How do textual features of L2 argumentative essays differ across proficiency levels? A multidimensional cross-sectional study

Jeong-eun Kim[1] · **Hosung Nam**[2,3]

Abstract

Using Biber's (1988) multidimensional analysis, this study investigates textual variation in second language (L2) learners' writing at different proficiency levels, and attempts to identify any developmental progression. The study used a corpus of 5200 argumentative essays written by 2600 students learning English as an L2. The results indicate that advanced L2 writing is fundamentally different from less advanced L2 writing: Advanced learners' writing is closer to native speakers' written discourse, while less advanced learners' writing is closer to native speakers' spoken discourse. The patterns of development vary across different sets of textual features. *Informational* (as opposed to *involved*) production and *impersonal* (as opposed to *nonimpersonal*) style showed gradual development as the learners' proficiency increases. *Nonnarrative* (as opposed to *narrative*) production, *elaborated* (as opposed to *situation-dependent*) *reference*, and *overt expression of persuasion* did not show significant differences across the proficiency levels. The article offers pedagogical implications for practices of L2 writing instruction.

Keywords Corpus linguistics · Cross-sectional research · Mixed-effects model · Multidimensional analysis · Argumentative writing · Second language writing

✉ Jeong-eun Kim
  jek48@jbnu.ac.kr

✉ Hosung Nam
  hnam@korea.ac.kr

1   Department of English Language and Literature, Chonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do 54896, Republic of Korea

2   Department of English Language and Literature, Korea University, 145 Anam-ro, Sungbuk-gu, Seoul 02841, Republic of Korea

3   Haskins Laboratories, 300 George St. #900, New Haven, CT 06511, USA

✆ Springer

Introduction

Acquiring native-like competence in second language (L2) writing is perhaps the greatest challenge faced by L2 learners. Much comparative and quantitative research in the past two decades has endeavored to discover linguistic features and characteristics of proficient L2 writing (e.g., Ferris, 1994; Frase, Faletti, Ginther, & Grant, 1999; Grant & Ginther, 2000; Jarvis, Grant, Bikowski, & Ferris, 2003; Lu, 2011; Ortega, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998). These studies commonly employ linguistic complexity (e.g., length of production units, lexical complexity, syntactic complexity) as a predictor for proficient L2 writing, but their results are sometimes inconsistent and/or inconclusive (e.g., Becker, 2010; Biber, Gray, & Poonpon, 2011; Ortega, 2003; Taguchi, Crawford, & Wetzel, 2013; Wolfe-Quintero et al., 1998). Biber (1988) argued that such measures do not necessarily represent developmental progression in L2 writing ability, and proposed that a function-based multidimensional analysis using interactions among various co-occurring linguistic features would be more useful. Unlike the previous approaches, Biber's multidimensional analysis does not restrict the focus of analysis to discrete occurrences of certain linguistic items. Rather, it focuses on the linguistic functions accomplished by the interactions among co-occurring linguistic items. This function-based multidimensional analysis may be better able to effectively capture distinct textual features of L2 writing by learners of different proficiency levels.

Using Biber's (1988) multidimensional analysis, the present research investigated distinctive textual features of L2 writing by learners at different proficiency levels, and identified any developmental progression evidenced by function-based sets of co-occurring linguistic features. The findings of the study have pedagogical implications for current practices of L2 writing instruction.

Literature review

Multidimensional textual analysis

Biber (1988) claimed that a bundle of linguistic features (e.g., private verbs, *that*-deletions, contractions, present tense verbs, etc.) that co-occur in a text can distinguish text types (genre or register) more effectively than individual features. He introduced multidimensional analysis, which employs factor analysis to identify the co-occurrence patterns of linguistic features. For example, a text with frequent co-occurrence of such linguistic features as infinitives, prediction modals, suasive verbs, conditional subordinations, necessity modals, and split auxiliaries can be interpreted as overtly persuasive. In other words, the six linguistic features, as a set, serve some common function in the text: If a text has a high frequency of these linguistic features, the text is likely to be more *overtly persuasive*; if, on the other hand, a text has a low frequency of the features, the text is

likely to be less *overtly persuasive*. Biber identified five such functionally defined sets, which he called "dimensions": Dimension 1, *involved versus informational production*; Dimension 2, *narrative versus nonnarrative discourse*; Dimension 3, *situation-dependent versus elaborated reference*; Dimension 4, *overt expression of persuasion*; and Dimension 5, *nonimpersonal versus impersonal style*. His multidimensional analysis involves computing a dimension score for each of the five, and using the scores to distinguish text types.

Table 1 provides a summary of Biber's (1988) five dimensions and factor analysis; the table also draws on information in Biber, Conrad, Reppen, Byrd, and Helt's (2002) study.

Dimension 1, *involved versus informational production*, distinguishes involved, noninformational texts with interactive and affective purposes from less involved, informational texts with precise content. Biber (1988) found that the positive linguistic features (e.g., private verbs, *that*-deletion, contractions) are more likely to co-occur in spoken texts (e.g., face-to-face conversation, spontaneous/prepared speeches), and thus, had higher mean scores (involved). The negative linguistic features (e.g., nouns, prepositions, attributive adjectives), on the other hand, are more likely to co-occur in written texts (e.g., academic prose, official documents, institutional editorials), and thus had lower mean scores (informational). These results indicate that spoken texts share the textual features of being interactive, affective, and involved while written texts are more densely informational and deliver exact informational content.

Dimension 2, *narrative versus nonnarrative production*, distinguishes between active, depictive, event-oriented texts and more static descriptive or expository texts. Biber (1988) found that the positive features (e.g., past tense verbs, third-person pronouns, perfect aspect verbs) are more likely to co-occur in spoken texts (narrative) than in written texts (nonnarrative), resulting in relatively higher mean scores for the spoken texts. For example, past tense verbs, perfect aspect verbs, and third-person pronouns are frequently used in sequential descriptions of past events involving specific animate participants, while public verbs are used for indirect, reported speech. While negative features were not defined for this dimension, the written texts scored lower because they rarely contain the positive features.

Dimension 3, *elaborated versus situation-dependent reference*, distinguishes between highly elaborated, context-independent reference and nonspecific, situation-dependent reference. For example, texts that are highly explicit and endophoric (i.e., involving text-internal reference) are characterized as situation-independent, while those that are exophoric (i.e., involving extensive reference to the physical and temporal situation) are considered situation-dependent. According to Biber's (1988) factor analysis, the positive features (e.g., time/place adverbials, adverbs) but not the negative features (e.g., *wh*-relative clauses in object/subject positions, phrasal coordination) of Dimension 3 tend to occur in elaborated texts. The opposite pattern was found in situation-dependent texts. A clear distinction was observed between written and spoken texts in Dimension 3, with higher mean scores for written and lower mean scores for spoken texts. This indicates that written texts are more likely to specify the identity of the referents within the text, whereas the referents in spoken texts are more likely to be found in the physical context of the discourse.

Table 1 Biber's multidimensional analysis (Biber, 1988; Biber et al., 2002, pp. 15–16)

Features	Examples	Factor loading
Dimension 1: involved versus informational production		
Positive features (involved production)		
Private verbs	think, know, believe	0.96
That-deletions	I think [0] he went	0.91
Contractions	can't, she's	0.90
Present tense verbs	is, likes, wants	0.86
Second-person pronouns	you	0.86
Do as pro-verb	so *did* Sandra	0.82
Analytic negation	that's *not* likely	0.78
Demonstrative pronouns	this shows…	0.76
General emphatics	really, a lot	0.74
First-person pronouns	I, we	0.74
Pronoun *it*	I didn't like *it*.	0.71
Be as main verb	that *was* sad	0.71
Causative subordination	because…	0.66
Discourse particles	well, anyway	0.66
Indefinite hedges	nothing, someone	0.62
General hedges	kind of, something like	0.58
Amplifiers	absolutely, extremely	0.56
Sentence relatives	Bob didn't study at all, *which is usual for him*	0.55
Wh-questions	Why did you go?	0.52
Possibility modals	can, could, may, might	0.50
Nonphrasal coordination	Sally was biking last weekend *and* then she…	0.48
Wh-clauses	Jill asked *what happened.*	0.47
Final preposition	the candidate that I was thinking *of*	0.43

Table 1 (continued)

Features	Examples	Factor loading
Negative features (informational production)		
Nouns	community, case	−0.80
Word length	–	−0.58
Prepositions	of, in, for	−0.54
Type/token ratio	–	−0.54
Attributive adjectives	good, possible	−0.47
Dimension 2: narrative versus nonnarrative discourse		
Positive features (narrative discourse)		
Past tense verbs	considered, described	0.90
Third-person pronouns	he, she, they	0.73
Perfect aspect verbs	had been, has shown	0.48
Public verbs	said, explain	0.43
Synthetic negation	no answer is good enough	0.40
Present participial clauses	Having established the direction, we can now…	0.39
Dimension 3: situation-dependent versus elaborated reference		
Positive features (elaborated reference)		
Wh-relative clauses in object positions	something *which everybody can do*	0.63
Pied piping constructions	the way *in which this happens*	0.61
Wh-relative clauses in subject positions	those *who retain inhibitions*	0.45
Phrasal coordination	salt and pepper	0.36
Nominalization	extension, proposition	0.36
Negative features (situation-dependent reference)		
Time adverbials	early, instantly, soon	−0.60
Place adverbials	above, beside, outdoors	−0.49

Table 1 (continued)

Features	Examples	Factor loading
Adverbs	always, significantly	−0.46
Dimension 4: overt expression of persuasion		
Positive features (overt expression of persuasion)		
Infinitives	hope to go	0.76
Prediction modals	will, would, shall	0.54
Suasive verbs	command, insist, propose	0.49
Conditional subordination	if you want	0.47
Necessity modals	must, should, have to	0.46
Split auxiliaries	should *really* be	0.44
Dimension 5: nonimpersonal versus impersonal style		
Positive features (impersonal style)		
Conjuncts	however, therefore	0.48
Agentless passives	The same mechanism *was analyzed* on each.	0.43
Past participial adverbial clauses	Directed by Twilling, the production is delightful	0.42
By passives	He *was surrounded by a ring of men.*	0.41
Past participial postnominal clauses	the course *chosen by the large majority*	0.40
Other adverbial subordination	since, while, whereas	0.39

Dimension 4, *overt expression of persuasion*, marks "the degree to which persuasion is marked overtly, whether overt marking of the speaker's own point of view, or an assessment of the advisability or likelihood of an event presented to persuade the addressee" (Biber, 1988, p. 111). The positive features (e.g., infinitives, prediction/necessity modals, suasive verbs) are commonly used to express the speaker's point of view or encode the speakers' attitude or stance towards certain propositions. For example, professional letters and editorials showed higher scores on this dimension whereas broadcasts and press reviews showed lower scores. According to Biber's interpretation, both professional letters and editorials are typical argumentative texts intended to persuade the reader, and thus the nature of the texts leads to a dense use of the positive features. Broadcasts and press reviews, on the other hand, are not intended to persuade, but directly report an event or present a person's opinion. Thus, due to their nature, these two text types scored low on Dimension 4. Unlike the previous dimensions, Dimension 4 does not demonstrate a patterned distinction between spoken versus written texts; rather, it distinguishes between persuasive and nonpersuasive texts.

Dimension 5, *impersonal versus nonimpersonal style*, marks informational discourse that is abstract, technical, and formal versus other types of discourse. Biber (1988) did not define negative features, but identified positive features (e.g., conjuncts, agentless passives, *by*-passives, past participle adverbial clauses) with impersonal style. The analysis showed higher scores for written texts and lower scores for spoken texts. This indicates that the written texts were highly informational, with abstract, conceptual, or technical topics, while the less impersonal spoken texts dealt with active, human participants and concrete topics.

Multidimensional textual analysis in L2 writing research

Biber, Gray, and Staples (2016) extended Biber's (1988) multidimensional analysis to L2 corpora, and found supportive evidence that text analyses based on linguistic co-occurrence patterns, represented by dimensions, offer more insightful descriptions and robust measures than analyses based on discrete linguistic features. Learners' spoken and written responses were obtained from the Test of English as a Foreign Language (TOEFL; the study used 2879 spoken responses and 960 written responses in total). For each response, information on task type (written vs. spoken) and learner's proficiency (high to low) was available. The analysis found strong linguistic differences between written and spoken texts. As for proficiency, however, the analysis did not find systematic linguistic differences, implying that proficiency is not a significant predictor of the variation described by individual linguistic elements. On the other hand, a multidimensional analysis of the co-occurring lexico-grammatical features found both task type and proficiency to be significant predictors. The results showed that the linguistic features frequently used in native speakers' written discourse appeared more in all written responses and in more proficient learners' responses (written: e.g., nouns, prepositional phrases, adjectives, word length, passives), while the features frequently occurring in native speakers' oral discourse were found more in all oral responses and in less proficient learners'

responses (oral: e.g., verbs, third-person pronouns, *that*-clauses, finite adverbial clauses). These results, therefore, illustrate that multidimensional analysis can serve as a productive alternative to traditional measures to better capture the developmental progression of L2 writing.

Indeed, several researchers have recently extended multidimensional analysis to explore L2 writing developmental progression (Biber & Gray, 2013; Biber et al., 2016; Crosthwaite, 2016; Friginal & Hardy, 2014; Friginal & Weigle, 2014; Staples, Biber, & Reppen, 2018; Weigle & Friginal, 2015). Unlike the previous studies focusing on the occurrence patterns of discrete linguistic elements, studies using the multidimensional approach assess patterns of linguistic-textual variation across time and/or proficiency groups based on a full set of linguistic complexity features. For example, Friginal and Weigle (2014) compared 207 L2 (English) college students' academic writing across proficiency groups (Low, Mid, and High) and timing of development (January, March, and April in a semester). Their factor analysis of the participants' writing production extracted four functional dimensions: (1) involved versus informational focus; (2) addressee-focused description versus personal narrative; (3) simplified versus elaborated description; and (4) personal opinion versus impersonal evaluation (for full descriptions of the dimensions, see Friginal & Weigle, 2014, pp. 85–90). The descriptive analysis of the mean dimension scores demonstrated that high-rated writing was more informational (Dimension 1), more personally narrative (Dimension 2), more elaborated (Dimension 3), and more impersonal (Dimension 4) than low-rated writing. The mean dimension scores showed the same trends increasing over the study period, indicating development from January to April.

In another meaningful study of L2 academic writing, Crosthwaite (2016) employed textual multidimensional analysis to evaluate the effectiveness of an English for academic purposes (EAP) undergraduate writing course (participants' age: 18–19). Unlike the previous studies, which conducted their own factor analyses and established new textual dimensions, Crosthwaite adopted Biber's (1988) five dimensions. For this analysis, Crosthwaite collected learners' essays and reports at three different time points: before EAP training (Pre-EAP, $n=87$), 9th week of EAP training (Post-EAP, $n=84$), and 13th week of EAP training (Final assessment, $n=86$). The multidimensional analysis on the texts from the three time points indicated that the learners' writing became more informational, more nonnarrative, more context-independent, and less overtly persuasive, illustrating the general trend that the students' writing more closely matched the norms of academic discourse towards the end of the EAP instruction.

Using multidimensional analysis, these previous studies attempted to provide more comprehensive descriptions of L2 writing and to identify any longitudinal developmental progression. Yet they have some limitations. The first limitation comes from the size of the corpora they employed. Valid, representative results require a sufficiently large corpus. The previous studies, however, employed corpora consisting of small numbers of L2 writings (e.g., 207 texts in Friginal & Weigle, 2014; 87 in Crosthwaite, 2016). Therefore, the results of the previous studies are difficult to generalize, and thus remain exploratory. The second limitation is due to the absence of systematic data analysis. For example, Friginal and Weigle did

not employ inferential statistical analysis; their group comparisons are descriptive. Crosthwaite attempted to use Biber's (1988) analytic method, which would enable systematic comparisons across studies, but the study limited its observation to the effects of EAP instruction, and did not control for any effects from learners' individual differences (e.g., nationality, gender, proficiency level, academic background).

This study

Despite its limitations, the previous research consistently suggests that the multidimensional approach can provide a useful window onto the characteristics of L2 writing written by different learners. Aiming to provide a more comprehensive description of how lower-level L2 writing differs from more advanced L2 writing, the present study addresses one primary research question: Within the five functional dimensions identified by Biber (1988), how do textual features of L2 argumentative essays differ across writers' proficiency levels?

Guided by this research question, the present study attempts to address the two major limitations found in the previous studies. First, as for the size of the corpus, the study employed the International Corpus Network of Asian Learners of English, consisting of 5200 English argumentative essays written by 2600 nonnative-English speakers at four different proficiency levels. Second, in regard to the data analysis and intervening effects from participant heterogeneity, the study adopted linear mixed-effects models as a statistical procedure when making comparisons among groups. By statistically factoring out random individual variables such as nationality and academic major, any mediating effects from these variables can be controlled for.

Method

Corpus

The study used the International Corpus Network of Asian Learners of English (ICNALE). The ICNALE, compiled by Ishikawa (2019), is one of the largest learner corpora publicly available online. Currently, it includes 1.3 million words of controlled essays written by 2600 college students (age: $M = 19.66$, $SD = 1.97$; gender: male $= 1127$; female $= 1473$; academic major: humanities $= 653$; social sciences $= 744$; sciences and technology $= 1007$; life sciences $= 92$; unknown $= 104$) in 10 Asian countries, and 200 English native speakers residing in English-speaking countries (age: $M = 29.22$, $SD = 8.33$; gender: male $= 119$; female $= 81$). Of the Asian countries, four are English as a second language (ESL) contexts (i.e., Singapore, Hong Kong, Pakistan, and the Philippines) and six are English as a foreign language (EFL) contexts (i.e., Thailand, Japan, Taiwan, Korea, Indonesia, and China). In a controlled condition ("Appendix 1"), each of the participants produced two argumentative essays in English on two topics: Topic 1 was *It is important for college students to have a part-time job*; and Topic 2 was *Smoking should be completely*

Table 2 ICNALE database

Countries	English	No. of writers	No. of essays		Average no. of words		Average no. of sentences	
			Topic 1	Topic 2	Topic 1	Topic 2	Topic 1	Topic 2
Hong Kong	L2	100	100	100	238.28	227.90	13.51	14.86
Pakistan		200	200	200	236.61	231.62	13.40	14.16
Philippines		200	200	200	251.91	242.33	12.52	13.15
Singapore		200	200	200	232.84	227.52	13.15	13.59
China		400	400	400	241.38	231.27	14.75	15.08
Indonesia		200	200	200	232.56	227.78	14.46	14.90
Japan		400	400	400	223.32	219.56	16.73	16.37
Korea		300	300	300	227.44	221.65	16.13	16.42
Thailand		400	400	400	224.70	223.15	13.96	14.44
Taiwan		200	200	200	233.76	223.00	13.53	13.79
USA, Canada, etc.	L1	200	200	200	232.09	227.68	13.95	14.44

Topic 1: Part-time Job; Topic 2: Smoking. Focusing on L2 writing development, the present study excluded the subcorpus of English native writers from the data analysis

banned at all the restaurants in the country. The distribution of writers' nationality and mean numbers of words/sentences for each topic are reported in Table 2. A paired samples *t* test revealed no statistical difference for number of words and sentences between the two topics across the writers from different countries ($p > .05$ for all).

Another advantage of this corpus, in addition to its size, is that the nonnative writers are classified into four proficiency levels based on objective, standard English proficiency measures, and that there are balanced numbers of learners across the four proficiency levels. The learners' proficiency levels are defined using their scores on major standardized English proficiency tests (e.g., TOEIC, TOEFL-PBT, TOEFL-iBT, IELTS) and/or a standard vocabulary size test (VST; Nation & Beglar, 2007). Using these scores, all learners were classified into three Common European Framework of Reference (CEFR)-linked proficiency bands, where A, B, and C refer to Basic Users (A1: Breakthrough; A2: Waystage), Independent Users (B1: Threshold; B2: Vantage), and Proficient Users (C1: Effective operational proficiency; C2: Mastery), respectively. These three major proficiency levels, each consisting of two sublevels, are reclassified into four groups in the ICNALE: Level 1 (A2: Waystage), Level 2 (B1_1: Threshold, Lower), Level 3 (B1_2: Threshold, Higher), and Level 4 (B2, C1, and C2: Vantage or Higher). In the cases in which the score-level mapping was determined by standard English proficiency test scores, official mapping guidelines offered by each of the official test administrators were used. In the cases in which the score-level mapping was determined using VST scores, no reliable conversion guidelines exist. Therefore, a linear regression modeling of 268 Asian participants who had taken both the TOEIC test and the VST was conducted to obtain a conversion formula (TOEIC $= 10.495 * VST + 289$ ($R^2 = .21$). Learners' VST scores were thus first converted into TOEIC scores, and then into the CEFR levels. After

Table 3 Ratios and distribution of writers at the four proficiency levels (%)

Countries	Level 1 (Way-stage)	Level 2 (Threshold, lower)	Level 3 (Threshold, higher)	Level 4 (Vantage or higher)
Hong Kong	1.0	30.0	52.0	17.0
Pakistan	9.0	45.5	44.0	1.5
Philippines	1.0	5.5	88.0	5.5
Singapore	0.0	0.0	67.0	33.0
China	12.5	58.0	26.3	3.3
Indonesia	16.0	41.1	41.5	1.5
Japan	38.5	44.8	12.3	4.5
Korea	25.0	20.3	29.3	25.3
Thailand	29.8	44.8	25.0	0.5
Taiwan	14.5	43.5	30.5	11.5

the score-level conversion, the corpus contains 480 Level 1 participants, 952 Level 2 participants, 936 Level 3 participants, and 232 Level 4 participants (Table 3).

Analysis

This study analyzed only essays written by nonnative speakers of English ($n = 2600$). A separate analysis was performed on essays on each of the two topics (i.e., Part-time Job and Smoking). The analysis on the corpus was performed using Nini's (2015) Multidimensional Analysis Tagger (MAT), version 1.3.[1] MAT is an open source software for implementing Biber's (1988) approach.[2] Biber's method includes various genres and styles of L1 English, and can therefore serve as a referential tool to appropriately assess the quality of L2 writing. Given L2 writing, MAT can thus produce the five dimensions' scores, which then can be compared to those of the various spoken and written text types (e.g., spontaneous speech, academic prose, official documents, institutional editorials) provided as references in Biber's study.

---

[1] When the reliability of the tool was tested using the Lancaster-Oslo-Bergen (LOB) and Brown corpora, Nini (2015) reported that, in both cases, MAT was largely successful in replicating the results of Biber's (1988) analysis (see the replicated mean dimension scores of 13 text types of the LOB corpus and Brown corpus, respectively, reported in Nini's manual, pp. 9–16). The correlation between the dimension scores computed by MAT and by Biber indicates a successful replication ($r = .973$, $p < .001$). Small discrepancies can be attributed to the difference of the POS taggers used by MAT (Stanford Tagger) and by Biber (unknown).

[2] Note that the current study did not conduct a new factor analysis and did not establish new textual dimensions based on the ICNALE corpus. Instead, similar to Crosthwaite (2016), this study uses Biber's (1988) factor analysis and his textual dimensions. Because the corpus used in this study consists only of argumentative essays, rather than various genres, applying a factor analysis to this highly constrained text corpus would be unlikely to inform us of more than identifying factors, yielding less generalizable interpretations.

After retrieving five mean dimension scores for each of the four proficiency groups from MAT, the present study conducted a statistical analysis to address its primary research question: Within the five functional dimensions identified by Biber (1988), how do textual features of L2 argumentative essays differ across writers' proficiency levels? The mean dimension scores among the four learner groups (Levels 1, 2, 3, and 4) were evaluated by linear mixed-effects models (LMM), a type of analysis that can specify fixed factors and random factors (Baayen, 2008). Unlike traditional general linear modeling (e.g., N-way factorial ANOVA), LMM generalize regression models to have random effects such as subjects or items as well as fixed effects. LMM can thus estimate and factor out any random effects that are uncorrelated with the independent variables in question. In this way, one can obtain more accurate estimations of the effects of the independent variables. In addition, LMM analysis is sufficiently flexible to not require design balance, and it therefore allows for unequal sample sizes between groups in the data. In addition to proficiency, which is the primary independent variable in question, the corpus used in this study provides other learner information that may affect the dimension scores as dependent variables: gender (male or female), academic major (humanities, social sciences, sciences & technology, or life sciences), nationality (10 Asian countries), and topic (two topics).

To obtain accurate analyses with linear mixed-effects models, fixed factors and random factors must be correctly specified. According to Winter (2013), a factor is fixed when the levels tested represent all levels of interest and the effect from a fixed factor "exhaust[s] the population of interest" (p. 18). A factor is random, on the other hand, when the levels tested represent random selections from some population of possible levels of interest, and thus, the effect from the random factor is "far away from exhausting the population" (p. 18). Also, Grace-Martin (2019) defined a factor as fixed if it is the primary treatment that the researcher wants to compare or the secondary control variable that the researcher wants to control for differences in this factor. A factor is random if the researcher is interested in quantifying how much of the overall variation to attribute to the factor, or if the researcher is not interested in knowing which means differ, but wants to account for the variation in the factor.

Given these definitions and descriptions, the present study specified the five variables as follows: (1) proficiency as a fixed factor because this is the primary variable in question, and the proficiency levels (Levels 1–4) relatively fully represent a full set of possible values in the population of interest; (2) gender and (3) academic major as fixed factors because they are the secondary control variables that the present study wants to control for differences, and again, the levels (male or female; humanities, social sciences, sciences & technology, or life sciences) fully or relatively fully represent the full sets of interest; (4) nationality and (5) topic as random factors because they represent random selections from some population of possible levels of interest (10 out of 48 Asian countries, and out of numerous numbers of non-English L1 countries; two topics out of numerous possible topics).

The present study used R, a programming language and software, and its lme4 and lmerTest packages to perform five separate LMM analyses of the effects of proficiency for each of the five dimension scores. The analyses

Table 4 Descriptive statistics for dimension 1: involved versus informational production

	Fixed factors	Levels	Text types	*M*	*SD*
Current study	Proficiency	Level 1 (n = 960)	Argumentative essays (w)	14.63	11.16
		Level 2 (n = 1904)		12.93	9.66
		Level 3 (n = 1872)		9.29	9.71
		Level 4 (n = 464)		6.33	9.26
	Gender	Male (n = 1127)		12.01	10.34
		Female (n = 1437)		10.84	10.18
	Academic major	Humanities (n = 653)		12.22	10.45
		Social Sciences (n = 744)		11.28	10.63
		Sciences and Technology (n = 1007)		11.25	9.78
		Life Sciences (n = 92)		12.14	11.30
Biber (1988)	–	Native	Academic prose (w)	− 14.9	6.0
			Official documents (w)	− 18.1	4.8
			Institutional editorials (w)	− 9.1	4.6
			Prepared speech (s)	2.2	6.7
			Spontaneous speech (s)	18.2	12.3
			Face-to-face conversation (s)	35.3	9.1

M mean, *SD* standard deviation, *w* written text, *s* spoken text

separately computed effects from the random factors and effects from the fixed factors. First, as fixed effects, Proficiency, Gender, and Academic Major (with interaction terms) were entered into the model. Using lsmeans, a Tukey post hoc analysis was carried out. Effect sizes for group differences were also calculated using MuMin to observe the magnitude of the effect from proficiency and/or gender by r^2, where $r^2 = .01$ indicates a small effect size, $r^2 = .09$ indicates a medium effect size, and $r^2 = .25$ indicates a large effect size (Rosenthal & Rosnow, 1984, p. 361). Next, as random effects, intercepts for Nationality and Topic were estimated.

Results

The results of the four learner groups' mean dimension scores were analyzed and are presented together with Biber's (1988) mean scores for six English text types in Tables 4, 5, 6, 7 and 8. Biber's text types represent English spoken discourse and written discourse, and thus, serve as general textual references. For spoken text types, face-to-face conversation, spontaneous speech, and prepared speech were chosen while for written text types, academic prose, official documents, and institutional editorials were selected as references.

Table 5 Descriptive statistics for dimension 2: narrative versus nonnarrative production

	Fixed factors	Levels	Text types	M	SD
Current study	Proficiency	Level 1	Argumentative essays (w)	− 3.10	3.08
		Level 2		− 2.82	3.12
		Level 3		− 2.60	3.02
		Level 4		− 2.39	3.13
	Gender	Male		− 2.78	3.07
		Female		− 2.74	3.06
	Academic major	Humanities		− 2.81	3.03
		Social Sciences		− 2.74	3.07
		Sciences and Technology		− 2.71	3.10
		Life Sciences		− 3.51	2.88
Biber (1988)	–	Native	Academic prose (w)	− 2.60	2.30
			Official documents (w)	− 2.90	1.20
			Institutional editorials (w)	− 0.60	4.40
			Prepared speech (s)	0.70	3.30
			Spontaneous speech (s)	1.30	3.60
			Face-to-face conversa-tion (s)	− 0.60	2.00

M mean, *SD* standard deviation, *w* written text, *s* spoken text

Table 6 Descriptive statistics for dimension 3: situation-dependent versus elaborated reference

	Fixed factors	Levels	Text types	M	SD
Current study	Proficiency	Level 1	Argumentative essays (w)	3.44	4.34
		Level 2		3.53	4.40
		Level 3		3.75	4.22
		Level 4		3.80	3.84
	Gender	Male		3.36	4.22
		Female		3.81	4.31
	Academic major	Humanities		3.51	4.09
		Social Sciences		3.53	4.21
		Sciences and Technology		3.38	4.26
		Life Sciences		3.88	3.76
Biber (1988)	–	Native	Academic prose (w)	4.20	3.60
			Official documents (w)	7.30	3.60
			Institutional editorials (w)	1.80	2.30
			Prepared speech (s)	0.30	3.60
			Spontaneous speech (s)	1.20	4.30
			Face-to-face conversa-tion (s)	− 3.90	2.10

M mean, *SD* standard deviation, *w* written text, *s* spoken text

Table 7 Descriptive statistics for dimension 4: overt expression of persuasion

	Fixed factors	Levels	Text types	*M*	*SD*
Current study	Proficiency	Level 1	Argumentative essays (w)	7.25	6.76
		Level 2		6.41	6.12
		Level 3		6.73	6.16
		Level 4		7.45	5.97
	Gender	Male		6.82	6.13
		Female		6.74	6.34
	Academic major	Humanities		6.89	6.37
		Social Sciences		7.07	6.31
		Sciences and Technology		6.62	6.07
		Life Sciences		8.43	6.23
Biber (1988)	–	Native	Academic prose (w)	−0.50	4.70
			Official documents (w)	−0.20	4.10
			Institutional editorials (w)	4.00	4.10
			Prepared speech (s)	0.40	4.10
			Spontaneous speech (s)	0.30	4.40
			Face-to-face conversation (s)	−0.30	2.40

M mean, *SD* standard deviation, *w* written text, *s* spoken text

Table 8 Descriptive statistics for dimension 5: impersonal versus nonimpersonal style

	Fixed factors	Levels	Text types	*M*	*SD*
Current study	Proficiency	Level 1	Argumentative essays (*w*)	1.52	4.70
		Level 2		1.76	4.49
		Level 3		3.34	5.65
		Level 4		6.02	5.91
	Gender	Male		2.62	5.25
		Female		2.70	5.29
	Academic major	Humanities		2.86	5.29
		Social Sciences		2.67	5.27
		Sciences and Technology		2.69	5.24
		Life Sciences		2.51	5.81
Biber (1988)	–	Native	Academic prose (*w*)	5.50	4.80
			Official documents (*w*)	4.70	2.40
			Institutional editorials (*w*)	0.10	1.80
			Prepared speech (*s*)	−1.90	1.40
			Spontaneous speech (*s*)	−2.60	1.70
			Face-to-face conversation (*s*)	−3.20	1.10

M mean, *SD* standard deviation, *w* written text, *s* spoken text

Dimension 1: involved versus informational production

Table 4 shows that the argumentative texts written by the L2 writers showed positive dimension scores regardless of their proficiency, gender, and academic major, which means that the texts share many traits with Biber's (1988) spoken texts. Nevertheless, a distinctive pattern was found when the scores were compared across proficiency levels. For example, texts by more advanced L2 writers (Level 3, $M=9.29$; $SD=9.71$; Level 4, $M=6.33$; $SD=9.26$) tended to score lower on this dimension than texts by less advanced L2 writers (Level 1, $M=14.63$; $SD=11.16$; Level 2, $M=12.93$; $SD=9.66$). This distinction indicates that the less proficient groups' L2 writings can be characterized as more involved and less informational.

As for the scores on Dimension 1, the LMM first computed effects from the three fixed factors: Proficiency, Gender, and Academic Major. The analysis found a significant main effect for Proficiency with an effect size between small and medium, $F(3, 4955.6)=12.1746$, $p<.001$, $r^2=0.06$. A Tukey's post hoc analysis showed significant differences among the different proficiency groups. (The significance level was set at $p=0.05$ for all post hoc analyses reported in this paper.) The scores of the Level 1 and Level 2 groups, which did not differ significantly, were all significantly higher than those of the Level 3 and Level 4 groups, which differed significantly from each other. Thus, the results can be schematized as follows: Level 1 = Level 2 > Level 3 > Level 4. Main effects for other fixed factors, such as Gender, $F(1, 4950.4)=.4614$, $p>.05$, $r^2<.01$, and Academic Major, $F(3, 4956.7)=.6246$, $p>.05$, $r^2<.01$, were not statistically significant. None of the interaction effects between fixed factors were statistically significant: Proficiency × Gender, $F(3, 4950.3)=1.2083$, $p>.05$, $r^2<.01$; Proficiency × Academic Major, $F(9, 4953.0)=1.9687$, $p>.05$, $r^2<.01$; Gender × Academic Major, $F(3, 4951.7)=1.3699$, $p>.05$, $r^2<.01$; and Proficiency × Gender × Academic Major, $F(9, 4950.6)=1.5021$, $p>.05$, $r^2<.01$.

Second, the analysis also computed the contribution of the random effects to the model. The analysis revealed that Nationality accounted for 16.69 of the variance; Topic accounted for of 0.27 of the variance. The residual of the full model including the random effects was 88.49 of the variance. In order to test the significance of the random effects, a likelihood ratio test was run. The test found that the difference between the full model with the random effects and the model without the random effects was significant, $\chi^2(2)=459.78$, $p<0.01$, indicating that the random effects of Nationality and Topic were significant.

In sum, these results indicate that the texts from the more advanced proficiency groups were more informational whereas those from the lower proficiency groups were less informational, but more interactive, affective, and involved. These results suggest that proficiency can be a significant predictor for Dimension 1.

Dimension 2: narrative versus nonnarrative production

Table 5 shows negative mean scores for all subgroups of Proficiency, Gender, and Academic Major on Dimension 2, aligning with Biber's (1988) score patterns for written texts (e.g., academic prose, official documents, institutional editorials).

In terms of the fixed effects, the LMM revealed no significant main effect for Proficiency, $F(3, 4934.5) = 1.3291$, $p > .05$, $r^2 < .01$; for Gender, $F(1, 4951.9) = 1.6695$, $p > .05$, $r^2 < .01$; or for Academic Major, $F(3, 4881.0) = .2812$, $p > .05$, $r^2 < .01$. No interaction effect was found between any of the fixed factors: Proficiency × Gender, $F(3, 4951.7) = 1.3771$, $p > .05$, $r^2 < .01$; Proficiency × Academic Major, $F(9, 4955.3) = 0.9416$, $p > .05$, $r^2 < .01$; Gender × Academic Major, $F(3, 4955.4) = 1.3082$, $p > .05$, $r^2 < .01$; and Proficiency × Gender × Academic Major, $F(9, 4952.8) = 1.8155$, $p > .05$, $r^2 < .01$.

As for the random effects, the analysis found that Nationality and Topic accounted for .34 and .09 of the variance, respectively. The residual of the full model including the random effects was 9.03 of the variance. The likelihood ratio test found that the difference between the full model with the random effects and the model without the random effects was significant, $\chi^2(2) = 148.18$, $p < 0.01$, indicating that the random effects of Nationality and Topic were significant.

The results together suggest that neither proficiency, gender, nor academic major is a significant predictor for Dimension 2; that is, regardless of the learners' proficiency level, gender, and academic major, their argumentative texts in general comply with norms for written texts, and can be characterized as nonnarrative.

Dimension 3: elaborated versus situation-dependent reference

Table 6 shows positive mean scores for all subgroups of Proficiency, Gender, and Academic Major on Dimension 3, aligning with Biber's (1988) academic prose, one of the written text types.

The LMM found no significant main effects for any of the fixed factors: Proficiency, $F(3, 4951.2) = .533$, $p > .05$, $r^2 < .01$; Gender, $F(1, 4952.3) = .243$, $p > .05$, $r^2 < .01$; and Academic Major, $F(3, 4927.7) = .998$, $p > .05$, $r^2 < .01$. No interaction effect was found between any of these fixed factors: Proficiency × Gender, $F(3, 4955.5) = 2.646$, $p > .05$, $r^2 < .01$; Proficiency × Academic Major, $F(9, 4957.1) = 1.908$, $p > .05$, $r^2 < .01$; Gender × Academic Major, $F(3, 4955.5) = 2.647$, $p > .05$, $r^2 < .01$; and Proficiency × Gender × Academic Major, $F(9, 4952.9) = 1.504$, $p > .05$, $r^2 < .01$.

In terms of the random effects, the analysis revealed that Nationality accounted for .92 of the variance, and Topic accounted for .01 of the variance. The residual of the full model including the random effects was 16.83 of the variance. The likelihood ratio test found that the difference between the full model with the random effects and the model without the random effects was significant, $\chi^2(2) = 153.38$, $p < 0.01$, indicating that the random effects of Nationality and Topic were significant.

Similar to those for Dimension 2, these results indicate that proficiency, gender, and academic major are not significant predictors for Dimension 3; the learners' argumentative texts, in general, comply with norms for written texts, and can be characterized as having elaborated reference.

Dimension 4: overt expression of persuasion

Table 7 indicates that all learners gained higher mean scores on Dimension 4 compared to the scores for institutional editorials in Biber's (1988) study. This study's highest scores were observed for this dimension, probably because the ICNALE texts are intended to be argumentative and persuasive.

The analysis found no significant main effect for Proficiency, $F(3, 4955.6) = .4131$, $p > .05$, $r^2 < .01$; Gender, $F(1, 4950.4) = .4782$, $p > .05$, $r^2 < .01$; or Academic Major, $F(3, 4956.7) = .8695$, $p > .05$, $r^2 < .01$. The interaction effects between any of the fixed factors were not significant: Proficiency × Gender, $F(3, 4950.3) = 0.2331$, $p > .05$, $r^2 < .01$; Proficiency × Academic Major, $F(9, 4953.0) = 1.6103$, $p > .05$, $r^2 < .01$; Gender × Academic Major, $F(3, 4951.7) = 0.3048$, $p > .05$, $r^2 < .01$; and Proficiency × Gender × Academic Major, $F(9, 4950.6) = .4986$, $p > .05$, $r^2 < .01$.

The analysis for the random effects showed that Nationality accounted for 6.34 of the variance; Topic accounted for 3.18 of the variance. The residual of the full model including the random effects was 33.53 of the variance. The likelihood ratio test found that the difference between the full model with the random effects and the model without the random effects was significant, $\chi^2(2) = 675.99$, $p < 0.01$, indicating that the random effects of Nationality and Topic were significant.

These results illustrate that all the learners, regardless of their proficiency, gender, and academic major, tended to use positive features for Dimension 4, such as prediction modals (e.g., *will, would, shall*), suasive verbs (e.g., *command, insist*), and necessity modals (e.g., *must, should*), and thus, their texts tend to read as overtly persuasive.

Dimension 5: impersonal versus nonimpersonal style

Table 8 shows positive mean scores for all subgroups of Proficiency, Gender, and Academic Major on Dimension 5, aligning with Biber's (1988) written text types. In particular, the texts of one of the proficiency subgroups, Level 4 ($M = 6.02$; $SD = 5.91$), showed quite similar scores to Biber's academic prose and official documents, indicating that these texts share more traits with Biber's written texts than with Biber's spoken texts.

In terms of the fixed effects, the analysis found a significant main effect for Proficiency with an effect size between small and medium, $F(3, 4954.5) = 4.9411$, $p < .001$, $r^2 = .06$ (post hoc: Level 1 = Level 2 < Level 3 < Level 4); a significant main effect for Gender with a negligible effect size, $F(1, 4950.3) = 4.453$, $p < .05$, $r^2 < .01$; and a significant main effect for Academic Major with a negligible effect size, $F(1, 4956.0) = 4.297$, $p < .05$, $r^2 < .01$. The minimal effect sizes and the low p values for the main effects of Gender and Academic Major might be false positives, due to confounding by the large sample size ($N = 5200$) of the study (Type I error). The analysis also found significant interaction effects between Proficiency and Gender, $F(3, 4950.2) = 6.9989$, $p < .05$, $r^2 < .01$, and between Gender and Academic Major, $F(3, 4951.3) = 3.6561$, $p < .05$, $r^2 < .01$, but, again, the negligible effect sizes indicate

a false positive (Type I error) caused by the large sample size. Other interaction effects were not significant: Proficiency × Academic Major, $F(9, 4952.2) = 1.909$, $p > .05$, $r^2 < .01$; and Proficiency × Gender × Academic Major, $F(9, 4950.4) = 1.8968$, $p > .05$, $r^2 < .01$.

In terms of the random effects, the analysis revealed that Nationality accounted for 6.03 of the variance; Topic accounted for .23 of the variance. The residual of the full model including the random effects was 22.69 of the variance. The likelihood ratio test found that the difference between the full model with the random effects and the model without the random effects was significant, $\chi^2(2) = 649.07$, $p < 0.01$, indicating that the random effects of Nationality and Topic were significant.

In sum, these results together suggest that proficiency can be a significant predictor for Dimension 5; higher proficiency writers tend to write their argumentative texts in an impersonal style while lower proficiency writers tend to write in a non-impersonal style. L2 writing, in other words, gradually becomes more impersonal as writers' English proficiency develops.

Discussion

Inspired by Biber's (1988) multidimensional analysis, the present study explored textual features of L2 writing by four different proficiency groups. In particular, the study investigated whether proficiency, or L2 development, could be a significant predictor for the differences shown in the proficiency groups' dimension scores. The results of the present study are summarized in Fig. 1.

The textual features of the learners' essays generally complied with native speaker norms for written discourse rather than spoken discourse. For some dimensions, however, these features were more likely to be observed in texts from the advanced learner groups than in texts from the less advanced groups: Dimension 1 (involved vs. informational production) and Dimension 5 (impersonal vs. non-impersonal style). As L2 proficiency develops, learners' writing becomes more likely to incorporate the linguistic features highly associated with native speakers' writing (i.e., more informational and more impersonal). Less advanced L2 learners, on the other hand, frequently used linguistic features in their L2 writing that are highly associated with native speakers' spoken discourse (i.e., more involved and less impersonal).

Such developmental trends of L2 writing are in line with previous findings on the occurrence frequency of a range of linguistic features in L2 writing (Biber et al., 2016; Friginal & Weigle, 2014; Grant & Ginther, 2000; Shaw & Liu, 1998; Taguchi et al., 2013). Friginal and Weigle (2014) also showed that learners' essays become more informational and impersonal over time and with higher proficiency. Grant and Ginther (2000) and Taguchi et al. (2013) reported that more advanced learners' essays contain higher numbers of nouns, attributive adjectives, and prepositional phrases (informational features) than less advanced learners' essays. Shaw and Liu (1998) demonstrated that less advanced learners' texts contain more contractions and first-person pronouns (involved features). They further identified more frequent use of passives, conjuncts, and adverbial expressions (impersonal features) in more
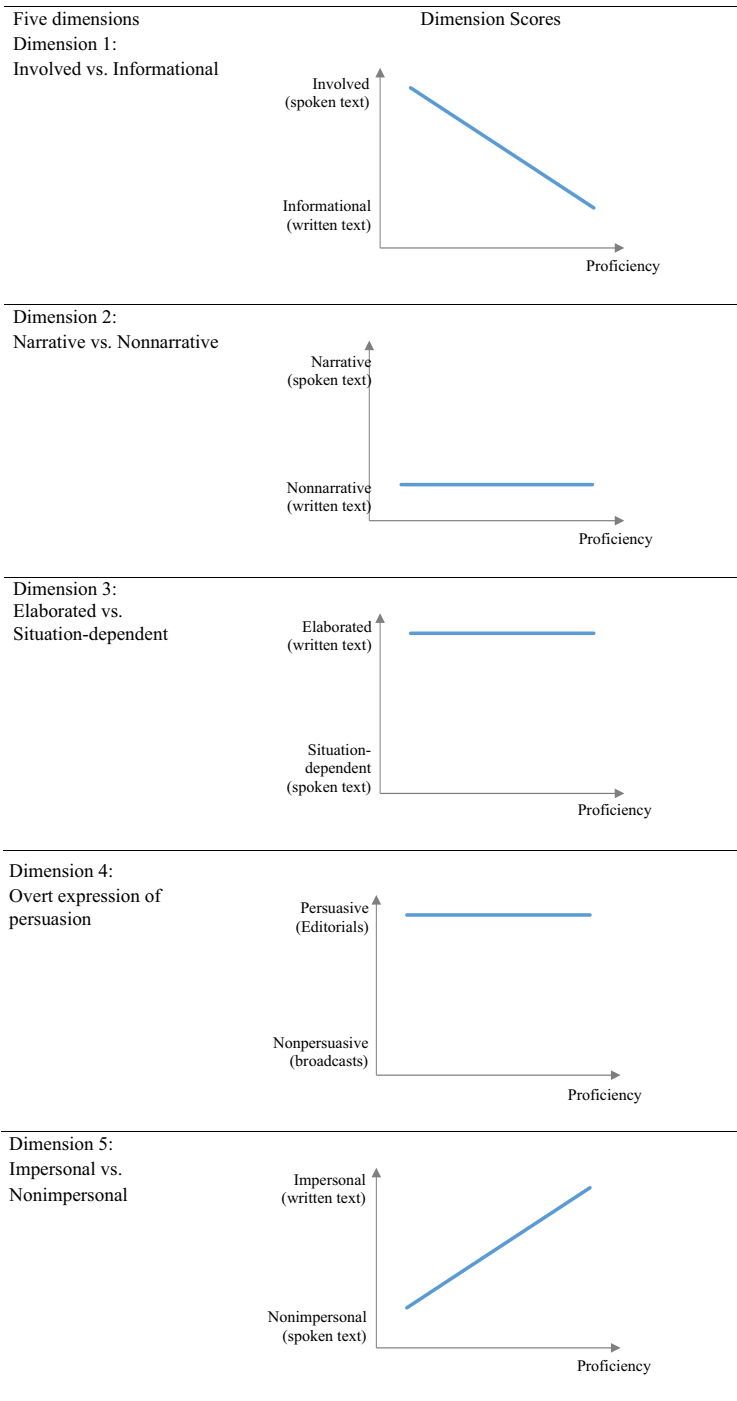
Five dimensions
Dimension 1:
Involved vs. Informational

Dimension Scores

Dimension 2:
Narrative vs. Nonnarrative

Dimension 3:
Elaborated vs.
Situation-dependent

Dimension 4:
Overt expression of
persuasion

Dimension 5:
Impersonal vs.
Nonimpersonal

Fig. 1 Changes in dimension scores across proficiency levels

advanced writing than in less advanced writing. Biber et al. (2016) found that high-rated responses involved more features of passivization, such as passive voice verbs and passive -*ed* relative clauses (impersonal features).

The pattern change observed for Dimension 1 (involved vs. informational production) in this study is thus in line with the results of previous research. The present study found that L2 writings gradually become more informational and less interactive as L2 proficiency develops. For example, lower-level learners tend to use the linguistic features that are often employed for describing actions occurring in the immediate context of interaction and overtly expressing writers' private attitudes, thoughts, and emotions, such as first-person pronouns (e.g., *I*), present tense verbs (e.g., *want*, *plan*, *learn*), and private verbs (e.g., *think*, *feel*), and thus, associated with interactive or affective purposes (e.g., *It's like when I want to buy drinks or foods or clothes, I just spend and very careless. Then I become broke*; see Example 1 and Example 2 in "Appendix 2"). These descriptive statements were expressed more impersonally in higher-level writers' essays (e.g., *For college students, they are mature enough to be financial independence.*). Also, the lower-level learners used such positive linguistic features as private verbs (e.g., *think, know*), *that*-deletion (e.g., *I think [0] it is*), first/second-person pronouns (e.g., *I, you*), contractions (e.g., *don't, can't*), present tense verbs (e.g., *is, have*), general emphatics (e.g., *just*), pronoun *it*, *be* as a main verb (e.g., *I <u>was</u> poor*), and amplifiers (e.g., *very*) more often than the higher-level learners. Frequent use of these linguistic features resulted in lower-level learners' higher mean scores on Dimension 1, and essays that can be characterized as verbal, highly involved, interactive, and affective.

In contrast, texts from more advanced learners show a low density of these positive features but a high density of negative features (e.g., nouns, prepositions, and attributive adjectives), which lead to a text's character as informational rather than interactive or affective (see Example 3 in "Appendix 2"). While for lower-level learners, the use of nouns and attributive adjectives was limited mainly to keywords given in the prompts (e.g., *important part-time job*), nouns occur frequently for more advanced learners, and the nouns are frequently modified by attributive adjectives within a prepositional phrase (e.g., *a complex enterprise, of our future career*). According to Biber (1988), these linguistic features function to package information about specific referents by elaborating or specifying the exact nature of the nominal information. High amounts of information can be integrated into a text in this way. On the other hand, very limited use of first person pronouns (e.g., *I*) and verbs creates an impression of objectivity.

Despite the gradual developmental pattern shown in Dimension 1, it should be noted that the Dimension 1 scores of Level 4, the most advanced learner group, were still above zero (Part-time Job: $M = 6.14$, $SD = 9.28$; Smoking: $M = 6.53$, $SD = 9.25$), thus distinguishing this group's writing from Biber's (1988) written discourse, which had negative mean scores (e.g., institutional editorials, $M = -9.1$, $SD = 4.6$; academic prose, $M = -14.9$, $SD = 6.0$; and official documents, $M = -18.1$, $SD = 4.8$). One might think that writing with high informational density and delivering exact informational content in an L2 are highly complex and difficult skills to achieve. The topics of the argumentative writing in the corpus, however, were nonacademic and closely related to the college students' real-life experience. Therefore, the writers'

positive scores on Dimension 1 (involved) can be attributed to the experiential, less formal nature of the topics (Weigle & Friginal, 2015). In fact, a separate, subsequent analysis that computed the scores for native speaker essays on the same topics (recall that ICNALE's native subcorpus [$n = 200$] was not included in the analysis of the present study) found the native group's Dimension 1 scores to be near 10 (Part-time Job: $M = 10.53$, $SD = 10.63$; Smoking: $M = 10.28$, $SD = 10.71$); that is, even higher than Level 4's scores. The lower scores of Level 4 compared to the native group further confirm the developmental features of L2 writing on Dimension 1.

In terms of Dimension 5 (impersonal vs. nonimpersonal style), the subsequent qualitative analysis found that the occurrence of the positive features for Dimension 5 was quite different across the L2 proficiency groups. On the other hand, the difference was particularly striking for the linguistic features associated with human agents: agentless passives and *by* passives. For example, the advanced writers' texts (see Example 6 in "Appendix 2") tended to be impersonal; human agents (e.g., *people, I, you*) rarely appear, and the verbs' agents are unspecified (e.g., *alternative measures should be introduced…*; *smoking is largely perceived…*). Using these passive constructions, the texts emphasize abstract, technical, and conceptual information over active, human participants and concrete topics. These features appear less densely in texts written by lower-level learners (see Examples 4 and 5, respectively, in "Appendix 2"). As proficiency increases, writers are less likely to use human agents in subject positions (e.g., *some agree*; *others are against*; *people say…*) and more likely, if they use *by* passives, to give them animate patients (e.g., *you will be fined…*).

Another interesting finding in the current research is that the trend of L2 writing development from spoken to written style was not observed for Dimensions 2, 3, and 4, as illustrated by the absence of significant mean score differences among the four proficiency groups. Instead, all groups' use of the textual features assessed by these dimensions appeared to be close to that of native speakers' written discourse (non-narrative style and elaborated reference) and sufficiently persuasive (overt expression of persuasion).

As for Dimension 3, for example, learners frequently used the positive features (e.g., *wh*-relative clauses in object position, *wh*-relative clauses in subject position, pied-piping constructions, phrasal coordination, and nominalizations), and thus, the texts feel referentially explicit and integrated (see Examples 7 and 8 in "Appendix 2"). For example, the *wh*-relative clauses and pied-piping constructions function to pack information into noun phrases (e.g., *place and manner <u>in which</u> …*; *the person <u>who</u> contra or disagree about…*). With these *wh*-relative clauses, nominal referents are explicitly identified in a text. The referential explicitness of the texts is further enhanced by elaboration and integration of information, which is accomplished through the use of phrasal coordination (e.g., *place <u>and</u> manner, the best <u>and</u> worst*) and nominalization (e.g., *restrictions, management*). Sharing these features, the texts were highly situation-independent and specific. In terms of Dimension 4, the learners tended to use the positive features quite effectively (see Examples 9 and 10 in "Appendix 2"), such as predictive modals (e.g., *would have to…*; *will…*), possibility modals (e.g., *can gain more pocket money…*; *may happen in their future…*), conditional subordination (e.g., *if the student is…*), and necessity modals (e.g., *we*

should earn money…), to evoke different perspectives and make supporting arguments. Modals and conditionals leading to a final conclusion suggest that the writer's opinion is the correct opinion.

No significant group differences in these dimensions might be explained by the writing topics. For example, unlike the present study, Friginal and Weigle's (2014) study found that high-rated L2 essays were more narrative than nonnarrative. The authors attributed this trend to the essay topics (i.e., the writers' current homes or experiences of good/bad teachers), which elicited descriptive writing. Descriptive writing asks writers to describe a particular experience, and therefore calls for personal narrative. The absence of a significant mean score difference, particularly in Dimension 2 (narrative vs. nonnarrative) and Dimension 4 (overt expression of persuasion), across proficiency groups in the current study seems likewise to be due to the writing topics, which elicited argumentative essays. The argumentative genre calls for evaluating evidence and establishing a position. The two genres of narrative discourse and argumentative essay, therefore, can be expected to employ distinct sets of linguistic features, and the lack of variation in this study's Dimension 2 (and Dimension 4[3]) scores could be attributed to the genre-specific characteristics of argumentative writing.

Another reason for the small difference among the proficiency groups found for these dimensions could be the relative ease of acquiring the relevant linguistic features. Recall that for Dimension 3, written discourse involves high frequency of *wh*-relative clauses, phrasal coordination, and nominalization (elaborated reference) and low frequency of adverbs (situation-dependent reference). Grant and Ginther (2000) found minimal differences in the frequency of *wh*-relative clauses used by three proficiency groups above a certain level. Ai and Lu (2013) also identified no significant difference in the use of subordination (dependent clauses per clause) or coordination (coordinate phrases per clause) in high-level versus low-level L2 writing. These findings may partially support the idea that acquisition of *wh*-relatives occurs relatively early. Therefore, while it is stereotypically associated with text complexity (Biber et al., 2011), clausal subordination may not be indicative of academic writing, and thus, might not be a significant predictor for L2 writing development.

Conclusions and implications

Using Biber's (1988) multidimensional analysis, this study aimed to investigate patterns of distinctive linguistic and functional characteristics in L2 writing at different levels of proficiency. The study found different linguistic and functional features across L2 proficiency levels. To be specific, the lower proficiency groups' writings

---

[3] It should be noted that the learners' scores on Dimension 4 (being overtly persuasive) are much higher than any found by Biber (1988) for any type of discourse, which could also be due to the topic effect (argumentative essay). The subsequent analysis of the native group's essays found scores of over 8 on Dimension 4 (Part-time Job: $M = 8.04$, $SD = 5.47$; Smoking: $M = 9.01$, $SD = 6.04$), which was also higher than any of the scores for Biber's text types. This finding indicates that explicit persuasive intent is also overtly shown in the essays in this corpus.

were more similar to native speakers' spoken text while the higher proficiency groups' writings were closer to native speakers' written text. These patterns were found for the dimensions associated with dense delivery of information (Dimension 1: involved vs. informational production) and academic use of language (Dimension 5: impersonal vs. nonimpersonal style). Moreover, the study's results suggest a gradual developmental progression for these dimensions. Thus, it offers supportive evidence that some aspects of L2 writing competence such as the ability to write with high informational density and about abstract, conceptual, or technical matters are incrementally acquired, requiring more effort for a longer period of time. The lack of difference across proficiency levels in the use of nonnarrative style (Dimension 2), elaborated reference (Dimension 3), and overt expression of persuasion (Dimension 4) can be explained by the relative ease of acquiring the associated linguistic structures and the effect of the argumentative-essay-eliciting topics chosen in this research. It should be noted that Dimensions 2, 3, and 4 might have exhibited a difference by proficiency in different conditions; that is, if the study had included even lower proficiency learners and/or nonacademic writing.

This study provides useful implications for L2 writing instruction. The results indicate that there is no single construct or feature that defines *good, native-like* writing; rather, L2 writing competence is a complex and multifaceted skill consisting of diverse subcompetences, such as understanding text types, understanding how linguistic features comply with written conventions, and understanding the linguistic functions achieved by interaction among those linguistic features. Besides teaching L2 vocabulary and grammar, writing instruction should include the description of a range of linguistic features in relation to their functions. Learners' slow progress in writing implies that writing well in the style of the target culture could be challenging for L2 learners, and hence is unlikely to be achieved incidentally in the context of meaning-focused classrooms. For this reason, L2 interventions and teaching materials should promote learners' ability to make meaningful connections among linguistic features and functions, which in turn will contribute to the development of global L2 writing competence.

Appendices

Appendix 1: The ICNALE's writing instructions

Do you agree or disagree with the following statements? Use reasons and specific details to support your opinion

Topic 1: It is important for college students to have a part-time job

Topic 2: Smoking should be completely banned at all the restaurants in the country

Instructions

1. Clarify your opinions and show the reasons and some examples

---

2. You can use 20–40 min for each essay. This means that you have 40 to 80 minutes to complete two essays. Do not finish too early or spend too much time

3. You must use MS Word or a similar word processor

4. Do not use dictionaries or other reference tools

5. Do not plagiarize anyone else's essays

6. The length of your single essay should be from 200 to 300 WORDS. Too short or too long essays cannot be accepted. You can check the length of your essay using the word count function of MS Word

7. You must run spell check before completing your writing

Appendix 2: L2 argumentative essay examples 1–10

Dimension 1 (Topic 1: part-time job): Examples 1–3

Example 1. L1 = Thai; Level 1

If **you**[SPP2] never **do**[VPRT] something **special**[JJ] in[PIN] a university **life**[NN], **you**[SPP2] should try to **find**[PRIV] **extra**[JJ] **work**[NN] to do more. They **have**[VPRT] the **money**[NN] to use **during**[PIN] **class**[NN]. **I**[FPP1] **know**[VPRT][PRIV][THATD] **it**[PIT]**'s**[VPRT][CONT] tired, if compared **to**[PIN] **those**[DEMP] who **study**[VPRT] only. […] It's like when **I**[FPP1] **want**[VPRT] to buy **drinks**[NN] or **foods**[NN] or **clothes**[NN], **I**[FPP1] **just**[EMPH] **spend**[VPRT] and **very**[AMP] **careless**[JJ]. Then **I**[FPP1] **become**[VPRT] **broke**[JJ]. **My**[FPP1] **parents**[NN] **do**[VPRT]**n't**[CONT] **know**[PRIV] that **I**[FPP1] **have**[VPRT] to borrow **money**[NN] **from**[PIN] **friends**[NN].

Example 2. L1 = Japanese; Level 2

I[FPP1] **think**[VPRT][THATD] **it**[PIT] **is**[VPRT][BEMA] **very**[AMP] **important**[JJ] **part-time**[JJ] **job**[NN] **for**[PIN] **me**[FPP1]. […] **I**[FPP1] experienced what **is**[VPRT] hospitality there. For example, **I**[FPP1] **study**[VPRT] a **way**[NN] **of**[PIN] talking and **choice**[NN] **of**[PIN] **words**[NN]. **I**[FPP1] **think**[VPRT] talking **skill**[NN] is[VPRT][BEMA] very[AMP] important **in**[PIN] communication or business. **I**[FPP1] have[VPRT] to talk **with**[PIN] many **customers**[NN]. **I**[FPP1] **was**[BEMA] poor **at**[PIN] talking but, **I**[FPP1] **become**[VPRT] **good**[JJ] **at**[PIN] talking gradually now.

Example 3. L1 = Mandarin Chinese; Level 4

By[PIN] going out **of**[PIN] the **school**[NN] and locating **ourselves**[FPP1] **at**[PIN] a **complex**[JJ] **enterprise**[NN], **we**[FPP1] are[VPRT] urged to meet and approach **with**[PIN] **our**[FPP1] **supervisor**[NN] and **boss**[NN], who **are**[VPRT] usually successful and complicated **in**[PIN] their **own**[JJ] **field**[NN]. They **can**[POMD] be **good**[JJ] **mentors**[NN] **of**[PIN] our[FPP1] **future**[JJ] **career**[NN] **development**[PIN]. In addition, **by**[PIN] working part-timely, **students**[NN] **can**[POMD] earn some **money**[NN], **which**[SERE] alleviates[VPRT] the **financial**[JJ] **burden**[NN] **of**[PIN] their **family**[NN]. **For**[PIN] **college**[NN] **students**[NN], they **are**[VPRT][BEMA] **mature**[JJ] **enough**[JJ] to be[BEMA] **financial**[JJ] **independence**[NN].

Note. Positive features are highlighted in red; negative features are highlighted in blue; [AMP], amplifier; [BEMA], *be* as main verb; [CONT], contraction; [FPP1], first-person pronoun; [JJ], attributive adjective; [NN], noun; [PIN], preposition; [PIT], pronoun *it*; [POMD], possibility modal; [PRIV], private verb; [SPP2], second-person pronoun; [VPRT], present tense; [THATD], subordinator *that*-deletion; [VPRT], present tense.

Dimension 5 (Topic 2: smoking): Examples 4–6

Example 4. L1 = Thai; Proficiency = Level 1

[…] When your parents were young, people could buy cigarettes and smoke pretty much anywhere-even in hospitals. Today we're more aware about how bad smoking is for our health. Smoking **is restricted**[PASS] or **banned**[PASS] in almost all public places and cigarette companies **are** no longer **allowed**[PASS] to advertise on TV, radio, and in many magazines. People start smoking for a variety of different reasons.

Example 5. L1 = Mandarin Chinese; Proficiency = Level 2

[…] And when it comes to the argument whether to forbid the smoking completely, the people can't make the same decision. Some agree when others are against. In my opinion, I think the cigarette should **be forbidden**[PASS]. The cost of cigarette is a large burden. Many people, including the poor man, spend thousands of money buying cigarette.

Example 6. L1 = Mandarin Chinese; Proficiency = Level 4

This **is exacerbated by the fact**[BYPA] that nicotine present in cigarettes makes it hard for smokers to kick the habit and **hence**[CONJ] they **are inclined**[PASS] to smoke every few hours and even before or after meals. […] Additionally, **while**[OSUB] smoking is largely **perceived**[PASS] as an unhealthy habit, smokers may argue that it is a form of activity that helps them to de-stress and a complete ban to smoke in restaurants is discriminatory. […] Alternative measures should **be introduced**[PASS] instead of applying a complete ban on smoking throughout the country. For example, depending on the design and location of the restaurant, partial smoking bans could **be implemented**[PASS] instead. For instance, restaurants with outdoor and indoor dining areas can segregate their dining areas such that smoking **is** only **permitted**[PASS] outdoors. […] Such a measure involving partial ban of smoking practices has already **be**en **implemented**[PASS] in Singapore.

Note. Positive features are highlighted in red; [BYPA], *by*-passive; [CONJ], conjunct; [OSUB], other adverbial subordinator; [PASS], agentless passive.

Dimension 3 (Topic 2: Smoking): Examples 7–8

Example 7. L1 = Indonesian; Proficiency = Level 2

We have debate in the class about Should **Government**[NOMZ] Ban Smoking In The Restaurant. I as the person **who**[WHSUB] contra or disagree about **intervention**[NOMZ] of the **government**[NOMZ] to ban smoking in the restaurant. **Actually**[RB] the **majority**[NOMZ] of people know about the dangerous of smoke and they know what is the best **and**[PHC] worst for them, however the **majority**[NOMZ] of them are smoker. I think if the **government**[NOMZ] want to reduce the smoker can use other ways such as give **elucidation**[NOMZ] and more **information**[NOMZ] about cigarette or ban smoking in the hospital, etc. As we know that restaurant have **management**[NOMZ] that manage it in order. I am sure that each **management**[NOMZ] in each restaurant will give the best for their restaurant.

Example 8. L1 = Thai; Proficiency = Level 3

Restrictions[NOMZ] on the time, place **and**[PHC] manner **in which**[PIRE] public smoking may occur have been increasing over the last several years. While the **early**[TIME] focus of anti-smoking initiatives was on consumer **education**[NOMZ] **and**[PHC] industry advertising **restrictions**[NOMZ], over the past two decades, smoking opponents have **increasingly**[RB] taken their battle to state and local **governments**[NOMZ], seeking **prohibitions**[NOMZ] on smoking in a wide variety of public **establishments**[NOMZ]. Advocates of these bans claim to be protecting the nonsmoking public **and**[PHC] workers from the adverse health effects of secondhand smoke. Opponents of smoking **restrictions**[NOMZ] dispute the existence and/or **severity**[NOMZ] of these adverse consequences and claim that bans have the unintended consequence of hurting **business**[NOMZ].

Note. Positive features are highlighted in red; negative features are highlighted in blue; [NOMZ], nominalization; [PHC], phrasal coordination; [PIRE], pied-piping relative clause; [TIME], time adverbial; [RB], total adverb; [WHSUB]; *wh*-relative clause on subject position.

Dimension 4 (Topic 2: smoking): Examples 9–10

Example 9. L1 = Japanese; Proficiency = Level 2

[…] Having a part time job **can**[POMD] train the teenagers how **to**[TOOO] earn money and let them know that it is not an easy thing **to**[TOOO] make money. Then they **will**[PRMD] use their money more and more carefully. Besides, I think that being a college student, **if**[COND] we want **to**[TOOO] buy something we really want, we **should**[NEMD] earn money by ourselves, not using the money of our parents. Using our own money, we **will**[PRMD] not feel guilty anymore. Or **if**[COND] the kids **can**[POMD] give the parents the money that is earned by them, their parents **will**[PRMD] be very happy. The second reason is that having a part time job, we **will**[PRMD] have the working experience and have the chance **to**[TOOO] adapt the different relationship between classmates and colleagues earlier than the people who do not have the part time job. Having the working experience **can**[POMD] help teenagers find their ideal job more easily in the future. And the teenagers **can**[POMD] **also learn**[SPAU] how **to**[TOOO] deal with the problems that **may**[POMD] happen in their future working circumstance. Third, teenagers **can**[POMD] **also expand**[SPAU] their interpersonal relationship by having a part time job. In the recent year, interpersonal relationship has become the most important factor in our working environment.

Example 10. L1 = Cantonese Chinese; Proficiency = Level 4

[…] First and foremost, it **can**[POMD] **be physically straining**[SPAU] for the students take up a part-time job. Having a part-time job **would**[PRMD] mean that a student **would**[PRMD] have **to**[TOOO] travel to the workplace, work for a **stipulated**[SUAV] number of hours and then travel back home. On top of that, the student **would**[PRMD] have **to**[TOOO] travel to school and sit in lectures for long periods of time. Over time, the student **would**[PRMD] feel physically tired trying **to**[TOOO] balance out work and school. This **could**[POMD] **negatively influence**[SPAU] their grades since the student is not in tip-top condition **to**[TOOO] study effectively. […] Lastly, one **may**[POMD] argue that a student **can**[POMD] gain more pocket money or support the family through working part-time. However, **if**[COND] the student is truly in need of finances, one **can**[POMD] **always apply**[SPAU] for financial assistance **to**[TOOO] alleviate their financial strain. Thus, there is no need for one **to**[TOOO] take up a part-time job while one is studying in college. In conclusion, the most important job for a student is **to**[TOOO] work hard in their studies so that they **can**[POMD] excel and in future, they **would**[PRMD] be able **to**[TOOO] land a proper full-time job.

Note. Positive features are highlighted in red; [COND], conditional adverbial subordinator; [POMD], possibility modal; [PRMD], predictive modal; [SPAU], split auxiliary; [SUAV], suasive verb; [TOOO], infinitive.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). Amsterdam: John Benjamins.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Becker, A. (2010). Distinguishing linguistic and discourse features in ESL students' written performance. *Modern Journal of Applied Linguistics, 2,* 406–424.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly, 36,* 9–48.

Biber, D., & Gray, B. (2013). *Discourse characteristics of writing and speaking task types on the TOEFL iBT test: A lexico-grammatical analysis (TOEFL iBT research report 19)*. Princeton, NJ: Educational Testing Service.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45,* 5–35.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37,* 639–668.

Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes, 22,* 166–178.

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28,* 414–420.

Frase, L. T., Faletti, A., Ginther, A., & Grant, L. (1999). *Computer analysis of the TOEFL test of written English*. Princeton, NJ: Educational Testing Service.

Friginal, E., & Hardy, J. A. (2014). Conducting Biber's corpus-based multi-dimensional analysis using SPSS. In T. Berber-Sardinha & M. Veirano-Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 297–316). Philadelphia, PA: John Benjamins.

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing, 26,* 80–95.

Grace-Martin, K. (2019). Specifying fixed and random factors in mixed models. Retrieved from https://www.theanalysisfactor.com/specifying-fixed-and-random-factors-in-mixed-models/. Accessed 27 Mar 2019.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*(2), 123–145.

Ishikawa, S. (2019). The ICNALE: The international corpus network of Asian learners of English. Retrieved from http://language.sakura.ne.jp/icnale/. Accessed 27 Mar 2019.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing, 12,* 377–403.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45,* 36–62.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Nini, A. (2015). Multidimensional analysis tagger (version 1.3). Retrieved from http://sites.google.com/site/multidimensionaltagger. Accessed 27 Mar 2019.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24,* 492–518.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.

Shaw, P., & Liu, E. T. (1998). What develops in the development of second-language writing? *Applied Linguistics, 19,* 225–254.

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal, 102,* 310–332.

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly, 47,* 420–430.

Weigle, S. C., & Friginal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes, 18,* 25–39.

Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses (Tutorial 2). Retrieved from www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf. Accessed 27 Mar 2019.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity (report no. 17)*. Honolulu: University of Hawaii Press.