Development of L2 disciplinary literacy: A multidimensional analysis

Abstract

Bilingual education programmes have become popular worldwide in part because they are believed to facilitate second/foreign/additional language (L2) learning, often with English as the target L2. In such programmes, students learn content and L2 simultaneously, but a common difficulty they encounter is mastering disciplinary literacy in the L2. However, research exploring bilingual students' development of L2 disciplinary literacy remains scarce. Therefore, this paper analyses a corpus of biology examination essays produced by a stratified sample of students studying in English medium education (EMI) in Hong Kong (204 essays totalling 45,823 words) using Multidimensional Analysis Tagger (Nini, 2015). The multidimensional analysis shows that this sample generally produced texts containing features closely related to scientific exposition. This paper also compares the features of essays written by students at different levels of academic performance as measured by their examination scores. The results show that the essays produced by students with better scores were more similar to academic prose, consisted of denser and more abstract information and used more academic vocabulary, nominalisation, conjunctions, and passive voice. This study demonstrates a potential relationship between students' mastery of disciplinary literacy and their academic achievement, thereby having implications for pedagogy in bilingual education programmes.

Keywords: multidimensional analysis; learner corpus; bilingual education; English medium instruction; disciplinary literacy

1. Introduction

Bilingual education programmes can be broadly characterised as students learning non-language content subjects through a second, foreign, or additional language (L2). With such a definition, bilingual education programmes cover a wide array of programmes including immersion programmes, English medium instruction (EMI), and Content and Language Integrated Learning (CLIL). While acknowledging differences among these bilingual education programme variations (e.g., Cenoz et al., 2014), it is undeniable that such programmes are flourishing worldwide, especially in non-Anglophone countries where students are expected to learn English, a common target L2, through EMI schools and/or universities (Macaro, 2018). Students in bilingual education programmes are thus expected to learn content and language simultaneously. The "language" here means the L2, which is often the language in which students are less proficient and includes both everyday communicative language during classroom interaction and academic language associated with content subjects i.e. disciplinary literacy (Llinares et al., 2012). Disciplinary literacy has been shown to pose challenges for students, yet this register of language is essential for students to construe content knowledge in a particular discipline (Lemke, 1990), and would even contribute to their "success in education, and thus their prospects in the wider world" (Martin, 2013, p. 23). Hence, considering the worldwide spread of bilingual education programmes, it is important to examine students' development of L2 disciplinary literacy and how it may have affected their academic achievement. Some previous attempts (e.g., Whittaker & McCabe, 2020) have revealed students' mastery of particular features of disciplinary literacy (e.g., subject-specific vocabulary, sentence patterns) while some others (e.g., Coetze-Lachmann, 2007) have demonstrated how students' disciplinary literacy may correlate with their academic achievement. This study seeks to address the research issues in a more comprehensive way by analysing the linguistic features of a corpus of essays written by secondary school students studying in English medium education (EMI) with multidimensional analysis and correlating such linguistic features with students' academic achievement. By doing so, this study can demonstrate the integral role of disciplinary literacy in students' academic achievement in bilingual education programmes, thereby yielding pedagogical implications to support this group of students.

2. Students' disciplinary literacy development in bilingual education programmes

There is a consensus that mastering a discipline involves mastering the knowledge involved as well as mastering how language is used to communicate that knowledge (Lemke, 1990). For example, scientists tend to express accurate and objective information with abstract and subject-specific vocabulary, passive voice, and complex nominal groups in genres like protocol and explanation (de Oliveira, 2010; Halliday & Martin, 1993); similarly, historians tend to analyse historical events by presenting their "voices" with subject-specific vocabulary, past tense, and appraisal resources in genres such as historical narrative, explanation, and exposition (Coffin, 2006; Myskow, 2017). Such language used to communicate discipline knowledge is often called "academic language" or "disciplinary literacy", which generally refers to a set of linguistic registers (words, grammar, and organisational strategies) used to convey complex ideas and abstract concepts in scholarly subjects (Schleppegrell, 2009; Zwiers, 2008).

Mastering disciplinary literacy has been shown to be a difficult and lengthy process for native speakers, not to mention for L2 learners (Cummins, 2000; Martin, 2013), and thus some studies have examined whether and how L2 learners in EMI or CLIL develop the necessary disciplinary literacy. In Whittaker et al.'s (2011) longitudinal study on the English development of a group of junior secondary students in CLIL programmes in Spain, the researchers collected history texts written annually by the same group of students over a four-year period. Their texts analysis focused on the appropriate use of nominal groups to create coherence. Results showed improvement in the students' control of reference (e.g., an increase in textual cohesion via direct reference to other textual elements) and development in their nominal group complexity, both of which could be evidence of students' enhanced mastery of disciplinary literacy. In another longitudinal study on Spanish students learning content subjects in English, Whittaker and McCabe (2020) analysed student essays in different disciplines (history, ecology, and art) longitudinally from the end of primary school to the end of their second year of a CLIL secondary school programme. Focusing on whether students could develop the ability to use appropriate language resources to "evaluate" (Dalton-Puffer, 2013) in different disciplines, the researchers revealed that the CLIL students were slowly developing disciplinary knowledge and meaning-making resources, a wide range of lexico-grammatical structures such as abstract nouns, to perform the evaluative function in various disciplines. This study thus illustrates the development of both content and language, or in these researchers' terminology, "coupling" (Whittaker & McCabe, 2020, p. 14).

3. Disciplinary literacy and academic achievement

This "coupling" of content and language is crucial for bilingual education programmes, where students are expected to learn content through their L2. It has been argued that content and language "create a symbiotic relationship; that is, the learning of content contributes to the learning of language and a mastery of language gives learners easier access to content" (Stoller, 2008, p. 59). Hence, disciplinary literacy plays an important role in mediating students' academic achievement since students are very often required to express their content knowledge through the medium of language, particularly in high-stakes examinations (Authors, XXXX). A wealth of research has been published on the academic performance of immigrant students who are learning through their L2 in mainstream schools, and generally reported that immigrant students did not perform as well as their native-speaking counterparts in their academic achievement (e.g., Mathematics), probably owing to the lack of appropriate academic language (Brown, 2005; Levin & Shohamy, 2008).

Shaw (2012) and Shaw and Imam (2013) conducted detailed text analyses of the questions, instructions, and students' responses to an internationally recognised examination (the IGCSE), which is taken by students who speak different L1s. They identified the linguistic demands that various subjects (biology, geography, and history) imposed on candidates and explored how these may interact with assessment performance. It was observed that candidates' low scores in the examinations were primarily the result of deficiencies in their subject knowledge rather than linguistic hindrances. However, when it came to questions requiring more developed answers (e.g., essay-type questions), candidates without sufficient linguistic resources might not receive the maximum scores. In other words, to excel in examinations of content subjects in an L2, students must possess a certain level of academic language including subject-specific vocabulary, general academic vocabulary and beyond (e.g., linguistic resources to organise and present their ideas coherently).

Contextualised in a European CLIL programme, Coetze-Lachmann (2007) examined the subject-specific, task-based written performance of 10th graders in Germany studying geography in English. These students' written performance was evaluated based on six geography tasks that involved the application of high order subject-specific competences and the production of extended written responses. The results suggested that students had not mastered the thematic patterns related to the geography topics. Their argumentation skills

were also unsatisfactory in the sense that they failed to use appropriate subject-specific vocabulary and formal language when expressing their ideas.

The above studies demonstrate the integral role of language in content learning, particularly how bilingual students' L2 disciplinary literacy or academic language may interact with their academic performance. This may explain why some scholars have underscored the importance of disciplinary literacy instruction, which can be considered a form of "socially just teaching" as it provides students with access to highly specialised discourse communities (Moje, 2015, p. 259). However, students' L2 disciplinary literacy as measured in previous studies seems either somewhat arbitrary or focused on specific lexico-grammatical features of students' writing (e.g., subject-specific vocabulary, sentence patterns) which do not present a holistic picture. Adopting corpus-based multidimensional analysis, the present study examines bilingual students' mastery of disciplinary literacy at the end of their secondary education and further enquires as to whether and how disciplinary literacy is correlated with academic performance.

4. Multidimensional analysis of L2 writing

Corpus-based analyses have been widely adopted among the numerous EAP studies on academic writing (e.g., Gardner & Nesi, 2013; Lancaster, 2016). Corpora refer to "collections of naturally occurring language data, stored in electronic form, designed to be representative of particular types of text and analysed with the aid of computer software tools" (Nesi, 2016, p. 206). Corpus analyses not only offer quantitative descriptions of a certain group of texts but also yield qualitative insights into the use of specific registers in context (ibid.), and thus the multidimensional (MD) analytical approach was developed to quantify both (Biber, 2019). Through factor analysis in MD, Biber (1988) has proposed six dimensions of a single text (e.g., "involved and informational discourse", "narrative and non-narrative discourse", "context-independent and context-dependent discourse") which help to differentiate one text type from another. The rationale behind the MD approach is to investigate or compare different registers, groups of writers, disciplines and so on to reveal patterns of similarity and difference among the variants (Sardinha & Veirano Pinto, 2019).

Earlier studies adopting MD analysis have focused on the variations of spoken and written English genres (e.g., Biber, 1988; Friginal & Mustafa, 2017; Gardner, Nesi, & Biber, 2019), which has been extended to the same registers of different languages such as Spanish

(Asencion-Delaney, 2014) and Korean (Kim & Biber, 1994). In terms of student writing, Gardner, Nesi and Biber (2019) drew from 2,760 undergraduate student assignments in the BAWE (British Academic Written English Corpus) and found that academic disciplines, student level, and genre families could be mapped with MD analysis and measured together with their different dimensions. For example, biology students' assignments tended to contain more compressed procedural information than stances towards others' works, described completed events slightly more than possible events, expressed projections of personal stance infrequently, and demonstrated high informational density. Furthermore, the higher the level of the student (e.g., postgraduate vs. undergraduate), the more likely that writer was to include compressed procedural information, lack projection of personal stance, and intensify information density.

Despite the above trends, discrepancies in register use among L2 learners of different proficiency levels have also been examined. For instance, in a corpus of 5,200 argumentative essays of non-native English speakers of four proficiency levels (from beginners to advanced), Kim and Nam's (2019) MD analysis suggested that advanced L2 learners' writings were closer to those of native speakers (i.e., containing more descriptive informational discourse and less involved narration), but were fundamentally different from those of less advanced L2 writers (i.e., more involved narration, less informational). Additionally, Weigle and Friginal's (2015) study compared the lexico-grammatical features of TOEFL essays and successful disciplinary writings in the MICUSP (Michigan Corpus of Upper-level Student Writing) and found that compared with disciplinary writings, timed exam essays embraced "more narrative, expressions of opinion and stance, situation-dependent discourse, and expressions of possibility" (p. 34). The dimension scores for L2 learners' written output on the TOEFL test were also correlated with their overall TOEFL exam scores and indicated that more proficient L2 learners could better expand their linguistic repertoires to include academic writing conventions.

These findings indicate that register use is subject to a variety of factors or conditions such as learners' proficiency levels, discourse domains, disciplines, and test essays versus disciplinary writings. To develop a more thorough understanding of the linguistic features of L2 writing, it is pivotal to investigate a wider range of texts produced by diverse groups of writers under varied conditions. However, most of the existing studies on student writing which adopt the MD approach involve only university student-produced academic written discourse (from popular corpora such as MICUSP), while the register use of secondary

students' written discourse, especially those enrolled in bilingual education programmes, has been under-explored. This student population is unique because they are on the developmental trajectory of academic enculturation, learn content subjects and take high-stakes examinations in an L2 (very often in English) while simultaneously having few opportunities to use the target L2 outside the classroom. Therefore, investigating their register use in high-stakes subject-content examination essays with the MD approach can offer insights into how this group of writers expresses their ideas with the target language, which in turn sheds light on their overall content and language learning. With this overarching aim, this study performs MD analysis on examination essays produced by EMI secondary school students in Hong Kong and explores the correlations between their mastery of disciplinary literacy and academic performance.

5. Method

5.1. Corpus sample

The learner corpus analysed in this study was constructed with data provided by the Hong Kong Examination and Assessment Authority (HKEAA). This learner corpus included the final essay question of the biology examination from 2012-2015 in the Hong Kong Diploma of Secondary Education (HKDSE) examination, which is the high-stakes university entrance examination taken by 12th Grade (17-18-year-old) students in Hong Kong secondary schools. The HKDSE biology examination consists of multiple-choice questions, structured questions which require short answers, and one essay question. This learner corpus only included students' responses to the essay question because it constituted the most coherent discourse for MD analysis. Eleven points are allocated to each essay question, and there is no word limit. The genres involved in the four selected years included comparative report (2012), explanation (2013) and discussion (2014 and 2015). There may be differences in their structure and communicative purposes, but all these essay questions required students to describe scientific concepts and explain how they are different from or related to other phenomena. A total number of 240 students' scripts from this period (2012-2015), 60 from each year, were randomly selected for our stratified sample. Each cohort consisted of texts that were assigned an overall biology subject grade ranging from U (lowest) to 5** (highest). However, some scripts, especially those written by students at the lowest level (U), were excluded from this corpus because they did not reach the threshold word number for MD

analysis (see section 4.3). The total number of texts included was 204, with a total word count of 45,823 words and an average length of 225 words per text (see Table 1 for details).

Table 1. Characteristics of the corpus

Level	No. of texts	Total no. of words	Mean length of a text
U	4	831	208
1	28	4132	148
2	33	4984	151
3	39	8748	224
4	40	9690	242
5	21	6060	289
5*	19	5239	276
5**	20	6139	307
Total	204	45823	225

This study sought to explore relationships between the disciplinary literacy and academic performance of bilingual students. The latter was operationalised as the overall subject grade those students achieved in the HKDSE and warrants more explanation here. The biology subject grade (ranging from U to 5**) is made up of the timed public examination (80%), from which the essays of this learner corpus were taken, and school-based assessment (20%). For the determination of the overall subject grade, standards-referenced reporting is adopted, meaning that students' levels of performance are reported with reference to a set of standards as defined by cut scores on the grading scale for that subject. There are five cut scores distinguishing five levels of performance (1–5), with 5 being the highest. A performance below the cut score for Level 1 is labelled as "Unclassified" (U). To further discriminate for university selection purposes, approximately the highest-achieving 10% of Level 5 students are categorised as 5** and the next 30% of Level 5 students are classified as 5* (CDI & HKEAA, 2015; HKEAA, 2011). For each of the five levels, a set of written descriptors developed by the examination authority describes what the typical student at that level can do. These descriptors include knowledge of biology concepts, cognitive skills (e.g., apply, synthesise, evaluate) and communicative skills.

5.2. Participants

Further information about this group of students and their educational background is useful in better understanding this study's results. Both Chinese and English are the official languages in Hong Kong, which is a former British colony and is now a special administrative region of China. Most primary schools in Hong Kong adopt Chinese (spoken Cantonese and written standard Chinese) as the medium of instruction (MoI). Secondary schools have some flexibility in determining their MoI for content subjects in junior grades (Grades 7-9), providing they fulfil some criteria including teachers' capacity, students' capacity, and the support measures of the school. Owing to the high socio-economic status of English, parents generally prefer schools using English as the MoI (i.e., EMI education). Around a quarter of secondary schools adopt English as the MoI for most students and subjects, while other schools often adopt English as the MoI for 1-2 subjects, with science being one popular subject taught in English. Senior secondary schools (Grades 10-12) have even more flexibility in their MoI arrangements out of consideration for their students' upcoming transition to EMI universities. All secondary 6 students (Grade 12) will take the HKDSE as the university entrance examination. For each subject in the HKDSE, except Chinese-related ones (e.g., Chinese language, Chinese literature, Chinese History) and English language, there are both Chinese and English versions of examination papers. Students will take these assessments according to the language through which they have studied each subject. Hence, by the time students take the HKDSE, they are 17-18 years old and should have studied English as a subject for at least 12 years. For those who take the content subject examinations in English, they should have experienced at least three years of EMI.

For biology, according to the HKEAA reports, around half of the candidates took the exam in English between 2012 and 2015. The content subject of biology was chosen for this study because it is offered by over 90% of secondary schools in Hong Kong and is regarded as the most language-oriented of the science topics.

5.3. Data analysis: Multidimensional Analysis Tagger

The exam essays were analysed using Nini's (2015) Multidimensional Analysis Tagger (MAT), which replicates Biber's (1988) MD analysis but can accommodate texts shorter than 400 words. We conducted a grammatical analysis to identify lexico-grammatical items of each text with the Stanford Tagger (Toutanova, Klein, Manning, & Singer, 2003), which generated part-of-speech tags (e.g., coordinating conjunction, personal pronoun, singular or plural norm). The software then expanded the scope of the Stanford Tagger tags and

identified 68 linguistic features from Biber's (1988) list, such as amplifiers (*completely*, *very*), private verbs (*accept*, *discover*), and second person pronouns (*you*, *your*). The occurrences of all these linguistic features were normalised to frequencies per 50 words since 50 words was set as the threshold for the inclusion of essays in the data analysis procedure[1]. MAT automatically computed the linguistic features in the corpus texts to generate scores for Biber's (1988) six functional dimensions, and the z-score of each feature. Therefore, the MAT software produced the following data sets: (1) normalised frequencies of all the linguistic features, (2) the z-score of each feature, (3) the score of each functional dimension, and (4) two sets of graphs that compared the linguistic features of the (sub-)corpus with eight written and spoken genres analysed in Biber's (1995) study (i.e., official documents, press reportage, academic prose, broadcasts, general fictions, prepared speeches, personal letters, and conversations) or eight text types (i.e., intimate interpersonal interaction, informational interaction, imaginative narrative, general narrative exposition, scientific exposition, learned exposition, situated reportage, and involved persuasion) to determine the closest text type to the (sub-)corpus. Parametric tests (ANOVA) and non-parametric tests (Kruskal Wallis tests) were conducted in SPSS to further examine whether differences existed among students with different overall subject achievement levels (from U to 5**).

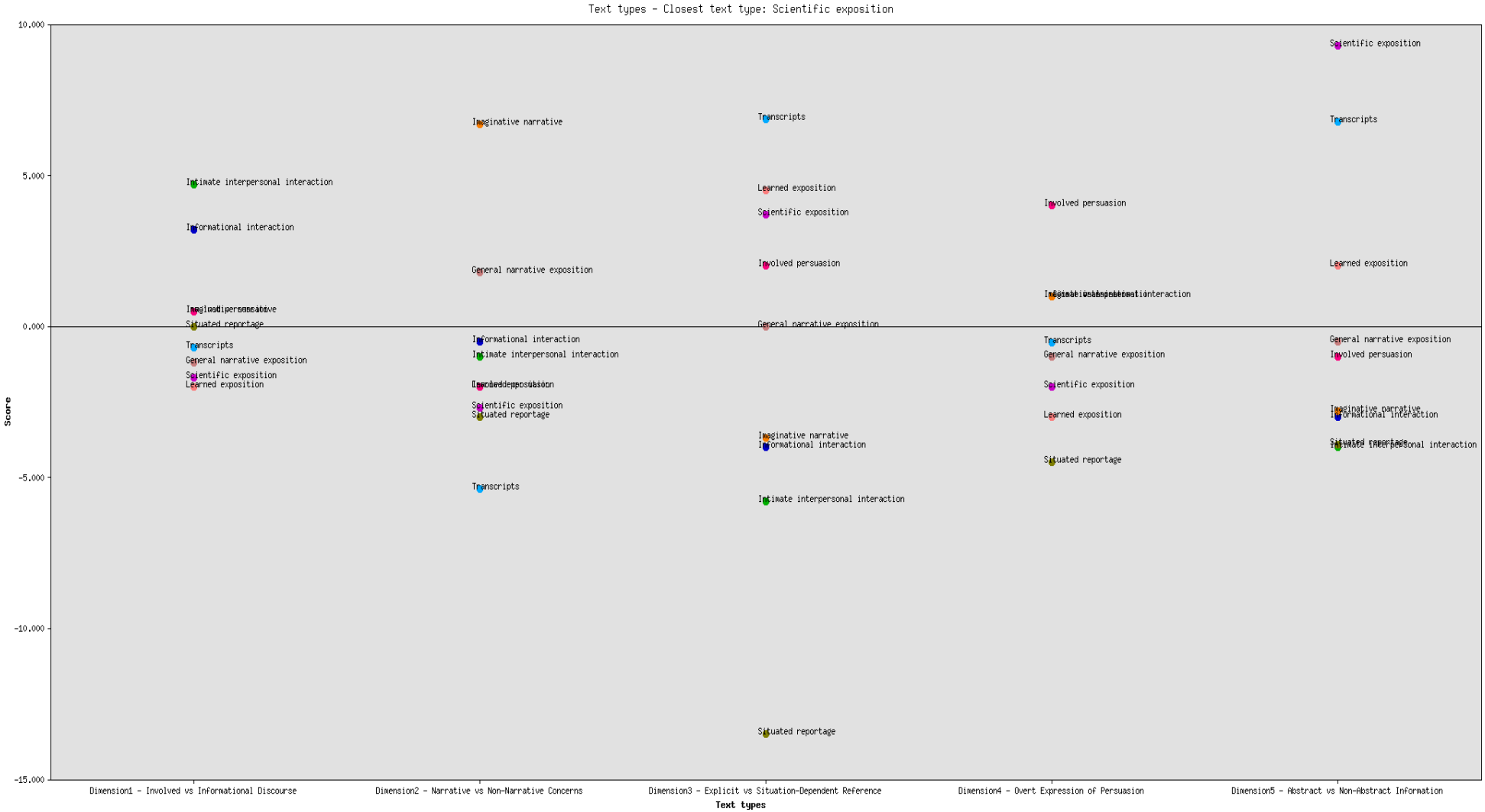
6. Findings

6.1 Overview of the corpus

We used the MAT software to analyse the linguistic features of the corpus, which automatically generated the score for each dimension. Figure 1 shows the output of the MAT analysis, comparing the dimensional scores of the whole corpus (the blue dots labelled "transcripts") with those of the eight text types. The different coloured dots reveal the mean dimensional scores of the text types. Given that the learner corpus of this study had a low score on D1, high score on D3, and high score on D5, the closest text type for the corpus is scientific exposition, which matches with the academic prose and official documents categories from Biber (1989). The results for each dimension are further elaborated below.

[1] The researchers read through all the essays and agreed that texts equal to or longer than 50 words could fully address the test questions, while those with fewer than 50 words or lacking content could not appropriately answer the exam questions. Therefore, 50 words was determined as the threshold for data analysis.

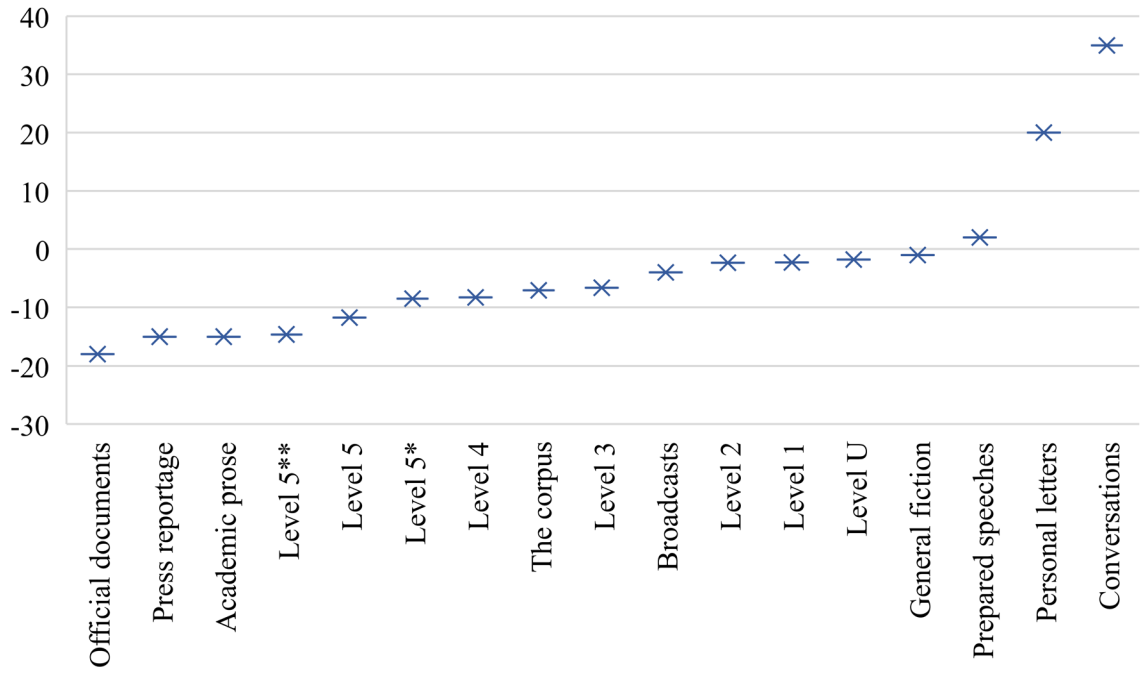
Figure 1. The closet text type



6.2 Dimension 1

Dimension 1 concerns the contrast between involved and informational discourse. Low scores indicate text informational density closer to academic prose, whereas texts with high scores contained more affective and interactional language closer to spoken discourse. Figure 2 shows a comparison of mean D1 scores between the entire learner corpus, the sub-corpora of different score levels, and other registers analysed in Biber's (1995) cross-linguistic studies of register variation. According to the MAT results, the essays of students with higher level grades (levels 3-5**) tended to share more similarities with academic prose while those of students with lower level grades (levels U-2) were closer to broadcast discourse and general fiction.

Figure 2. Dimension 1- Involved versus Informational Production



This trend was also corroborated by the significant differences among mean D1 scores of different levels revealed by a Kruskal Willis test (Table 2). Table 2 depicts the mean scores and standard deviations (in brackets) of the essays at each level. Follow-up Mann-Whitney tests, which compared the mean scores of any two groups, revealed significant differences in the scores between level 5** students and their counterparts of the five lower bands (from U

to 4) ($p < .0083$[2]), and between level 5 students and their levels 1 and 2 peers ($p < .0083$) although no significant differences were found for the other comparisons ($p > .0083$).

Table 2. Statistical analyses of Dimension scores across subject grade levels

	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
U	-1.81 (5.80)	-5.28 (2.26)	6.19 (3.31)	4.27 (8.17)	0.27 (3.25)	0.92 (2.98)
1	-2.29 (11.17)	-5.75 (1.90)	5.69 (5.31)	0.48 (5.49)	3.02 (7.67)	-1.64 (2.24)
2	-2.35 (10.94)	-6.12 (1.98)	6.62 (6.37)	-0.53 (5.47)	4.47 (9.37)	-2.07 (2.19)
3	-6.62 (9.24)	-5.43 (1.77)	6.03 (3.66)	0.30 (4.29)	4.71 (6.96)	-2.78 (0.97)
4	-8.29 (6.69)	-5.36 (1.92)	7.19 (5.44)	-1.11 (4.45)	7.83 (7.96)	-1.90 (2.04)
5	-11.75 (5.73)	-5.17 (1.92)	6.34 (3.66)	-0.90 (4.58)	8.22 (9.59)	-2.59 (1.20)
5*	-8.52 (7.83)	-4.55 (2.62)	8.78 (4.69)	-1.51 (3.27)	11.89 (9.92)	-1.85 (2.23)
5**	-14.65 (4.91)	-4.73 (2.16)	8.86 (5.98)	-2.13 (3.19)	12.60 (9.03)	-2.79 (1.25)
Chi-Square	37.119	25.495	8.578	7.368	32.264	15.881
p	< .001*	.030	.284	.392	< .001*	.026

The variations in mean D1 scores of the different levels could be further explained by the differences in the texts' linguistic features. Biber (1988, cited in Kim & Nam, 2019) listed 23 positive linguistic features and five negative ones which were loaded on Dimension 1, and ANOVA or Kruskal Willis tests were conducted with each of these linguistic features[3]. The differences in D1 scores could be attributed to significant differences in (1) average word length ($F = 5.321$, $p < .001$), (2) type-token ratio (Chi-Square = 74.936, $p < .001$), (3) attributive adjectives ($F = 4.554$, $p < .001$), (4) total prepositional phrases ($F = 2.089$, $p = .046$), (5) demonstrative pronouns (Chi-Square = 20.255, $p = .005$), and (6) the pronoun *it* (Chi-Square = 26.832, $p < .001$).

Table 3 presents the means and standard deviations (in brackets) for each of the above features across the eight levels. Word length (mean length of the words in the text in orthographic letters) and type-token ratio (dividing the number of word types by the total number of words) demonstrate high information density according to Biber (1995). Longer

[2] Bonferroni correction was conducted, so the alpha level was adjusted to .0083.
[3] ANOVA tests were conducted for normally distributed data sets, whereas Kruskal Willis tests were performed for data that did not follow normality.

words tend to bear more specific meaning than shorter ones (Zipf, 1949; cited in Biber, 1995), and a higher type-token ratio, demonstrating the use of different word types, indicates careful word choices and precise presentation of information. Table 3 shows a trend of increasing word length from levels 1 to 5** with the exception of level U. Levels 4-5** students tended to present information more densely than their level 1-2 counterparts ($p<.05$). Furthermore, since attributive adjectives (i.e., adjectives followed by another adjective or a noun) are employed to elaborate nominal references (Biber, 1995), the increasing mean frequency of attributive adjectives from levels 1-2 to the higher levels also reflected the more informative nature of the texts written by higher-level students. Prepositional phrases (e.g., *beside the point*, *in fact*) also specify and elaborate referential identity (Biber, 1995), and thus more frequent use of such prepositional phrases (at levels 5-5**) indicates a text that is more informational in nature. Compared with the above four positive features of informational texts, pronouns reflect the interactive, involved, or affective feature of texts. Decreases in the prevalence of the pronoun *it* were found from the texts of lower levels (levels U-4) to those of higher ones (levels 5-5**).

Table 3. Linguistic features loaded on Dimension 1

Level	Average word length	Type-token ratio	Atributive adjective	Total prepositional phrases	Demonstrative pronouns	Pronoun *it*
U	4.71 (0.37)	51.25 (29.04)	8.75 (1.99)	10.82 (3.35)	0.45 (0.34)	1.75 (0.57)
1	4.63 (0.33)	36.55 (12.49)	6.91 (2.63)	10.90 (3.28)	0.12 (0.29)	1.91 (1.53)
2	4.66 (0.34)	35.08 (13.15)	6.45 (3.07)	11.01 (4.15)	0.22 (0.39)	1.35 (1.22)
3	4.80 (0.34)	50.23 (13.84)	8.77 (2.45)	10.83 (2.51)	0.19 (0.27)	1.56 (1.30)
4	4.83 (0.38)	50.94 (13.70)	8.80 (2.98)	11.39 (2.26)	0.34 (0.60)	1.12 (0.94)
5	4.91 (0.31)	63.48 (19.17)	8.49 (2.39)	13.02 (2.16)	0.26 (0.41)	0.91 (0.85)
5*	5.02 (0.41)	56.24 (11.83)	8.45 (2.73)	12.12 (2.83)	0.51 (0.47)	0.98 (1.06)
5**	5.11 (0.30)	67.88 (13.64)	9.79 (2.17)	12.67 (1.87)	0.38 (0.34)	0.41 (0.42)

Excerpt 1 (level U) and Excerpt 2 (level 5**) below can provide some insights into these findings. The Level U student primarily used be-verbs as the main verbs (a positive measure of D1), whereas the Level 5** student included more types of verbs. Furthermore, in Excerpt

2, the student's efforts to seek synonyms are noticeable, for example "*but,*" "*while,*" and "*whereas*" were employed to present contrastive relationships. In addition, Excerpt 2 contains longer words (e.g., *homologous*, *chromosomes*, *separation*), which are not found in Excerpt 1. Thus, it appears that the level 5** student is not only more knowledgeable about biology but is also a more competent L2 user with more extensive linguistic resources.

Excerpt 1:

To ensure the continuity of life, mitosis and meiosis ***are*** important processes. [There][4] ***are*** two different process, mitosis and meiosis. These process[es] ***are*** ensured in the count[n]ious of life. If they are [disappear] then they will not hope for life. (Level U; file no.: 2012_37)

Excerpt 2:

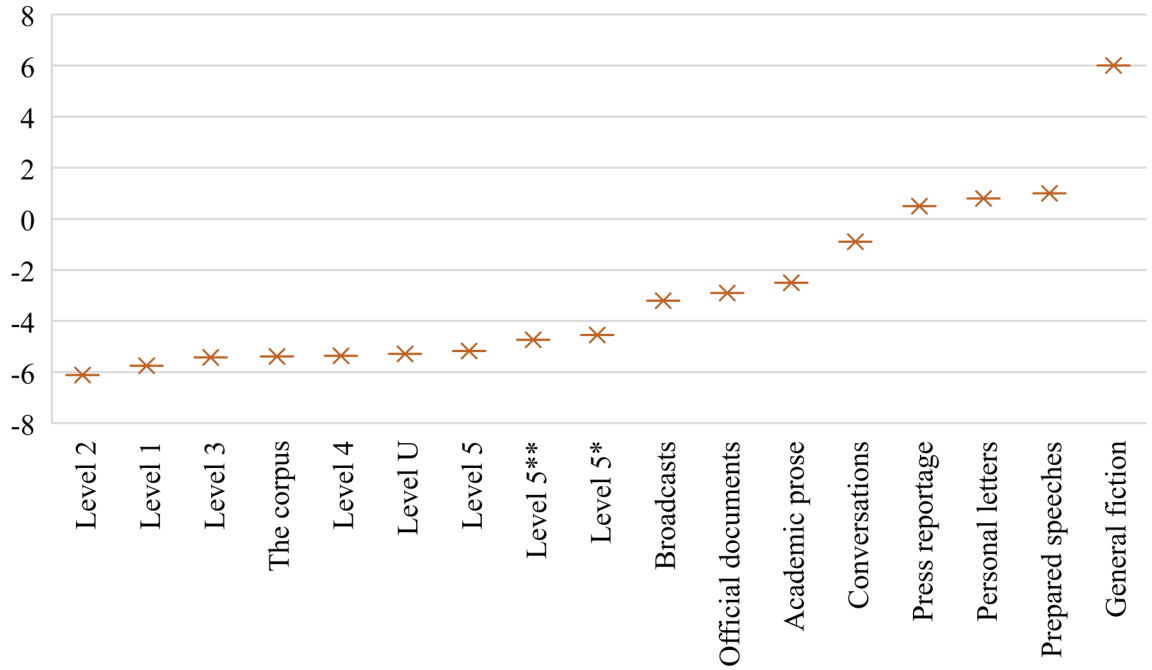
For the differences between the 2 processes, firstly, the pair of homologous chromosomes ***do*** not pair up at the equator during mitosis, <u>but</u> they ***do*** during meiosis. Secondly, mitosis only ***involves*** the separation of identical chromatids in a chromosome, <u>while</u> meiosis ***involves*** both the separation pairs of homologous chromosomes and identical chromatids. Thirdly, mitosis ***produces*** daughter cells that ***are*** genetically identical to the parent cell <u>whereas</u> meiosis ***produces*** daughter cells that ***are*** genetically different from the parent cell. (Level 5**; file no.: 2012_05)

6.3 Dimension 2

Dimension 2 concerns the degree to which a text is narrative, with high scores suggesting more features of narration. Figure 3 compares the mean D2 scores of the corpus, sub-corpora, and other registers. It shows that the essays in our learner corpus, regardless of the level grade assigned, are non-narrative in nature. No significant differences were found in D2 scores across different levels (Table 2).

[4] When any spelling and grammatical mistakes were noticed in text, we revised them accordingly and the revised forms were displayed in brackets [ ].

Figure 3. Dimension 2-Narrative versus Non-narrative Discourse



Although all the groups obtained similar D2 scores, significant differences were found (Chi-Square = 20.663, p = .004) in a follow-up analysis of the use of public verbs, i.e., actions that can be observed publicly or introduce indirect (and reported) statements (e.g., *say*, *explain*), as shown in Excerpt 3. This is a positive indicator for the narrative nature of texts. Table 4 reports the means and standard deviations (in brackets) for public verbs, which demonstrates that the mean frequency for this feature is relatively low in texts of all levels. Despite this, students of the lower levels (1-2) used public verbs less often than higher-level students (3-5**) did (p < .01).

Table 4. Public verbs (Dimension 2)

Level	U	1	2	3	4	5	5*	5**
Mean (SD)	0	0.07 (0.22)	0.05 (0.14)	0.16 (0.23)	0.16 (0.24)	0.23 (0.35)	0.30 (0.58)	0.17 (0.21)

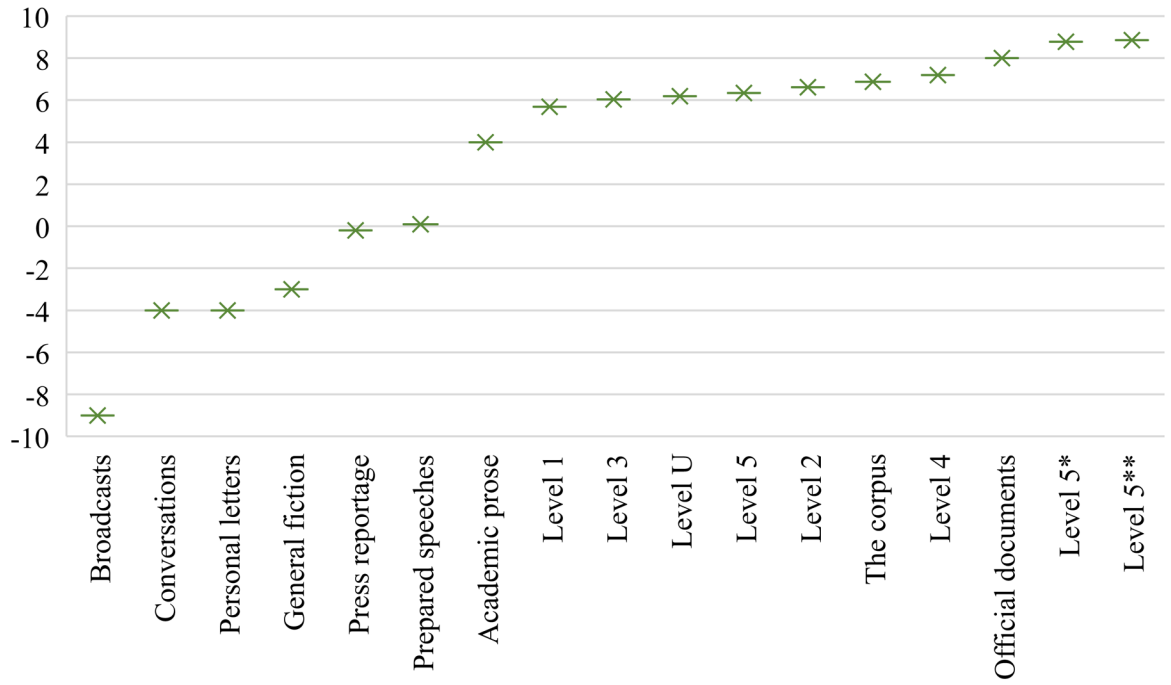
Excerpt 3:

Protein molecules can have different conformations and different functions. Let me ***explain*** in the followings. (Level 3; file no.: 2014_20)

6.4 Dimension 3

Dimension 3 reflects whether a text is context-independent or context-dependent, with higher scores referring to the former, which is closer to academic prose. Figure 4 demonstrates that the essays are generally context-independent (positive scores), and the mean scores for the sub-corpora are close to that of academic prose and official documents. The mean D3 scores for the eight subject level groups were not significantly different from each other (Table 2).

Figure 4. Dimension 3-Situation-dependent versus Elaborated Reference



Concerning the linguistic features loaded on D3, significant differences among the student groups were reported in total adverbs ($F = 2.254$, $p = .032$) and nominalisations ($F = 4.832$, $p < .001$). Table 5 depicts the means and standard deviations (in brackets) of each feature across the groups. Adverbs, which can be used as references to time, place, manner and so on, are context-dependent, as the understanding of the adverbs requires the knowledge of the surrounding environments or contexts (Biber, 1995). In this corpus, level 2 students employed fewer adverbs than their peers ($p < .05$), reflecting that they were less likely than others to depend on the situational context when writing. In contrast, the more nominalisations the students adopted (i.e., nouns ending in *-ment*, *-tion*, *-ness*, and *-ity*, and the plural forms), the more explicit and context-independent the discourse was. Level 1-3

students produced less context-independent discourse with fewer nominalisations than those of highest bands (5* and 5**).

Table 5. Linguistic features loaded on Dimension 3

	U	1	2	3	4	5	5*	5**
Nominalisations	1.41 (1.28)	0.65 (0.54)	0.75 (0.77)	0.93 (0.76)	1.13 (0.66)	1.28 (0.58)	1.51 (0.81)	1.48 (0.83)
Total adverbs	1.97 (0.73)	1.57 (0.82)	1.00 (0.68)	1.42 (0.69)	1.34 (0.66)	1.55 (0.63)	1.42 (0.56)	1.41 (0.88)

An example is provided to supplement the above findings. Excerpt 4 is taken from a level 5** student's text whose D3 score was 13.13. The nominal elaboration features which make the text more context-independent have been tagged. On the other hand, Excerpt 5 (from a level U student) has a much lower dimension score (6.59) as it includes some context-dependent markers (as underlined).

Excerpt 4:

Lean meat contains high ***proportion*** of protein and low ***proportion*** of fat and sugar. According to the food pyramid, fat and sugar are at the top of the pyramid, meaning we should consume the least because high ***consumption*** of fat may cause ***obesity***, ***deposition*** of cholesterol on wall of blood vessels leading to coronary heart disease and hypertension, while high sugar intake may also cause ***obesity*** and diabetes. (level 5**; file no.: 2014_01)

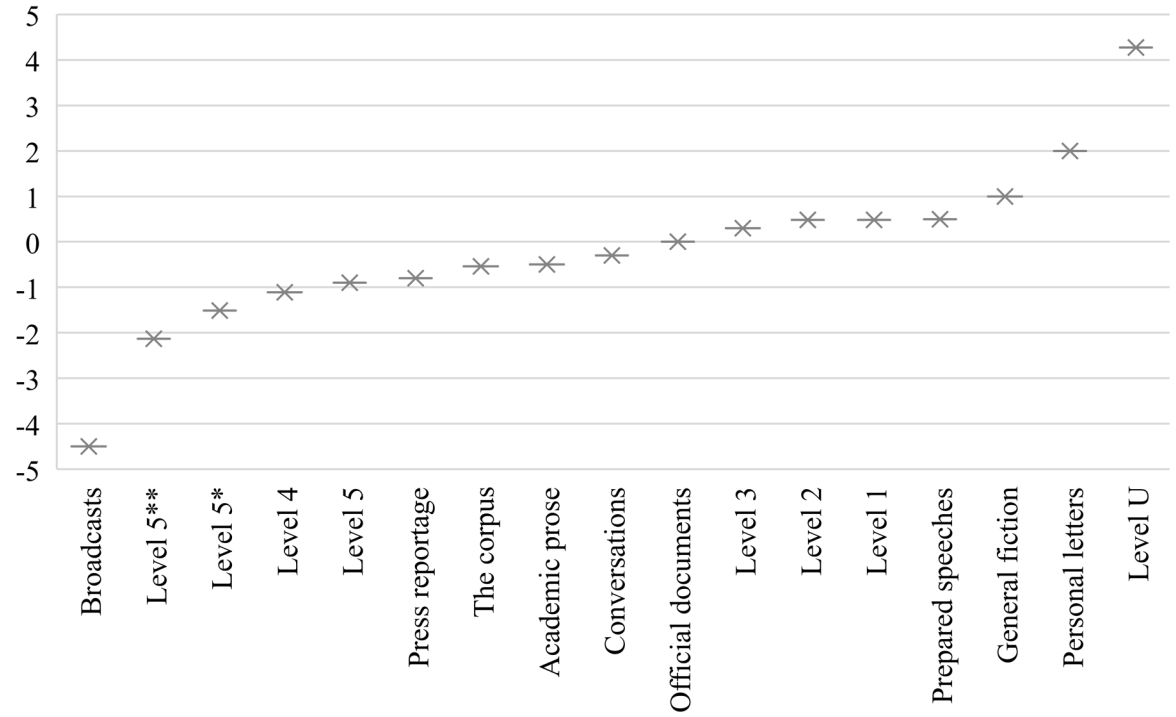
Excerpt 5:

In the [year] of 21st century. ***Nowadays*** peoples have very health awareness. Some people have adopted a diet rich in lean meat in order to close weight and build muscle. They may be able to achieve these aims but they are some health problems associated with this diet…***First*** we people have no healthy lifestyle for example only consume lean meat in order to lose weight and build muscle. (Level U; file no.: 2014_7)

6.5 Dimension 4

Regarding Dimension 4, overt expression of persuasion, high scores indicate that the writer's point of view and assessment of likelihood/certainty were explicitly marked. The dimension score is positively correlated to the presence of seven linguistic features: infinitives, prediction modals, conditional subordination, suasive verbs, necessity modals, split auxiliaries, and possibility modals (Biber, 1995). While in professional letters and editorials authors' viewpoints and assessments of probability are frequent, academic prose and official documents are unmarked regarding this dimension (Biber, 1995). In Figure 5, although the mean score of the level U group was the highest, those of other groups were near 0 or even negative, which indicates that the students seldom projected their viewpoints or stance when addressing these examination questions. No differences ($p > .0083$) were spotted among the groups (Table 2).

Figure 5. Dimension 4-Overt Expression of Argumentation



When breaking down the different linguistic features, the occurrence of suasive verbs most significantly differed among the groups (Chi-Square = 18.781, $p = .009$). Suasive verbs are verbs "which intend to effect a change of some sort" (Quirk et al., 1985; also cited in Grant & Ginther, 2001, p. 131) such as *suggest* or *recommend*. Excerpt 6 contains an example of a suasive verb which is also considered a marker to express the test-taker's stance (Biber, 1988), in this case their assertiveness of the necessity of proteins. Mann-Whitney test results

showed that level 2 students included fewer suasive verbs than their level U peers and those of higher levels (3-5**) ($p < .05$) (see Table 6). Less frequent uses of suasive verbs were found in texts of 4 and 5** groups than in those of the U group. Other relevant linguistic features such as infinitives, prediction modals, necessity modals, and conditional subordination can be found in texts to express persuasion, but with these features, no obvious group differences were found ($p > .05$).

Table 6. Suasive verbs (Dimension 4)

Level	U	1	2	3	4	5	5*	5**
Mean	0.52	0.11	0.07	0.22	0.24	0.23	0.24	0.19
(SD)	(0.56)	(0.23)	(0.18)	(0.29)	(0.34)	(0.33)	(0.33)	(0.17)

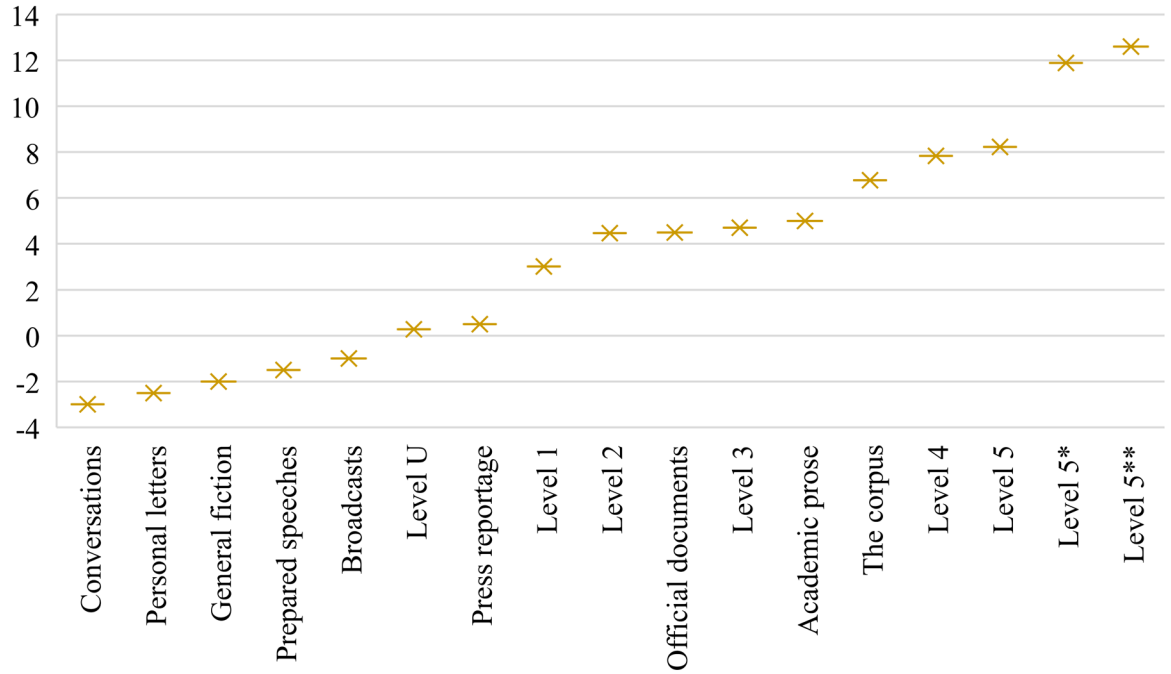
Excerpt 6:

Thus, having [a] diet rich in lean meat can help maintain a constant rate of protein metabolism. For the bad points regards to the protein metabolisms, proteins are ***required*** in a small proportion compare[d] to other nutrients. (Level 3; file no.: 2014_50)

6.6 Dimension 5

Dimension 5 discusses whether a text contains abstract or non-abstract information, with high scores suggesting that the text "provides information in a technical, abstract and formal way, as for example in scientific discourse" (Biber, 1988; cited in Nini, 2015, p. 5). Figure 6 shows that the essays of level U students were closer to press reportage, whereas the others contained features more like official documents or academic prose. As seen in Table 2, the D5 scores of the eight groups significantly varied. The mean scores increased as students received higher level scores in the examination, implying a pattern in which more abstract information was included in texts written by the higher-level groups. Follow-up Mann-Whitney tests revealed variations between levels U-3 and 5** ($p < .0083$), and between Levels U-3 and 5* ($p < .0083$). No obvious differences were observed concerning other comparisons ($p > .0083$).

Figure 6. Dimension 5-Abstract versus Non-abstract Style



The variations across the groups could be due to the significant differences in the prevalence of five linguistic features: (1) by-passives (Chi-Square = 20.467, p = .005), (2) conjuncts (Chi-Square = 17.007, p = .017), (3) agentless passives (Chi-Square = 25.830, p = .001), (4) other adverbial subordinators (Chi-Square = 16.237, p = .023), and (5) past participial WHIZ deletion relatives (Chi-Square = 20.831, p = .004). Their means and standard deviations (in brackets) are presented in Table 7. Higher-level students (5-5**) used more other adverbial subordinators (e.g., *since*, *whereas*) than lower level students (1-2) (p < .05). Furthermore, agentless passives appeared more frequently in level 5-5** writings than in level 1-2 ones (p < .05). Levels 5 and 5* students also deployed more by-passives (agentless passives+ preposition *by*) than those in levels 1-2 (p < .05). In addition, level 5** students used more conjunctions than levels 1-2 (p < .05). Although past participial WHIZ deletion relatives (e.g., The solution *processed by this process*…; Nini, 2015, p. 31) rarely occurred in the texts, Levels 5** and 5* students were the most likely to employ this structure in their essays. According to Biber (1995, p. 164), academic prose and official documents, which are most marked by D5 features, contained frequent use of conjuncts, main clause passive constructions (i.e., agentless and by-passives), and dependent clause passive constructions. From our research findings, it can be inferred that higher-level students (5-5**) were more likely to produce essays in a technical, abstract, and formal way.

Table 7. Linguistic features loaded on Dimension 5

Level	By-passives	Conjuncts	Agentless passives	Other adverbial subordinators	WZPAST
U	0	0.21 (0.24)	0.34 (0.21)	0	0.09 (0.11)
1	0.06 (0.16)	0.23 (0.37)	0.32 (0.35)	0.17 (0.31)	0
2	0.03 (0.14)	0.30 (0.43)	0.34 (0.67)	0.17 (0.34)	0.08 (0.17)
3	0.10 (0.19)	0.29 (0.31)	0.37 (0.36)	0.19 (0.31)	0.07 (0.16)
4	0.10 (0.20)	0.35 (0.31)	0.52 (0.49)	0.29 (0.38)	0.07 (0.14)
5	0.17 (0.20)	0.34 (0.32)	0.60 (0.29)	0.30 (0.51)	0.09 (0.21)
5*	0.13 (0.18)	0.46 (0.34)	0.62 (0.42)	0.40 (0.54)	0.13 (0.16)
5**	0.12 (0.23)	0.48 (0.24)	0.66 (0.43)	0.42 (0.47)	0.12 (0.17)

To illustrate texts containing abstract and non-abstract information, excerpts from levels 2 (Excerpt 7) and 5* (Excerpt 8) students are provided below. Excerpt 7 is a typical example of the texts of the lower-level groups where linguistic features for abstract discourse were seldom found; statements were written in the active voice. In comparison, in Excerpt 8, the level 5* student employed agentless passives (marked as PASS), one by-passive (BYPA), and one conjunction (CONJ) to express or link the ideas, thus making the text more abstract.

Excerpt 7:

For advantage, people can get enough fat from the [lean] meat and the amount is enough for one day activity. It can ensure that they will not eat excess amount. It can help them to maintain a good protein metabolism process. For disadvantage, people can't get the other nutritional needs for their daily live, like vitamin C, which may make them suffer from [other] disease. (Level 2; 2014_18)

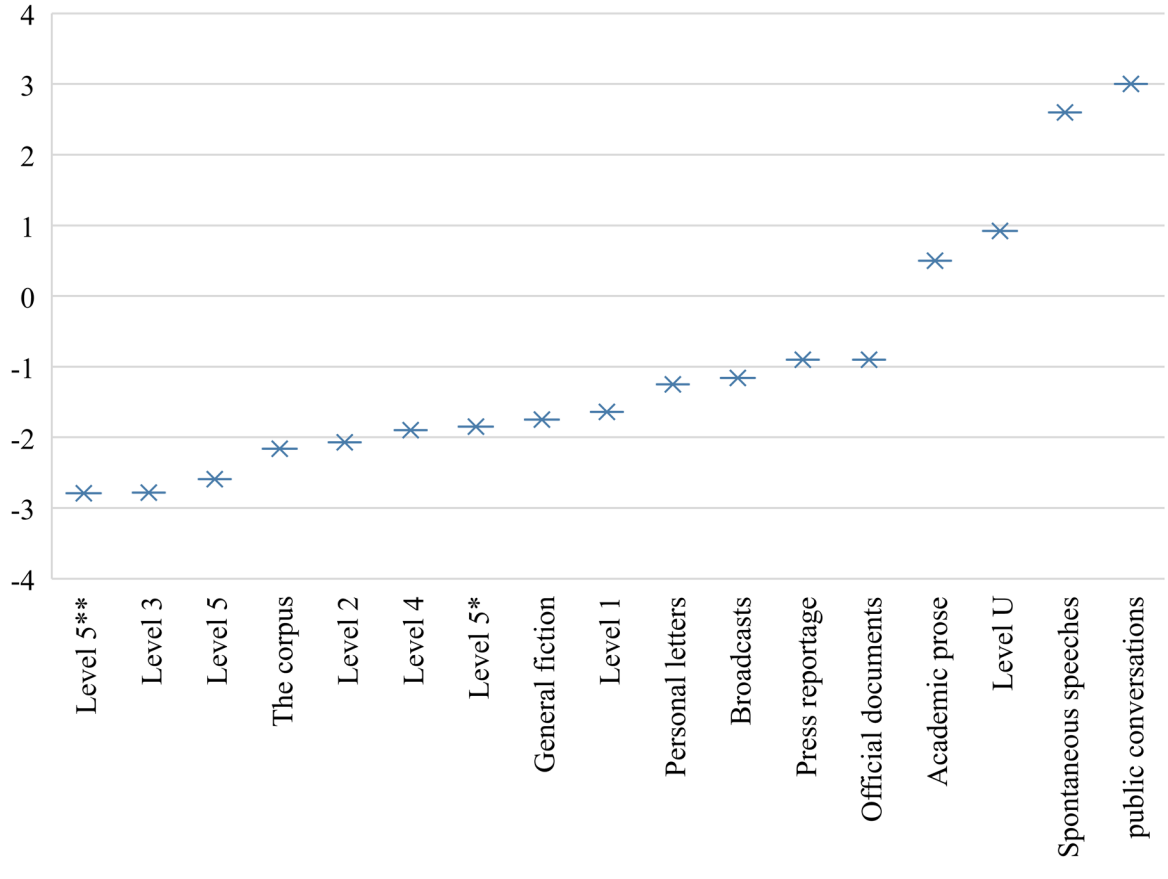
Excerpt 8:

Besides, the protein will ***be broken down to*** [PASS] amino acid in the body. Excess amino acid[s] will ***be deaminated*** [PASS] and some parts will ***be converted to*** [PASS] carbohydrate[s] and lipid[s] to provide energy to the body. ***Moreover*** [CONJ], proteins ***are [converted] to*** [PASS] amino acid[s] which provide us with certain type[s] of amino acid which cannot ***be produced by*** [BYPA] ourselves. (Level 5*; 2014_14)

6.7 Dimension 6

Dimension 6 indicates on-line informational elaboration. If a text is informational but is produced under time pressure (e.g., prepared and spontaneous speeches), high scores are observed. As seen in Figure 7, informational spoken registers, such as prepared speeches and public conversations, obtain high D6 scores while other spoken registers are unmarked regarding this dimension. D6 features such as that-complement clauses on verbs, on adjectives, and that-relative clauses on object positions are not common in academic prose and official documents (Biber, 1995). Except for level U, the entire corpus and its sub-corpora obtained negative D6 scores. This may imply that although the corpus was generated during an examination, the students planned deliberately when composing the texts, and the texts were academic. Furthermore, we did not find significant differences in the mean dimensional score of students at different levels ($p > .0083$; see Table 2).

Figure 7. Dimension 6-On-line Informational Elaboration Marking Stance



Four linguistics features are positively loaded on this dimension – *that* complement clauses on verbs and on adjectives (e.g., He never expected *that she would come*), *that* relative

clauses on object position (e.g., the cat *that I saw*), and demonstratives (e.g., *these*, *this*). For these features, their normalised occurrences in texts were similar among all the groups ($p > .05$). Excerpt 9 provides an example of how demonstratives were used.

Excerpt 9:

These [processes] have different [functions]. And they have different values. … ***This*** process usually [takes] place in the human life. (U Level text; 2012_37_involved persuasion)

To summarise, while the corpus with low D1 and high D3 and D5 scores generally shares similar linguistic features to academic prose and official documents, variations were found among essays written by students at different levels. Higher-level scoring students tended to perform better in D1 and D5 than others, which indicates that they were more likely to produce abstract informational discourse. No significant differences among the eight groups were found in D2, D3, or D4 scores.

7. Discussion

This study examined the essays written by a stratified sample of students studying through EMI in Hong Kong secondary schools. It is unique in two ways. First, it employed MD analysis with MAT to provide holistic text analyses. Instead of focusing on the use of specific linguistic features, MD analysis and MAT enable analyses of various linguistic features which together contribute to different dimensions and registers (Biber, 2019). It is interesting to note that although the texts included in this learner corpus were written under time pressure by grade 12 students, they demonstrated features closer to scientific exposition than general narrative exposition or informational interaction. In other words, the students in this study, who had studied biology through an L2 (English) for at least 3 years, were aware of the features of academic prose required in the discipline. They refrained from producing non-academic texts (e.g., narratives) or spoken discourse. More support for this can be found by examining the scores of different dimensions. In general, the texts in this corpus were information rich (negative scores in Dimension 1), non-narrative (negative scores in Dimension 2), context-independent (positive scores in Dimension 3), without explicit projection of the writer's stance (low scores in Dimension 4), and rich in technical and abstract information (positive scores in Dimension 5). These all show that the texts more closely resembled academic written discourse than everyday spoken discourse. Corroborating

previous studies on the writing skills of EMI and CLIL students (e.g., Whittaker et al., 2011; Whittaker & McCabe, 2020), the EMI students in this study had to some extent mastered disciplinary literacy by, for example, using demonstrating pronouns and conjuncts to enhance text cohesion; using the pronoun "it" and passive voice to make the text less subjective; using nominal groups to make the text less context-dependent. Hence, they were able to demonstrate their content knowledge in the form of expository scientific texts. These findings look more encouraging than those revealed by Coetze-Lachmann (2007), where the German students could not construct their ideas or demonstrate argumentation skills with appropriate subject-specific vocabulary and academic language in English in response to some geography tasks. Such contrasting findings may be due to the fact that the students in our study had studied biology through an L2 (English) for at least 3 years when they sat for the public examination, in addition to at least 12 years of learning English as an additional language. Hence, they have been apprenticed to disciplinary literacy and could make use of appropriate linguistic resources even under high-stakes test conditions. While Whittaker and her colleagues (2011, 2020) have tracked students' development through collecting their writings in several years, this study collated a stratified sample of essays written in a different condition (i.e., the testing condition), thereby contributing more empirical evidence to the disciplinary literacy development of students at the end of their bilingual secondary education.

Second, this study investigated the linguistic features of the essays at different levels of academic performance (i.e., the overall grade from level U to 5**) to reveal any potential relationships between students' academic achievement and their mastery of disciplinary literacy, which would thereby demonstrate the extent to which content and language learning occurred simultaneously. Although the MAT analysis revealed that scientific exposition was the text type closest to this corpus, students who had received different level scores performed differently in different dimensions. Specifically, students with higher levels were more likely to produce informationally dense (Dimension 1) and more abstract texts (Dimension 5), and they tended to use longer words, higher type-token ratios, more attributive adjectives, prepositional phrases, demonstrative pronouns, passives, conjuncts, and other adverbial subordinators. All these linguistic features have been identified as typical features of academic prose, which can thus explain the differences in the dimensional scores that different groups of students produced.

Similar findings have previously been reported (Whittaker et al., 2011), but this study presents a more comprehensive picture of the differences in the registers and lexico-grammatical features written by different groups of students. More importantly, our findings suggest a potential relationship between mastery of disciplinary literacy and content subject achievement, thereby revealing the integral role of academic language in content learning. Although this data cannot show absolute causal relationship between the two variables, i.e., whether mastery of disciplinary literacy contributes to academic performance or the other way round, our findings at least speak to the inextricable relationship between the two. The learner corpus constructed for this study consists of the essays in a biology examination which requires high order thinking skills (e.g., synthesis of information and discussion of various perspectives) and elaborate answers. To succeed in this task, students were expected to possess both content knowledge and linguistic resources to demonstrate their learning outcomes, similar to what Shaw and Imam's (2013) study has shown. As Whittaker and McCabe (2020) argued, students in bilingual education programmes need a coupling of content and language (disciplinary literacy) development, which in turn underscores the importance of assisting students in developing this disciplinary literacy within such programmes.

At the same time, we acknowledge some inconsistent patterns concerning the use of various linguistic features which may have contributed to the differences between texts produced by different level groups. For instance, the sub-corpus of level U students did not follow the normal patterns in Dimensions 2 and 6, and the higher-level students used more public verbs (a positive indicator for narrative texts) than lower-level students. There could be two explanations for these. First, methodologically, the learner corpus analysed in this study was not very large. In keeping with the overall proportion of students who scored at the different biology grade levels, the number of essays written by Level U and 5** students was comparatively lower. Some of the essays written by Level U students were further excluded because they did not reach the minimum word count threshold of 50 for inclusion in this analysis. This may have contributed to some inconsistent patterns. Second, it is also possible that bilingual students in secondary schools are still developing their disciplinary literacy. While some linguistic features such as passives or conjunctions may be easier to use or manifest, others like nominalisations may be more difficult. Hence, some of the linguistic features might not exhibit a regular trajectory from Level U to 5**.

It should also be noted that while the MD analysis depicted the systematic and generalisable features of the student essays, such kind of approaches do not consider the interrelation between context and linguistic choices. Heller and Morek (2015) pointed out that academic language not only serves as a tool for transferring knowledge (communicative function) but also represents a way of thinking about the world (epistemic function) and displays a sense of belonging and identity to the respective community of practice (socio-symbolic function). Therefore, the findings of this study, though focusing on EMI students' developmental trajectory in disciplinary writing, neglect the epistemic and socio-symbolic functions of academic language. More evidence is needed from different perspectives (e.g., textual, contextual) before a conclusion can be made.

8. Conclusion

This corpus-based study with MD analysis sought to investigate the texts produced by EMI students and explore any relationships between content learning and disciplinary literacy in bilingual secondary education programmes. Although students could generally express their content knowledge through scientific exposition, students who achieved higher grade levels demonstrated more lexico-grammatical features closely related to academic prose. Such findings imply that developing L2 disciplinary literacy is not intuitive for some L2 learners and that there is a need for teachers to provide more explicit scaffolding and instruction in this area. The goal of such instruction is to help students express content knowledge through appropriate registers and linguistic features. Some proposals for pedagogical frameworks of integrating content, language, and literacy instruction in bilingual education programmes can be found in the work of Cammarata (2016), who proposed a framework for teachers to set content, language, and literacy objectives, and in authors (XXXX), who illustrated how to plan and deliver lessons based on the teaching/learning cycle of unpacking content and academic language, modelling, text analysis, and joint construction (Rose & Martin, 2012). Authors (XXXX) also recently highlighted the potential of using "semantic waves" to guide classroom interaction between teachers and students to achieve both unpacking and repacking of academic language. It is recognised that universities usually offer English for academic purposes (EAP) and English for specific purposes (ESP) courses to help new arrivals bridge the gap between general English and EAP and ESP (Liu et al., 2011). If such bridging efforts could be offered earlier at secondary school level, it would better facilitate students'

transition into higher education, where they will face even more specialised content and the need for disciplinary literacy.

This study also demonstrates the potential of implementing MD analysis and MAT to analyse texts written by L2 students in bilingual education programmes. While existing corpus-based studies on L2 student writing has differentiated students with different L2 proficiency levels (e.g., Kim & Nam, 2019; Weigle & Friginal, 2015), this study reveals differences in the texts written by students with different academic content subject performance levels. Although the texts written by this group of learners may be short compared with essays written by university students, holistic analyses with MAT are still both possible and fruitful. Future research could potentially investigate the developmental trajectory of this same group of students' written texts over a period of time with MD analysis.

[Word count: 8191]

References

Authors (XXXX)

Asencion-Delaney, Y. (2014). A multi-dimensional analysis of advanced written L2 Spanish. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 214-275). Amsterdam/Philadephia, PA: John Benjamins.

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

Biber, D. (1989). A typology of English texts. *Linguistics, 27*, 3-43.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. (2019). Multi-dimensional analysis: A historical synopsis. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 11-26). London: Bloomsbury Academic.

Brown, C. L. (2005). Equity of literacy-based math performance assessments for English language learners. *Bilingual Research Journal, 29*(2), 337-363.

Cammarata, L. (2016). Foreign Language education and the development of inquiry-driven language programs: Key challenges and curricular planning strategies. In L. Cammarata (ed.), *Content-based foreign language teaching: Curriculum and pedagogy for developing advanced thinking and literacy skills* (pp. 123-143). New York: Routledge/Taylor Francis.

Curriculum Development Council (CDI), & Hong Kong Examinations and Assessment Authority (HKEAA). (2015). *Science Education Key Learning Area: Biology Curriculum and Assessment Guide (Secondary 4-6).* Hong Kong: CDI and HKEAA.

Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics, 35*(3), 243-262.

Coetze-Lachmann, D. (2007). *Assessment of subject-specific task performance of bilingual geography learners: Analysing aspects of subject-specific written discourse*. Unpublished PhD thesis. Osnabrück: Osnabrück University.

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics, 1*(2), 216-253.

de Oliveira, L. C. (2010). Enhancing content instruction for ELLs: Learning about language in science. In D. Sunal, C. Sunal, M. Mantero, & E. Wright (eds.), *Teaching science with Hispanic ELLs in K-16 classrooms* (pp. 135-150). Charlotte, NC: Information Age.

Friginal, E., & Mustafa, S. S. (2017). A comparison of U.S.-based and Iraqi English research article abstracts using corpora. *Journal of English for Academic Purposes, 25*, 45-57.

Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics, 34*(1), 25-52.

Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics, 40*(4), 646-674.

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

Heller, V., & Morek, M. (2015). Academic discourse as situated practice: An introduction. *Linguistics and Education, 31*, 174-186. https://dx.doi.org/10.1016/j.linged.2014.01.008

HKEAA. (2011). *Grading Procedures and Standards-referenced Reporting in the HKDSE Examination*. Hong Kong: HKEAA.

Kim, J., & Nam, H. (2019). How do textual features of L2 argumentative essays differ across proficiency levels? A multidimensional cross-sectional study. *Reading and Writing, 32*, 2251-2279.

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes, 23*, 16-30.

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Westport, CT: Ablex.

Levin, T., & Shohamy, E. (2008). Achievement of immigrant students in mathematics and academic Herbrew in Israeli school: A large-scale evaluation study. *Studies in Educational Evaluation, 34*(1), 1-14.

Liu, J. Y., Chang, Y. J., Yang, F. Y., & Sun, Y. C. (2011). Is what I need what I want? Reconceptualising college students' needs in English courses for general and specific/academic purposes. *Journal of English for Academic Purposes*, *10*(4), 271-280.

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge: Cambridge University Press.

Macaro, E. (2018). *English Medium Instruction*. Oxford: Oxford University Press.

Moje, E. B. (2015). Doing and teaching disciplinary literacy with adolescent learners: A social and cultural enterprise. *Harvard Educational Review, 85*, 254–278.

Nesi, H. (2016). Corpus studies in EAP. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 206-217). Abingdon: Routledge.

Nini, A. (2015). Multidimensional analysis tagger (Version 1.3). Available at https://sites.google.com/site/multidimensionaltagger

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Rose, D., & Martin, J. R. (2012). *Learning to write/reading to learn: Genre, knowledge and pedagogy in the Sydney school*. London: Equinox.

Sardinha, T. B., & Veirano Pinto, M. (2019). *Multi-dimensional analysis: Research methods and current issues*. London: Bloomsbury Academic.

Schleppegrell, M. J. (2009). Language in academic subject areas and classroom instruction: What is academic language and how can we teach it? *Paper presented at workshop on The role of language in school learning sponsored by the National Academy of Sciences, Menlo, CA.* Retrieved from https://www.rcoe.us/educational-services/files/2012/08/What_is_Academic_Language_Schleppegrell.pdf

Shaw, S. (2012). International assessment of geography through the medium of English: Analysing the language skills required. In P. Charzyński, K. Donert & Z. Podgórski (eds.), *Bilingual teaching – Globalization, regional geography and English integration* (pp. 24-44). Toruń: Association of Polish Adult Educators.

Shaw, S., & Imam, H. (2013). Assessment of international students through the medium of English: Ensuring validity and fairness in content-based examinations. *Language Assessment Quarterly, 10*(4), 452-475.

Stoller, F. L. (2008). Content-based instruction. In N. Van Deusen-Scholl & N. H. Hornberger (eds.), *Encyclopedia of Language and Education (2nd ed.) Vol. 4: Second and Foreign Language Education* (pp. 59-70). New York: Springer Science/Business Media.

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.

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343-362.

Whittaker, R., & McCabe, A. (2020). Expressing evaluation across disciplines in primary and secondary CLIL writing: a longitudinal study. *International Journal of Bilingual Education and Bilingualism.* https://doi.org/10.1080/13670050.2020.1798869