium journals which, according to Hu and Cao, can be attributed to distinct culturally preferred rhetorical norms in the Chinese and Anglo-American academic communities, respectively. Thus, the use of interactional devices may be culture-specific.

Similar to the L2 English findings, two studies on L2 Chinese writing have reported that, in comparison with native Chinese writers, L2 Chinese writers used a lower number or a narrower range of cohesive devices. Using a corpus-based approach, Li (2014) compared lexical cohesion in 50 argumentative compositions produced by advanced L1-English Chinese learners in a proficiency test with that in 50 native Chinese argumentative essays produced for the Chinese National College Entrance Examination. Li investigated various lexical cohesion features, including simple and complex repetitions, simple and complex paraphrases, superordinate and hyponymy, co-reference, and bond density (i.e., lexical repetitions across sentences). He found that, compared to native writers, L2 Chinese writers applied a lower frequency of simple and complex paraphrases and superordinate and hyponym relations, as well as a lower ratio of bond-forming sentences. Yang (2013) investigated the use of textual conjunctives and topicalizers in 30 written summaries produced by three fourth-year college Chinese learners. He compared the usages with those in the original texts produced by native Chinese authors and found that L2 Chinese learners applied a narrower range of cohesive devices.

On the other side, the findings, however, suggest that more is not necessarily better. Crossley and McNamara (2012) discovered that higher-proficiency L2 English writers produced texts with fewer cohesive devices than lower-proficiency writers. They explained their findings as a reverse cohesion effect: More proficient writers may assume that their audience includes high-knowledge readers, who benefit more from lower-cohesion texts (p. 130). Kennedy and Thorp (2007) similarly reported that compared to learners who received lower band scores, more proficient English learners applied many fewer lexico-grammatical and enumerative markers and subordinators in their argumentative essays, which appeared to be more similar to native-speaker use. In the study on L2 Chinese discussed earlier, Yang (2013) found that compared to the original texts produced by native authors, L2 Chinese learners overused certain types of cohesive devices in their written summaries, such as adversative (e.g., *but*, *danshi*), causative (e.g., *therefore*, *suoyi*), and additive (e.g., *but also*, *erqie*) connectives. Two previously discussed L2 English studies found that a higher ratio of pronouns was associated with low writing proficiency. Reid (1992) discovered that, in comparison with native writers, ESL writers used significantly higher percentages of pronouns and coordinate conjunctions, which was similar to the register of interactive or oral English communications. Crossley et al. (2016a) found that a higher pronoun/noun ratio negatively predicted human judgments of writing quality of academic English essays. Together, these findings suggest that more proficient L2 writers likely use certain cohesive devices more concisely, such as enumerative markers, subordinators, or pronouns.

The findings thus far have enabled us to better understand how L2 writers with different proficiencies may use metadiscourse features in distinct patterns. For instance, we know that higher-level writers deploy a greater range of cohesive devices and use them more accurately (Chen and Baker 2016; Ferris 1994; Li 2014; Yang 2013). We also know that as L2 learners grow their writing skills, they may rely less on coordinate connectives, subordinators, or pronouns, and resort more frequently to lexical cohesive devices or prepositions (Crossley et al. 2016a; Kennedy and Thorp 2007; Li 2014; Reid 1992; Yang 2013). Despite increased knowledge of L2 learners' textual organizational skills, the understandings we have obtained are derived from different studies that have used different writing genres, tasks, or learner proficiencies. It is, thus, difficult to compare the results across studies and develop more integrated knowledge. We do not yet know in a systematic way how various textual organizational features interrelate to influence writing.

Another gap in previous research is that the majority of the studies have examined either linguistic features or discourse features in L2 texts, but not both, which does not allow us to observe how L2 writers pull together linguistic resources to form global meaning. Only a handful of studies

include both linguistic and discourse features in their analyses. A previously discussed study by Crossley and McNamara (2012) examined the predictive effects of text cohesion and linguistic sophistication on L2 writing proficiency among high school English learners. Their results showed that highly proficient writers produced essays that were linguistically more sophisticated, but not more cohesive. Several linguistic and cohesive features, including lexical diversity, word frequency, word meaningfulness, aspect repetition, and word familiarity, significantly predicted writing proficiency. Kormos (2011) investigated the effects of task complexity on linguistic and discourse characteristics of narrative texts produced by upper-intermediate secondary school English learners. She found that a task variable—whether learners had to narrate a story with predetermined content or plan their own story—did not result in substantial linguistic or cohesive differences. The task conditions exerted a major impact on only one measure of lexical sophistication and had a minor effect on the explicit signaling of temporal cohesion. Guo et al. (2013), discussed earlier, examined both linguistic and cohesion features in integrated summary writing and independent argumentative writing, regarding their predictive power for human judgments of writing quality. They found that cohesive features predicted the writing quality of integrated essays only (see Section 2.2 for more details). Ferris (1994) compared lower-level and higher-level ESL texts using 28 linguistic and textual organizational measures. He found that the 28 variables divided the subjects into groups with 82% accuracy, and that higher-level students used a greater variety of lexis, syntactic constructions, and cohesive devices. Thus, these studies have examined written linguistic and cohesion features mainly in terms of their predictive capacity for human judgments of writing quality. They provide little knowledge regarding the interrelations between linguistic and discourse performances in L2 writing. Without such knowledge, we will not understand appropriately the dynamic development of L2 writing ability as a whole.

Furthermore, the previous studies have mainly analyzed argumentative writing (e.g., Chen and Baker 2016; Kennedy and Thorp 2007; Lee and Deakin 2016; Li 2014; Yang and Sun 2012). We still need to explore how learners apply organizational features in other types of writing, such as descriptive writing. The current study examines textual organizational features in L2 Chinese descriptive writing, as well as how organizational features differ between proficiencies. How various organizational features interrelate with each other, as well as how organizational features correlate with linguistic features were also investigated. Three questions guided this study:

1. What kinds of textual organizational features exist in low-score, middle-score, and high-score L2 Chinese descriptive essays, respectively, and how are the organizational features different across the groups?
2. What are the interrelations among various textual organizational features in L2 Chinese descriptive essays?
3. How do textual organizational features relate to linguistic features in L2 Chinese descriptive essays?

3. Methods

3.1. Participants and Dataset

The participants in the current study were 62 L1-English college Chinese learners from the United States, who were in China on a study-abroad program when the data were collected. The dataset comprised 62 descriptive Chinese essays produced by the participants during the placement test administered by the program. There were 27 females and 35 males, with ages from 19 to 22 years.

Students hand-wrote their essays within 30 min, based on the topic of introducing one's home university to one's Chinese friends. A descriptive writing task was used because it was suitable for both lower-level and higher-level learners. The essays were scored on a 6-point holistic scale (see Appendix C). Scores 1–2, 3–4, and 5–6 correspond roughly to the Novice, Intermediate, and Advanced levels of the proficiency scale of American Council on the Teaching of Foreign Languages (ACTFL), respectively (ACTFL 2012). The scale focused on overall writing quality and included general descriptors of the overall quality of language, content, and organization. Specific criteria on linguistic accuracy or complexity were not included; instead, they were incorporated into the descriptors of overall language and content quality. The author and a second rater evaluated the essays. Both raters were experienced college Chinese language educators. For 56 of the 62 essays (90.32%), the two raters' ratings were identical or they differed by one point, which were considered acceptable scores. The two raters' scores were averaged to derive the final score for each essay. Therefore, the final score may be an integer or a 0.5 value. For the six essays whose ratings differed by two points, the final ratings were determined through discussion.

The 62 essays had three score levels: 19 low-score (1.0–2.5), 20 middle-score (3.0–4.0), and 23 high-score (4.5–6.0). According to the program's placement results and course syllabi, the low-score students were mostly placed into Chinese first-year and second-year part I classes, roughly equivalent to the ACTFL Novice Low to Novice High levels; the middle-score students were mostly placed into Chinese second-year part II and third-year classes, roughly equivalent to the ACTFL Intermediate Low to Intermediate High levels; and the high-score students were often placed into Chinese fourth-year and fifth-year classes, roughly equivalent to the ACTFL Advanced Low to Advanced Mid levels. The low-score, middle-score, and high-score essays had a mean length of 152, 230, and 298 characters, respectively, and they contained a total of 2895, 4603, and 6864 Chinese characters, respectively. Table 1 provides a summary of the dataset in this study.

Table 1. Dataset in the current study.

	Low-Score	Middle-Score	High-Score
Score range	1.0–2.5	3.0–4.0	4.5–6.0
Number of essays	19	20	23
Mean essay length (characters)	152	230	298
Total number of characters	2895	4603	6864

3.2. Measures of Textual Organizational Features

To obtain a comprehensive picture of the metadiscourse choices that the learners made in their descriptive writing, a range of theoretically driven indices related to text cohesion and interpersonal features were designated as variables for the data analysis. First, Hyland's (2005) interactive and interactional metadiscourse framework was adopted to capture both text structure and interpersonal characteristics. Second, to analyze organizational features on a finer level, following the methods used in Crossley et al. (2016a), the interactive metadiscourse features were further classified into local (between/within clauses/sentences), global (across idea units), and text indices (across a text), based on the specific metadiscourse functions that individual indices served. Measures drawn from situation models (Kintsch 1998; Kormos 2011; Van Dijk and Kintsch 1983; Zwaan and Radvansky 1998) were also used to classify the metadiscourse features into logical categories. Categories not found in the current dataset, including interactive metadiscourse features, such as evidential, code-

gloss, and endophoric markers and interactional metadiscourse features such as hedge, booster, and attitude markers, which are more relevant to genres such as argumentative or academic writing, were not included in the current analysis. Measures not relevant to the Chinese language were also excluded, for example, cohesions that concern aspect and tense. Since there are no effective computerized tools for analyzing textual organizational features in Chinese texts, features that are difficult to analyze manually, such as lexical and semantic overlap, were not included. Kormos (2011) also argued that cohesive features, such as semantic overlap, co-reference, or latent semantic analysis, may not be fit well with short texts (p. 155), such as the essays in the current dataset.

In particular, local markers denote logical relations between or within clauses and adjacent sentences. Local transitional conjunctions, adverbs, and phrasal bundles were coded into logical categories, including continuative/additive (e.g., moreover, also, next), comparison/contrast (e.g., but), causative (e.g., therefore), and conditional/hypothetical (e.g., only if, if) markers. Adopting the misuse category in Li and Wharton (2012), incorrectly used logical devices that expressed an inaccurate semantic relation between clauses or adjacent sentences were categorized as misuse (p. 348). Moreover, the frequency of preposition usage in an essay was computed to further understand intra-clausal cohesion (Crossley et al. 2016a; Reid 1992; Smith and Frawley 1983). For example, the two prepositions (bold and underlined) in sentences (1) and (2), 跟 *gen* ‘with’ and 对 *dui* ‘to’, connect 我 *wo* ‘I’ and 朋友们 *pengyoumen* ‘friends’, and 我 *wo* ‘I’ and 我的大学 *wodedaxue* ‘my university’, respectively.

(1)

我	星期五	晚上	跟	朋友们	吃	很	好吃	的	饭。
Wo	<i>xingqiwu</i>	<i>wanshang</i>	<u>gen</u>	<i>pengyoumen</i>	<i>chi</i>	<i>hen</i>	<i>haochi</i>	<i>de</i>	<i>fan.</i>
I	Friday	evening	with	friends	eat	very	delicious	Auxiliary	food
I eat very delicious food with friends on Friday evenings.									

(2)

我	对	我的	大学	有	很	深	的	了解。
Wo	<u>dui</u>	<i>wode</i>	<i>daxue</i>	<i>you</i>	<i>hen</i>	<i>shen</i>	<i>de</i>	<i>liaojie.</i>
I	to	my university	have	very	deep			Auxiliary knowledge
I have very deep knowledge of my university.								

Global cohesive devices signal interconnectedness between idea units. In particular, frame markers that signal discourse acts, sequences or stages, or introduce new topics/subtopics were identified (e.g., first of all, finally, in conclusion).

Cohesion across a text was examined by evaluating the amount of given information. The proportion of third-person pronouns, as well as the third-person pronoun/noun ratio in an essay, were calculated to observe givenness and referentiality in a text (Crossley et al. 2016a, 2016b; Kyle and Crossley 2017; Reid 1992; Yang and Sun 2012). A greater proportion of third-person pronouns will indicate a higher amount of given information in a text.

Since the current writing task involved a topic of introducing one’s institution to friends, an examination of interactional metadiscourse usages is relevant for understanding how the writer established authorial presence and engaged the reader. Following Hyland’s (2005) framework, interactional devices were categorized into self-mention (referencing to the author; e.g., I) and engagement markers (address the reader; e.g., you). Possessive first-person pronouns including 我的 *wo de* ‘my’ and 我们的 *women de* ‘our’ were not counted, since these pronouns would be naturally needed to address the current topic and, thus, may not necessarily represent authorial voice.

The analysis of organizational performances was also supplemented with an investigation of their relationships with linguistic performances. Linguistic performances were evaluated for both accuracy and complexity. Complexity was analyzed for lexical and syntactic complexity. Lexical

complexity was operationalized as lexical diversity; syntactic complexity was evaluated by clause length. See Table 2 for details on how the measures were operationalized. Accuracy was analyzed by the ratio of correct clauses in a text. Clauses containing lexical or syntactic errors were counted as incorrect clauses. Since the current writing task was timed (30 min) handwriting, students had to write fast and may produce imperfect characters with inaccurate or missing strokes. These types of errors, however, often did not prevent effective character recognition. Since character accuracy concerns a rather unique language ability and it is not the focus of the current study, characters with incorrect or missing strokes that did not affect recognition were corrected during the transcribing process. More significant character errors that made characters unrecognizable were marked with the symbol *.

3.3. Analysis

Since the essays varied in length, ratios and frequencies were calculated to control for the effect of length. Specifically, the proportions of metadiscourse markers in each category against the total number of interactive or interactional metadiscourse markers were computed to observe which types of metadiscourse features were most or least used to organize ideas or to engage the reader. The percentages of prepositions and third-person pronouns and the ratios of third-person pronoun/noun and correct clauses in an essay were also calculated.

To answer RQ 1, one-way MANOVA test was employed to identify significant differences in textual organizational features among the groups. To answer RQs 2 and 3, Pearson correlations were calculated among the metadiscourse indices, as well as between the metadiscourse and linguistic indices. Descriptive statistical analysis was also conducted. The measures and methods of analysis are summarized in Table 2.

Table 2. Measures used in the current study.

Indices	Analysis Methods
Interactive Metadiscourse Markers	
Local cohesion between/within clauses/sentences	
Transitional markers:	
- continuative/additive (e.g., moreover)	Proportion of the number of transitional markers in each category against the total number of interactive metadiscourse markers (i.e., total number of transitional and frame markers) in an essay
- comparison/contrast (e.g., but, however)	
- causative (e.g., therefore)	
- conditional/hypothetical (e.g., if)	
- misuse: inaccurate logical relations	
Percentage of prepositions: relations among clausal constituents	Proportion of the number of prepositions to the total number of words in an essay
Global cohesion across idea units	
Frame markers: signal discourse acts, sequences, and stages (e.g., first, finally)	Proportion of the number of frame markers to the total number of interactive metadiscourse markers in an essay
Text cohesion	
Givenness: proportion of given to new information	Third-person pronoun/noun ratio: number of third-person pronouns divided by the number of nouns in an essay Third-person pronoun density: proportion of the number of third-person pronouns to the total number of words in an essay
Interactional Metadiscourse Markers	
Self-mention: referencing to the author (e.g., I); Engagement: addressing and involving the reader (e.g., you)	Proportion of the number of interactional markers in each category to the total number of interactional metadiscourse markers (i.e., total number of self-mention and engagement markers) in an essay
Linguistic Indices	
Linguistic accuracy: ratio of correct clauses	Number of error-free clauses divided by the total number of clauses in an essay

Lexical complexity: lexical diversity	Number of word types divided by the square root of the total number of word tokens in an essay
Syntactic complexity: clause length	Total number of words divided by the total number of clauses in an essay

The data were coded by the author and a second rater. The two raters independently coded the same 20% data sample, reaching interrater agreement of 88.24–97.59% for the coding of interactive and interactional metadiscourse measures. Clause accuracy had a lower interrater agreement of 86.84%. Given the challenges in achieving high interrater reliability on accuracy (Polio 1997; Polio and Shea 2014), these relatively low agreement values were considered acceptable.

4. Results

4.1. RQ1: Textual Organizational Features in the Essays

Table 3 presents the mean values of the textual organizational measures. In comparison with the low-score group, the middle-score and high-score groups produced notably higher percentages of conditional/hypothetical, frame, and engagement markers, as well as lower percentages of misuse and self-mention markers. The high-score group also produced the highest percentage (28.76%) of causative markers. The percentage of third-person pronouns and the third-person pronoun/noun ratio consistently increased from the low-score to the high-score group. Across the groups, the continuative/additive, comparison/contrast, and causative markers displayed high percentages. Figure 1 provides a visual illustration of the use of various organizational markers in the three essay groups.

Table 3. Textual organizational measures descriptive statistics.

	Low (<i>n</i> = 19) <i>M</i> (<i>SD</i>)	Middle (<i>n</i> = 20) <i>M</i> (<i>SD</i>)	High (<i>n</i> = 23) <i>M</i> (<i>SD</i>)
Interactive Metadiscourse Markers			
Local cohesion between/within clauses/sentences:			
Continuative/additive marker	17.97% (0.2318)	17.99% (0.2086)	15.47% (0.1371)
Comparison/contrast marker	16.18% (0.2648)	22.20% (0.2398)	17.96% (0.1795)
Causative marker	21.67% (0.2383)	20.42% (0.1849)	28.76% (0.1988)
Conditional/hypothetical marker	1.50% (0.0451)	5.90% (0.0872)	7.64% (0.0951)
Misuse marker	26.36% (0.3262)	9.53% (0.1213)	8.23% (0.0868)
Preposition	3.29% (0.0209)	3.20% (0.0160)	3.27% (0.0115)
Global cohesion across idea units:			
Frame marker	5.80% (0.1076)	23.95% (0.2217)	21.93% (0.1473)
Text cohesion:			
Third-person pronoun	1.49% (0.0118)	1.56% (0.0140)	2.21% (0.0120)
Third-person pronoun/noun ratio	5.54% (0.0466)	5.82% (0.0525)	7.84% (0.0454)
Interactional Metadiscourse Markers			
Self-mention marker	83.49% (0.3194)	77.00% (0.2181)	76.26% (0.2367)
Engagement marker	5.98% (0.1260)	23.00% (0.2181)	19.40% (0.1737)

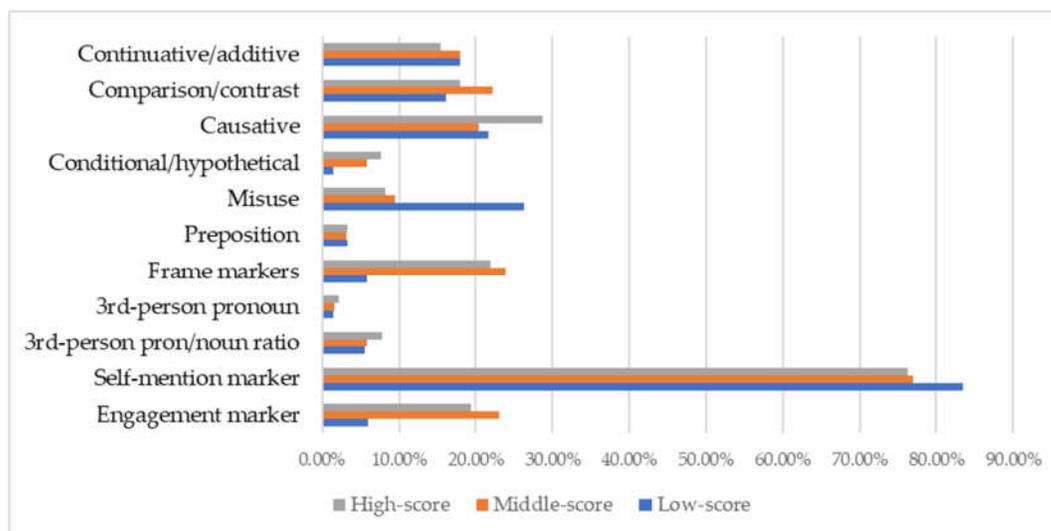


Figure 1. Use of textual organizational markers by groups.

The MANOVA analysis revealed a significant multivariate effect (see Table 4), Wilks' $\Lambda = 0.494$, $F(22, 98) = 1.882$, $p = 0.019$, partial $\eta^2 = 0.297$. The tests of between-subjects effects (see Table 5) showed that the percentages of conditional/hypothetical, frame, misuse, and engagement markers had significant differences among the groups. The frame markers, $F(2, 59) = 7.079$, $p = 0.002$, $\eta^2 = 0.194$, and misuse markers, $F(2, 59) = 5.079$, $p = 0.009$, $\eta^2 = 0.147$ displayed the highest significance level.

Table 4. Textual organizational measures multivariate test.

	Value	F	Df	Error df	p	Partial eta Squared
Pillai's trace	0.567	1.799 *	22	100	0.027	0.284
Wilks' Λ	0.494	1.882 *	22	98	0.019	0.297
Hotelling's trace	0.900	1.963 *	22	96	0.013	0.310
Roy's largest root	0.730	3.316 *	11	50	0.002	0.422

* $p < 0.05$.

Table 5. Textual organizational measures tests of between-subjects effects.

Source	Dependent Variable	Df	F	p	Partial eta Squared
Interactive Metadiscourse Markers					
Local cohesion between/within clauses/sentences:					
Groups	Continuative/additive marker	2	0.122	0.886	0.004
	Comparison/contrast marker	2	0.364	0.696	0.012
	Causative marker	2	1.021	0.367	0.033
	Conditional/hypothetical marker	2	3.150 *	0.050	0.096
	Misuse marker	2	5.079 *	0.009	0.147
	Preposition	2	0.015	0.985	0.001
Global cohesion across idea units:					
	Frame marker	2	7.079 *	0.002	0.194
Text cohesion:					
	Third-person pronoun	2	2.151	0.125	0.068
	Third-person pronoun/noun ratio	2	1.468	0.239	0.047
Interactional Metadiscourse Markers					

	Self-mention marker	2	0.468	0.628	0.016
	Engagement marker	2	4.995 *	0.010	0.145
Error	Continuative/additive marker	59			
	Comparison/contrast marker	59			
	Causative marker	59			
	Conditional/hypothetical marker	59			
	Misuse marker	59			
	Preposition	59			
	Frame marker	59			
	Third-person pronoun	59			
	Third-person pronoun/noun ratio	59			
	Self-mention marker	59			
	Engagement marker	59			

* $p < 0.05$.

The post-hoc analysis results with the Bonferroni correction showed that the middle-score and high-score groups produced significantly greater percentages of frame markers and lower percentages of misuse markers than the low-score group (see Table A1 in Appendix A). The high-score group also produced a significantly higher percentage of conditional/hypothetical markers than the low-score group. The middle-score and high-score groups produced a significantly higher ($p = 0.012$) or near-significantly ($p = 0.053$) higher percentage of engagement markers, respectively, than the low-score group. The differences in the other textual organizational measures were non-significant across the groups, including continuative/additive, comparison/contrast, causative, and self-mention markers, prepositions, third-person pronouns, and the third-person pronoun/noun ratio. Thus, compared to the low-proficiency writers, more proficient writers used organizational devices more accurately, applied a higher number of frame markers to signal topics, expressed conditional/hypothetical meaning more frequently, and engaged the reader more often.

4.2. Interrelations among Textual Organizational Features

Table 6 displays the interrelations among the textual organizational measures. Since causative marker was not significantly correlated with other organizational measures, it was not included for space limitations. The correlation analysis results demonstrated several interesting patterns.

First, several textual organizational measures revealed significant negative correlations with each other, suggesting that a decrease in particular organizational measures was accompanied by an increase in some other organizational measures. In particular, the percentage of continuative/additive markers correlated negatively with the percentages of both comparison/contrast markers and prepositions, indicating that the writers who used more comparison/contrast markers and prepositions tended to use fewer continuative/additive markers. The percentage of conditional/hypothetical markers correlated negatively with the percentage of self-mention markers. Thus, the writers who used more conditional/hypothetical markers reduced their use of first-person accounts in their writing. The percentage of frame markers correlated negatively with the percentage of misuse markers, implying that the writers who were better at signaling their topics tended to use organizational features more accurately. Interestingly, the two types of interactional metadiscourse indices—self-mention (e.g., I) and engagement (e.g., you) markers—correlated negatively with each other ($r = -0.566, p < 0.001$), indicating that as the writers became more skillful at engaging the reader, they reduced their use of first-person voice.

Second, positive correlations were displayed among several textual organizational measures. The percentage of third-person pronouns and the third-person pronoun/noun ratio correlated positively with frame marker ($r = 0.272, 0.257, p < 0.05$), indicating that the writers who employed more third-person pronouns (e.g., he/they/it) to describe their schools were also better at signaling their topics. In addition, the percentage of conditional/hypothetical markers correlated positively with the percentage of engagement markers ($r = 0.598, p < 0.001$), suggesting that the learners who used more conditional/hypothetical markers also better engaged their readers.

Last, not surprisingly, the percentage of third-person pronouns and the third-person pronoun/noun ratio correlated strongly with each other ($r = 0.962, p < 0.001$), suggesting that they may signal rather similar constructs. Using one of the two measures may satisfy the relevant analysis purposes.

Table 6. Correlations among textual organizational measures ($n = 62$).

	1	2	3	4	5	6	7	8	9
1. Continuative/additive marker	—								
2. Comparison/contrast marker	−0.402 *	—							
3. Conditional/hypothetical marker	−0.094	−0.038	—						
4. Frame marker	−0.034	−0.143	−0.005	—					
5. Misuse marker	−0.093	−0.207	−0.189	−0.336 *	—				
6. Preposition	−0.285 *	0.151	−0.067	−0.032	0.211	—			
7. Third-person pron.	−0.039	−0.186	0.028	0.272 *	0.046	0.054	—		
8. Third-person pron./noun ratio	0.012	−0.184	0.129	0.257 *	0.071	0.050	0.962 **	—	
9. Self-mention marker	0.091	0.100	−0.318 *	0.021	0.008	0.077	0.123	0.118	—
10. Engagement marker	−0.062	−0.112	0.598 **	0.165	−0.092	0.001	0.055	0.087	−0.566 **

** $p < 0.001$, * $p < 0.05$.

4.3. Interrelations between Textual Organizational and Linguistic Features

Before discussing the correlation results between the organizational and linguistic measures, the results for the linguistic measures are summarized to facilitate an understanding of the relationships between organizational and linguistic performances. The results show that the mean values of all three linguistic measures—ratio of correct clauses, lexical diversity, and clause length—consistently increased from the low-score to the high-score group (see Table A2 in Appendix B). The test of between-subjects effects revealed significant group differences for all three measures (see Table A3 in Appendix B). The results of the post-hoc analysis revealed that the high-score group produced a significantly higher ratio of correct clauses than the middle-score and low-score groups. The high-score group also produced significantly greater lexical diversity and clause length than the middle-score group, which also had greater values in both measures than the low-score group (see Table A4 in Appendix B).

Table 7 presents the correlation results between the textual organizational measures and the linguistic measures. The results demonstrate that the percentage of conditional/hypothetical markers correlated positively with lexical diversity, suggesting that the learners who used more conditional/hypothetical markers also applied more diversified lexis. The percentage of frame markers correlated positively with clause length, indicating that an ability to signal topics/subtopics was aligned with an ability to produce lengthier clauses in writing. The percentage of misused markers correlated negatively with the ratio of correct clauses and lexical diversity, suggesting that when the learners improved their accurate use of organizational features, their ability to use more accurate clauses and more diversified lexis also improved. The percentage of engagement markers correlated positively with lexical diversity and clause length. Thus, the writers who used more devices to engage the reader also produced more diversified lexis and lengthier clauses. In sum, the learners’ ability to use more diversified lexis in writing was positively associated with their ability to apply more accurate textual organizational devices, to use more conditional/hypothetical markers, and to better engage the reader. The learners’ ability to produce lengthier clauses also aligned well with their skills to signal topics/subtopics and apply devices to engage the reader.

Table 7. Correlations between organizational measures and linguistic measures ($n = 62$).

	Ratio of Correct Clauses	Lexical Diversity	Clause Length
Continuative/additive marker	0.026	0.018	-0.045
Comparison/contrast marker	0.098	0.084	-0.047
Causative marker	0.244	0.162	0.219
Conditional/hypothetical marker	0.007	0.263*	0.194
Misuse marker	-0.269*	-0.296*	-0.188
Preposition	-0.088	0.009	0.204
Frame marker	0.150	0.182	0.266*
Third-person pronoun	0.052	0.171	0.214
Third-person pronoun/noun ratio	0.025	0.116	0.143
Self-mention marker	0.201	-0.157	-0.053
Engagement marker	0.016	0.296*	0.337*

* $p < 0.05$.

5. Discussion

The current findings revealed differential textual organizational features for L2 writers at different proficiencies on local, global, and text levels. Multiple organizational features also display significant negative correlations with each other. Textual organizational features characteristic of advanced writers demonstrate some positive associations with linguistic performances.

5.1. RQ1: Textual Organizational Features in the Essays

For local cohesion, the findings demonstrate that the learners across levels frequently employed continuative/additive, comparison/contrast, and causative markers to signal transitions and establish cohesion between clauses/sentences. Thus, the learners seem to already possess the ability to deploy these transitional markers to shape their writing at an early stage of development. The higher-level writers showed a significantly stronger ability to use conditional/hypothetical markers in their texts. This finding contradicts those of previous studies of L2 English, in which conditional connectives negatively predicted the quality of EFL argumentative essays (Guo et al. 2013). The discrepancy in the findings may relate to the different genres examined in the two studies, i.e., argumentative essays in Guo et al. (2013) and descriptive essays in the current study. The use of conditional connectives may be more relevant and, therefore, more needed in the current descriptive writing task. The discrepancy may also be associated with the current analysis method of aggregating conditional and hypothetical markers into one analysis category. Additionally, the higher-proficiency writers used organizational devices more accurately than the lower-level writers, which corroborates L2 English findings that fourth-year Chinese-L1 college English learners used more accurate cohesive devices than second-year learners in argumentative writing (Yang and Sun 2012). One possible explanation is that low-proficiency writers need to focus more on low-level linguistic encoding, which may have taken away attentional resources that would otherwise be available for appropriate signaling of cohesion (Halliday and Hasan 1976; Kormos 2011; McCutchen 1996).

With respect to global cohesion, the middle-score and high-score writers have demonstrated a greater ability to use frame markers to signal relations between idea units, suggesting that advanced writers are more capable of connecting ideas on a higher textual level. This result corroborates the previous findings that English learners increased their use of global cohesive devices in academic writing over a semester-long course (Crossley et al. 2016a). Regarding text cohesion, although non-significant, the percentage of third-person pronouns and the third-person pronoun/noun ratio consistently increased from the low-score group to the high-score group, suggesting that more proficient writers are able to describe their schools beyond merely discussing their first-person experience. This finding contradicts previous L2 studies, which reported that less proficient writers used a significantly higher percentage of pronouns in comparison with native writers (Reid 1992) and that pronoun-to-noun ratio negatively predicted human judgments of essay quality (Crossley et al. 2016a). The reason for the divergent findings may lie in the fact that Reid (1992) and Crossley et al.

(2016a) counted all pronouns, whereas the current study only counted third-person pronouns to meet the needs of the analysis.

Concerning the use of interactional metadiscourse markers, the middle-score and high-score writers have demonstrated a stronger ability to use engagement markers to involve the reader. The differences in the number of self-mention markers, however, were non-significant among the groups, suggesting that low-level writers refer to themselves as frequently as more advanced writers while describing their institutions. This result differs from the findings in Lee and Deakin (2016) that Chinese-L1 English learners were overwhelmingly resistant to establishing an authorial identity in their argumentative essays. There are two possible reasons. First, Lee and Deakin (2016) examined Chinese-L1 English learners whose writing may have been influenced by the Chinese rhetorical tradition of preferring indirect authorial presence, whereas the current study analyzed English-L1 Chinese learners whose essays may have been impacted by the English rhetorical norms of advocating more direct authorial identity. Second, Lee and Deakin (2016) examined argumentative writing, whereas the current study investigated descriptive writing on a personalized topic. The latter may have naturally elicited a higher level of authorial presence to address the current topic. Thus, frequent use of self-mention markers in the current descriptive writing may not necessarily reflect better writing quality.

No statistical differences were found across the groups in several interactive and interactional organizational features, including the percentages of continuative/additive, comparison/contrast, causative, and self-mention markers, the percentages of prepositions and third-person pronouns, as well as the third-person pronoun/noun ratio. There may be two reasons for these insignificant differences. First, the learners may have learnt how to use continuative/additive, comparison/contrast, causative, and self-mention markers from an early stage of learning, which may have resulted in a similar number of usages of across the groups. Second, the topic used in the current task, i.e., introducing one's institution, may have allowed a limited context for applying prepositions and third-person pronouns and the resulted low frequency of usages (e.g., 1.49% third-person pronouns, 3.29% prepositions in the low-score group) may have weakened the statistical power to detect differences between the groups.

5.2. Interrelations among the Textual Organizational Features

Regarding the interrelations among the textual organizational features, both positive and negative correlations were identified. Both of the text-level cohesive markers—percentage of third-person pronouns and the third-person pronoun/noun ratio—correlate positively with the percentage of frame markers. Thus, learners' ability to provide discussions beyond first-person accounts is positively associated with their ability to signal change of topics in writing. Given the current finding that the high-score group uses significantly more frame markers than the low-score group, more frequent use of third-person pronouns to refer to given information is likely to result in stronger cohesion.

More interesting interrelation findings lie in that multiple pairs of organizational features demonstrate significantly negative correlations, implying connected decreases and increases in specific textual organizational features. In particular, significant negative correlations were found in the following pairs of measures: (a) continuative/additive and comparison markers; (b) continuative/additive markers and preposition; (c) conditional/hypothetical and self-mention markers; (d) frame and misuse markers; and (e) self-mention and engagement markers. Thus, a higher use of continuative/additive markers is accompanied by a lower use of comparison markers and prepositions. A more frequent use of conditional/hypothetical markers is associated with a reduced use of first-person discussions. When learners become more adept at signaling their topics/sub-topics, their misuse of organizational features also declines. The negative relation between self-mention and engagement markers seems to be somewhat intuitive. When learners downplay first-person experiences, they become more aware of involving their readers. Given that the high-score group uses significantly more conditional/hypothetical, frame, and engagement markers than

the low-score group, we may infer that their corresponding negative correlators—use of self-mention and misuse markers—may be characteristics of low-proficiency writers.

Combining the findings of RQs 1 and 2, we can see that in the current descriptive writing task, compared to low-proficiency writers, more advanced writers use organizational features more accurately, apply more conditional/hypothetical transitional markers, provide more third-person discussions, signal their topics/subtopics more effectively, and engage the reader more actively.

5.3. Interrelations between Textual Organizational Features and Linguistic Features

The analysis shows that the organizational features characteristic of more advanced writers, including the use of conditional/hypothetical, frame, and engagement markers, third-person accounts, and accuracy of organizational features, display positive relationships with the linguistic measures. Specifically, the conditional/hypothetical marker correlates positively with lexical diversity; the frame marker correlates positively with clause length; and the engagement marker correlates positively with both lexical diversity and clause length. Misuse marker correlates negatively with clause accuracy and lexical diversity, suggesting a connected growth between the ability to control the accuracy of organizational features and the ability to produce accurate clauses and use diversified lexis in writing.

We can see that learners' ability to apply diversified lexis in writing, an indicator of lexical complexity, is positively associated with multiple textual organizational features: accurate use of metadiscourse devices, application of devices to engage the reader, and use of conditional/hypothetical markers. These findings suggest that learners' effective use of metadiscourse devices and conditional/hypothetical markers in particular may relate to their lexical skills in the complexity dimension. Learners' ability to produce lengthier clauses, an indicator of syntactic complexity, aligns well with their metadiscourse skills in framing new topics and engaging the reader. These results indicate that as learners become more capable of developing complex clauses, they are also more skillful at signaling topic shifts and involving the reader, thus better guiding the reader's interpretations of the text towards their preferred ones (Hyland 2005). In contrast, although the high-score learners produced significantly higher ratio of correct clauses than the other two groups, clause accuracy has non-significant correlations with all metadiscourse features, except for misuse marker. This finding suggests that linguistic accuracy develops somewhat independently from the development of written metadiscourse skills.

Although only a few linguistic measures have been analyzed in the current study, the findings have provided useful knowledge regarding the connections, or lack thereof, between L2 textual organizational skills and linguistic skills. For example, the findings demonstrate that strong lexical skills are connected with effective skills to establish writer–reader interactions. The development of linguistic accuracy, however, lacks a clear connection with the development of meta-discourse skills.

6. Implications

The current study adds knowledge to our understanding of how L2 writers at different proficiencies employ metadiscourse features to shape their written discourses, as well as how various textual organizational performances relate to each other and correlate with linguistic performances.

This study has limitations that should be taken into consideration in future research in this area. The first limitation arises from the writing task used in the current study, i.e., descriptive writing with a single topic, which may have provided a limited context for applying certain metadiscourse features. For instance, the use of propositions and third-person pronouns is limited across the groups. Researchers may consider investigating whether other genres or topics may generate distinct outcomes with respect to these features. This line of research will deepen an understanding of the effects of genres and topics on textual organizations. The current findings also prove that several interactive and interactional metadiscourse markers successfully distinguish writing proficiencies. Multiple negative correlations also exist among various organizational features. These discriminative and correlational patterns will deserve additional investigations in different research contexts, such as different types of genres or other L2s. Second, the current study examined timed handwritten essays, a research condition that may have affected the composing process. Future research may explore whether type-written essays demonstrate differential textual organizational performances than handwritten ones. Finally, the current study examined only finished written products. Future studies may focus on process-oriented research to document the process of writing from beginning to completion. This research focus will help us know better the micro- and macro-level mechanisms that L2 writers go through to organize a text.

The findings also inform L2 writing pedagogy. First, they show that the development of L2 metadiscourse skills may follow specific complex patterns and may need to be nurtured in its own right. For instance, the results demonstrate that learners increase and decrease their use of particular organizational features with increased proficiency. Language instructors could consider providing more explicit guidance regarding how more or less use of specific metadiscourse features may boost coherence and organizational quality. Second, the current findings indicate that low-proficiency writers may not possess a capacity to engage the reader effectively. Language instructors may want to provide clear instructions to students from an early stage of learning on how to compose with an audience in mind. Third, the low-proficiency writers used few conditional/hypothetical and frame markers. Given that L2 learners may have learnt the linguistic items related to conditional/hypothetical markers or frame markers at lower levels of instruction, language instructors may consider designing writing activities to guide learners to practice using a variety of logical operators to express logical meaning more effectively.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Post-Hoc Analysis Results for Organizational Measures

Table A1. Multiple comparisons with Bonferroni correction.

	(I) Groups	(J) Groups	Mean Difference (I–J)	Std. Error	<i>p</i>	95% CI	
						Lower Bound	Upper Bound
Continuative/ additive marker	low	mid	0.0002	0.0620	1.000	−0.1529	0.1525
		high	0.0250	0.0600	1.000	−0.1228	0.1727
	mid	low	0.0002	0.0620	1.000	−0.1525	0.1529
		high	0.0252	0.0591	1.000	−0.1205	0.1709
	high	low	−0.0250	0.0600	1.000	−0.1727	0.1228
	mid	−0.0252	0.0591	1.000	−0.1709	0.1205	
Comparison/ contrast marker	low	mid	−0.0602	0.0730	1.000	−0.2401	0.1196
		high	−0.0179	0.0706	1.000	−0.1919	0.1562
	mid	low	0.0602	0.0730	1.000	−0.1196	0.2401
		high	0.0424	0.0697	1.000	−0.1293	0.2141
	high	low	0.0179	0.0706	1.000	−0.1562	0.1919
	mid	−0.0424	0.0697	1.000	−0.2141	0.1293	
Causative marker	low	mid	0.0125	0.0665	1.000	−0.1514	0.1763
		high	−0.0709	0.0643	0.825	−0.2294	0.0877
	mid	low	−0.0125	0.0665	1.000	−0.1763	0.1514
		high	−0.0833	0.0635	0.582	−0.2397	0.0730
	high	low	0.0709	0.0643	0.825	−0.0877	0.2294
	mid	0.0833	0.0635	0.582	−0.0730	0.2397	
Conditional/ hypothetical marker	low	mid	−0.0440	0.0257	0.278	−0.1073	0.0194
		high	−0.0613 *	0.0249	0.050	−0.1227	−0.00001
	mid	low	0.0440	0.0257	0.278	−0.0194	0.1073
		high	−0.0174	0.0245	1.000	−0.0778	0.0431
	high	low	0.0613 *	0.0249	0.050	0.00001	0.1227
	mid	0.0174	0.0245	1.000	−0.0431	0.0778	
Misuse marker	low	mid	0.1683 *	0.0641	0.033	0.0104	0.3262
		high	0.1812 *	0.0620	0.015	0.0285	0.3340
	mid	low	−0.1683 *	0.0641	0.033	−0.3262	−0.0104
		high	0.0130	0.0612	1.000	−0.1377	0.1637
	high	low	−0.1812 *	0.0620	0.015	−0.3340	−0.0285
	mid	−0.0130	0.0612	1.000	−0.1637	0.1377	
Preposition	low	mid	0.0009	0.0052	1.000	−0.0120	0.0137
		high	0.0002	0.0050	1.000	−0.0122	0.0126
	mid	low	−0.0009	0.0052	1.000	−0.0137	0.0120
		high	−0.0007	0.0050	1.000	−0.0129	0.0116
	high	low	−0.0002	0.0050	1.000	−0.0126	0.0122
	mid	0.0007	0.0050	1.000	−0.0116	0.0129	
Frame marker	low	mid	−0.1816 *	0.0531	0.003	−0.3123	−0.0508
		high	−0.1614 *	0.0514	0.008	−0.2880	−0.0348
	mid	low	0.1816 *	0.0531	0.003	0.0508	0.3123
		high	0.0202	0.0507	1.000	−0.1047	0.1450
	high	low	0.1614 *	0.0514	0.008	0.0348	0.2880
	mid	−0.0202	0.0507	1.000	−0.1450	0.1047	
Third-person pronoun	low	mid	−0.0007	0.0040	1.000	−0.0107	0.0093
		high	−0.0072	0.0039	0.210	−0.0169	0.0024
	mid	low	0.0007	0.0040	1.000	−0.0093	0.0107
		high	−0.0065	0.0039	0.289	−0.0160	0.0030
	high	low	0.0072	0.0039	0.210	−0.0024	0.0169
	mid	0.0065	0.0039	0.289	−0.0030	0.0160	
Third-person pronoun/noun ratio	low	mid	−0.0028	0.0154	1.000	−0.0408	0.0353
		high	−0.0230	0.0149	0.387	−0.0598	0.0138
	mid	low	0.0028	0.0154	1.000	−0.0353	0.0408
		high	−0.0202	0.0147	0.524	−0.0566	0.0161
	high	low	0.0230	0.0149	0.387	−0.0138	0.0598
	mid	0.0202	0.0147	0.524	−0.0161	0.0566	
Self-mention marker	low	mid	0.0649	0.0831	1.000	−0.1400	0.2697
		high	0.0723	0.0804	1.000	−0.1259	0.2706
	mid	low	−0.065	0.0831	1.000	−0.2697	0.1400

	high	high	0.0075	0.0793	1.000	-0.1880	0.2030
	high	low	-0.0723	0.0804	1.000	-0.2706	0.1259
		mid	-0.0075	0.0793	1.000	-0.2030	0.1880
Engagement marker	low	mid	-0.1701 *	0.0568	0.012	-0.3100	-0.0302
		high	-0.1341	0.0549	0.053	-0.2695	0.0013
	mid	low	0.1701 *	0.0568	0.012	0.0302	0.3100
		high	0.0360	0.0542	1.000	-0.0975	0.1695
	high	low	0.1341	0.0549	0.053	-0.0013	0.2695
		mid	-0.0360	0.0542	1.000	-0.1695	0.0975

* $p < 0.05$.

Appendix B. Post-Hoc Analysis Results for Linguistic Measures

Table A2. Linguistic measures descriptive statistics.

	Low ($n = 19$) <i>M (SD)</i>	Middle ($n = 20$) <i>M (SD)</i>	High ($n = 23$) <i>M (SD)</i>
Ratio of correct clauses	0.5209 (0.2397)	0.5510 (0.1196)	0.6979 (0.0852)
Lexical diversity	5.02 (0.90)	6.41 (0.89)	7.06 (0.79)
Clause length	5.53 (0.89)	6.51 (0.86)	7.73 (0.83)

Table A3. Linguistic measures tests of between-subjects effects.

Source	Dependent variable	<i>Df</i>	<i>F</i>	<i>p</i>
Groups	Ratio of correct clauses	2	7.779 *	0.001
	Lexical diversity	2	30.13 **	<0.001
	Clause length	2	34.685 **	<0.001
Error	Ratio of correct clauses	59		
	Lexical diversity	59		
	Clause length	59		

** $p < 0.001$, * $p < 0.05$.

Table A4. Multiple comparisons with Bonferroni correction.

	(I) Groups	(J) Groups	Mean Difference (I–J)	Std. Error	<i>p</i>	95% CI	
						Lower Bound	Upper Bound
Ratio of correct clauses	low	mid	-0.0301	0.0505	1.000	-0.1546	0.0943
		high	-0.1770 *	0.0489	0.002	-0.2974	-0.0566
	mid	low	0.0301	0.0505	1.000	-0.0943	0.1546
		high	-0.1469 *	0.0482	0.010	-0.2657	-0.0281
	high	low	0.1770 *	0.0489	0.002	0.0566	0.2974
		mid	0.1469 *	0.0482	0.010	0.0281	0.2657
Lexical diversity	low	mid	-1.3891 **	0.2747	<0.001	-2.0661	-0.7121
		high	-2.0405 **	0.2659	<0.001	-2.6956	-1.3853
	mid	low	1.3891 **	0.2747	<0.001	0.7121	2.0661
		high	-0.6513 *	0.2622	0.048	-1.2974	-0.0052
	high	low	2.0405 **	0.2659	<0.001	1.3853	2.6956
		mid	0.6513 *	0.2622	0.048	0.0052	1.2974
Clause length	low	mid	-0.9734 *	0.2744	0.002	-1.6496	-0.2972
		high	-2.1962 **	0.2656	<0.001	-2.8506	-1.5419
	mid	low	0.9734 *	0.2744	0.002	0.2972	1.6496
		high	-1.2228 **	0.2619	<0.001	-1.8682	-0.5775
	high	low	2.1962 **	0.2656	<0.001	1.5419	2.8506
		mid	1.2228 **	0.2619	<0.001	0.5775	1.8682

** $p < 0.001$, * $p < 0.05$.

Appendix C. Rating Scale for the Essays

Table A5. Rating scale for the essays.

Scores	Proficiency Level	Score Criteria
1	Lower-Beginning	Limited content is presented. The meaning is difficult to understand. Limited formulaic language, such as familiar words or phrases, may be used. No discernible writing structure can be identified.
2	Higher-Beginning	Undeveloped content is presented. The meaning is generally comprehensible, but gaps in comprehension occurs. Formulaic language, such as familiar words or phrases, may be used. A very basic and undeveloped writing structure is available.
3	Lower-Intermediate	Simple and unsophisticated content is presented. A basic writing structure is available, but it lacks effective cohesion and coherence. The writing style resembles oral discourse and the writing communicates limited information to the audience.
4	Higher-Intermediate	Some variety of ideas is presented, but is often unsophisticated. A basic writing structure is available with some coherence and cohesion. The writing style resembles oral discourse and the writing communicates some basic information to the audience.
5	Lower-Advanced	A good variety of ideas is presented with some elaboration. An organized writing structure is presented with good coherence and cohesion. An introduction, elaboration, and conclusion on the topic are often presented. The writing communicates clear information to the audience.
6	Higher-Advanced	A good variety of well-developed ideas is presented. A clear and organized writing structure is evident with effective coherence and cohesion. An effective introduction, elaboration, and conclusion on the topic are presented. The writing communicates very clear information to the audience.

References

- ACTFL. 2012. ACTFL Proficiency Guidelines. Available online: http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf (accessed on 2 January 2020).
- Bardovi-Harlig, Kathleen. 1990. Pragmatic word order in English composition. In *Coherence in Writing: Research and Pedagogical Perspectives*. Edited by Ulla Connor and Ann M. Johns. Alexandria: TESOL, pp. 43–65.
- Chen, Yu-Hua, and Paul Baker. 2016. Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics* 37: 849–80. doi:10.1093/applin/amu065.
- Chiang, Steve. 2003. The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System* 31: 471–84. doi:10.1016/j.system.2003.02.002.
- Connor, Ulla. 1990. Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English* 24: 67–87.
- Crossley, Scott A., and Danielle S. McNamara. 2012. Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading* 35: 115–35. doi:10.1111/j.1467-9817.2010.01449.x.
- Crossley, Scott A., Rod Roscoe, and Danielle S. McNamara. 2011. Predicting human scores of essay quality using computational indices of linguistic and textual features. In *Artificial Intelligence in Education: AIED 2011*. Edited by Gautam Biswas, Susan Bull, Judy Kay and Antonija Mitrovic. Berlin and Heidelberg: Springer, pp. 438–40.
- Crossley, Scott A., Kristopher Kyle, and Danielle S. McNamara. 2016a. The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing* 32: 1–16. doi:10.1016/j.jslw.2016.01.003.
- Crossley, Scott A., Kristopher Kyle, and Danielle S. McNamara. 2016b. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods* 48: 1227–37. doi:10.3758/s13428-015-0651-7.
- Evola, Jill, Ellen Mamer, and Becky Lentz. 1980. Discrete point versus global scoring for cohesive devices. In

- Research in Language Testing*. Edited by John W. Oller and Kyle Perkins. Rowley: Newbury House, pp. 177–81.
- Ferris, Dana R. 1994. Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly* 28: 414–20. doi:10.2307/3587446.
- Graesser, Arthur C., Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 36: 193–202. doi:10.3758/BF03195564.
- Guo, Liang, Scott A. Crossley, and Danielle S. McNamara. 2013. Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing* 18: 218–38. doi:10.1016/j.asw.2013.05.002.
- Halliday, Michael Alexander Kirkwood, and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Harman, Ruth. 2013. Literary intertextuality in genre-based pedagogies: Building lexical cohesion in fifth-grade L2 writing. *Journal of Second Language Writing* 22: 125–40. doi:10.1016/j.jslw.2013.03.006.
- Hu, Guangwei, and Feng Cao. 2011. Hedging and boosting in abstracts of applied linguistics articles: A comparative study on English- and Chinese-medium journals. *Journal of Pragmatics* 43: 2795–809.
- Hyland, Ken. 2005. *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Jafarpur, Abdoljavad. 1991. Cohesiveness as a basis for evaluating compositions. *System* 19: 459–65. doi:10.1016/0346-251X(91)90026-L.
- Kennedy, Chris, and Dilys Thorp. 2007. A corpus-based investigation of linguistic responses to an IELTS academic writing task. In *IELTS Collected Papers: Research in Speaking and Writing Assessment*. Edited by Lynda Taylor and Peter Falvey. Cambridge: Cambridge University Press, pp. 316–77.
- Kintsch, Walter 1998. *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press.
- Kormos, Judit. 2011. Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing* 20: 148–61. doi:10.1016/j.jslw.2011.02.001.
- Kyle, Kristopher, and Scott Crossley. 2017. Tool for the automatic analysis of cohesion. Available online: https://www.linguisticanalysisistools.org/uploads/1/3/9/3/13935189/_taaco_1.5_user_manual_7-16-17.pdf (accessed on 5 January 2020).
- Lee, Joseph J., and Lydia Deakin. 2016. Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing* 33: 21–34. doi:10.1016/j.jslw.2016.06.004.
- Li, Shouji. 2014. The gap in the use of lexical cohesive devices in writing between native Chinese speakers and second language users. *Journal of the Chinese Language Teachers Association* 49: 25–47.
- Li, Ting, and Sue Wharton. 2012. Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes* 11: 345–56.
- Liu, Meihua, and George Braine. 2005. Cohesive features in argumentative writing produced by Chinese undergraduates. *System* 33: 623–36. doi:10.1016/j.system.2005.02.002.
- Louwerse, Max. 2002. An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics* 12: 291–315. doi:10.1515/cogl.2002.005.
- Mazgutova, Diana, and Judit Kormos. 2015. Syntactic and lexical development in an intensive English for academic purposes programme. *Journal of Second Language Writing* 29: 3–15. doi:10.1016/j.jslw.2015.06.004.
- McCutchen, Deborah. 1996. A capacity theory of writing: Working memory in composition. *Educational Psychology Review* 8: 299–325. doi:10.1007/BF01464076.
- McNamara, Danielle S., Eileen Kintsch, Nancy Butler Songer, and Walter Kintsch. 1996. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction* 14: 1–43. doi:10.1207/s1532690xci1401_1.
- McNamara, Danielle S., Scott A. Crossley, and Philip M. McCarthy. 2010. Linguistic features of writing quality. *Written Communication* 27: 57–86. doi:10.1177/0741088309351547.
- McNamara, Danielle S., Scott A. Crossley, and Rod Roscoe. 2013. Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods* 45: 499–515. doi:10.3758/s13428-012-0258-1.
- Neuner, Jerome L. 1987. Cohesive ties and chains in good and poor freshman essays. *Research in the Teaching of English* 21: 92–105.
- Polio, Charlene G. 1997. Measures of linguistic accuracy in second language writing research. *Language Learning* 47: 101–43.
- Polio, Charlene, and Mark C. Shea. 2014. An investigation into current measures of linguistic accuracy in second

- language writing research. *Journal of Second Language Writing* 26: 10–27.
- Reid, Joy. 1992. A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing* 1: 79–107. doi:10.1016/1060-3743(92)90010-M.
- Sanders, Ted J.M., Wilbert P.M. Spooren, and Leo G.M. Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15: 1–35.
- Smith, Raoul N., and William J. Frawley. 1983. Conjunctive cohesion in four English genres. *Text* 3: 347–74. doi:10.1515/text.1.1983.3.4.347.
- Van Dijk, Teun Adrianus, and Walter Kintsch. 1983. *Strategies of Discourse Comprehension*. New York: Academic Press.
- Yang, Chunsheng. 2013. Textual conjunctives and topic-fronting devices in CFL learners' written summaries. *Journal of the Chinese Language Teachers Association* 48: 71–89.
- Yang, Wenxing, and Ying Sun. 2012. The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education* 23: 31–48. doi:10.1016/j.linged.2011.09.004.
- Zhao, Cecilia Guanfang. 2013. Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing* 30: 201–30. doi:10.1177/0265532212456965.
- Zwaan, Rolf A., and Gabriel A. Radvansky. 1998. Situation models in language comprehension and memory. *Psychological Bulletin* 123: 162–85. doi:10.1037/0033-2909.123.2.162.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).