ANALYZING DISCOURSE COMPETENCE IN ARGUMENTATIVE ESSAYS

BY

YU-WEN LIEN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Teaching of English as a Second Language
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Master's Committee:

Associate Professor Xun Yan, Chair
Research Assistant Professor Rurik Tywoniw

ABSTRACT

This study investigates discourse competence in second language (L2) argumentative essays. Discourse competence has long been treated holistically in most rating rubrics in large-scale, high-stakes, writing tests. To investigate the detailed sub-constructs of discourse competence in second language (L2) writings, this study analyzed 120 argumentative essays of different proficiency levels on an English-as-a-second-language (ESL) placement test at a Midwest university in the United States with an adapted fine-grained discourse competence rating scale from Wang and Xie's study (2022). The study found that among the nine discourse features, F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development had strong and significant correlations with the EPT placement levels, and the F7 Connective complexity, F8 Connective accuracy, and F9 Complexity of hedges and boosters had very weak and insignificant correlations with the placement levels. The findings of the study presented the nature of discourse competence across proficiency levels and provided practical implications for ESL academic writing curriculum development.

ACKNOWLEDGMENTS

While it is a preliminary research project and could be further improved, I really appreciate having the opportunity to apply what I learned in our program in the two years in this project. I am also highly grateful for all the people who have helped me in the process.

Ping-Lin, Huiying, and Yulin, thank you so much for helping me study Rasch analysis and encouraging me to explore the field of language testing these two years. It was my great honor to meet and learn with you in the learning group. Rosana, thank you so much for your astute suggestions and help in refining my research method and rating scale. It is you who made this project possible. I am also thankful for my friends, classmates, professors, ESL family, and EJP Language Partners. Your help, support, and encouragement gracefully helped me through the hard times. The two years in this program are one of the best memories of my life.

Most importantly, I would like to acknowledge my grandpa, my parents, and my sister for your unconditional love and support. Thank you so much for tolerating my untraditional life plans and being proud of me no matter which path I chose. I also want to thank my college friend, Zong-Ying, for always supporting me and being my shelter in the United States. Without you, I could never have come so far.

Finally, I could never express enough gratitude to my committee, Dr. Xun Yan and Dr. Rurik Tywoniv. Without your insightful feedback, support, and guidance in research methods and language testing, I would never be able to finish this project. Any errors, great or small, that remain are entirely my own.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Discourse competence is essential in writing an effective essay for second language learners, and it contributes to coherence, cohesion, text organization, and the flow of argumentation in the text. However, it is challenging to specify instructions regarding discourse competence in English as a Second Language (ESL) writing courses since discourse competence requires multiple linguistic and semantic management skills. Furthermore, discourse competence has been commonly rated holistically as "organization" in current large-scale high-stakes writing tests, which provided limited feedback for the test takers. Thus, this study aims to analyze the discourse competence features in 120 argumentative essays on an English placement test at a US university with a fine-grained scale adapted from Wang and Xie's study (2022) to provide implications for test development and ESL writing instructions.

CHAPTER 2: LITERATURE REVIEW

2.1 Theoretical constructs of discourse competence

According to Canale and Swain (1980), discourse competence is "the mastery of how to combine and interpret meanings and forms to achieve a unified text in different modes by using (a) cohesion devices to relate forms and (b) coherence rules to organize meanings" (p. 339). The two elements, cohesion and coherence, are theoretically different but conceptually inter-connected. More specifically, coherence is about the structural relations, such as information structure and thematic patterns, that make a text unified, meaningful, and can be understood, while cohesion refers to the non-structural semantic relations, including lexical and grammatical mechanisms that organize relevant information orderly (Witte & Faigley, 1981; Mortensen et al., 2009).

In Halliday and Hassan's Cohesion in English (1976), the five linguistic devices commonly used to build cohesion in written and spoken English are reference, conjunction, lexical cohesion, substitution, and ellipsis. First, reference includes using pronouns, demonstrative phrases (e.g., these, such, neither), or comparative references (e.g., identical, different) to imply whether the information is mentioned in the text before. Second, conjunction refers to the semantic connection between two textual elements. It consists of additive (e.g., and, moreover), causative (e.g., because, therefore), temporal (e.g., next, then, before), and adversative (e.g., but, however) conjunctions. Third, lexical cohesion is achieved by using semantically related words throughout the text, including reiteration/ repetition, synonym (e.g., content – joyful), antonym (e.g., appear – disappear), superordination (e.g., airplane – vehicle), and collocation (e.g., take a risk). Fourth, substitution is using a general term (e.g., one, so, do) to replace previously mentioned information. For instance, he rejected to accept the job offer but

encouraged his friend to do so. Finally, the ellipsis omits repeated words. For example, she intended to propose the idea but [she] didn't [propose it]. In writing, reference, conjunction, and lexical reiteration and collocation are more commonly used. Following previous scholars, Grabe and Kaplan (1996) later suggested that discourse knowledge in writing is knowing how to construct a cohesive text, such as understanding the main topics, organizing schemes, and using transition markers, information structure, and semantic relations. Since most of the earlier studies only focused on the textual perspective, Chalhoub-Deville (2003) proposed to include reader-writer interactions to acknowledge the mutual interaction in the reading and writing process. These concepts build up the preliminary framework of discourse competence.

2.2 Discourse competence in writing assessment research

Recent studies attempted different approaches to analyze components of discourse competence in writing. Among the components, cohesion, and coherence were the most frequently discussed discourse features in writing assessment research.

For instance, Crossley and McNamara (2010) researched the roles of cohesion and coherence in essay evaluations. Expert raters used a holistic scale of scores 1-6 to rate 184 essays from the corpus at Mississippi State University. They found that expert raters evaluated coherence mostly relying on the absence of connectives instead of their presence, and coherence is an important feature attributing to the overall essay quality. Crossley and McNamara (2016) later validated their findings in 2010 by investigating the relationships between essay quality, text elaboration, and text cohesion. In the study, 35 students wrote and elaborated two essays in fifteen minutes. Then, a discourse comprehension expert revised the four essays, original and elaborated ones, to improve the cohesion. According to the experts' ratings, a combination of

elaboration and improved cohesion contributed to higher scores in general quality and coherence. Furthermore, Crossley et al. (2016) examined the relationships between writing quality and local, global, and text cohesive devices and their effects on human judgment. The researchers analyzed fifty-seven L2 university students' essays in a semester-long upper-level academic English course, yet no clear cohesion pattern was found between longitudinal analysis and human judgments of essay quality. Two global cohesive devices, the adjacent overlap of function words and nouns between paragraphs, and one local cohesive device, the adjacent overlap of function words between sentences, were found to have the strongest predictive power to the human rating of overall essay quality. Raters' judgment of text organization was also found to be influenced by the function words and connectives, explaining 36% of the variance of the human rating. While the focus of the three studies was slightly different, their results suggested that global coherence and certain local cohesive devices were important indicators of the essay's overall quality and rating scores. However, it should be noted that some mixed results were found across L1 and L2 studies. For example, Crossley and McNamara (2011) found that global coherence, not local cohesion, could better explain the human ratings of coherence in L1 speakers' essays.

Since argumentative essays have been regarded as the most difficult model in both L1 and L2 compared with narrative, descriptive, and exposition essays (Gleason, 1999), some previous studies analyzed the cohesion and coherence in argumentative essays specifically. For instance, Connor (1984) examined the cohesion and coherence in argumentative essays on six prompts and compared the features in ESL learners' essays and native speakers' essays. They found that ESL writers used fewer varieties of cohesive devices than native speakers. Basturkmen and von Randow (2014) also compared the native speakers' and L2 learners' use of cohesive devices. They analyzed 20 postgraduate students' argumentative essays at a university

in New Zealand; 15 of them were ESL speakers. The categories of textual meta-discourse used to analyze the coherence included announcements, illocutionary markers, code glosses, logical markers, reminders, sequencers, and topicalizers. The result showed that essays with higher scores demonstrated well in guiding the readers to recreate coherence by using illocutionary markers, code glosses, and logical markers. Different from the previous two studies comparing ESL speakers' and native speakers' argumentative essays, Yang and Sun (2012) focused on analyzing undergraduate Chinese EFL learners' use of cohesive devices in argumentative essays. The result indicated that the correct use of cohesive devices and the amount of lexical cohesion had a significant positive correlation with the writing quality and L2 learners' proficiency. When L2 learners' proficiency levels increased, they were more capable of using various lexical choices to maintain cohesion in their writing, and the number of cohesion errors dropped significantly between intermediate and advanced L2 learners.

Automated measures have also been used to analyze the cohesive features in essays recently. For example, Tabari et al. (2023) used the Tool for Automatic Analysis of Cohesion (Crossley et al., 2018) to examine the effects of task complexity and task repetition on L2 writers' use of cohesive devices. They found that task complexity had a more significant effect on L2 writers' use of textual and local cohesive devices. Another study by Tian et al. (2024) used natural language processing (NLP) to predict L2 writing fluency with three cohesive devices (reference, conjunction, and lexical cohesion) in argumentative essays. They found that higher text productivity was associated with more unattended demonstratives (e.g., this without associated nominal), but more pronouns and more verbs and adjective overlap between sentences were related to lower productivity. They also noticed that the more conjunctions or unattended demonstratives were used, the more revisions the writers tended to have, and the chance

decreased if the writers used more attended demonstratives and negative connectives. Since attended demonstratives and negative connectives were usually related to complex argumentation, they required more processing effort during production. In a timed writing task, writers might not have sufficient cognitive resources to employ more sophisticated cohesive devices. On the contrary, using unattended demonstratives, basic connectives, and high-frequency words, was generally easier, requiring less processing effort during L2 text production (Hu & Li, 2015). Therefore, writers tended to use less complex cohesion features in production and revision with diminished pause behaviors.

Although several studies investigated coherence or cohesion in writing, most of them only focused on some of the cohesive devices or used exploratory methods; few studies have used a fine-grained rating scale to comprehensively analyze discourse competence in essays. The most recent studies that utilized a comprehensive rating scale for discourse competence are from Xie and Lei (2021) and Wang and Xie (2022).

First, Xie and Lei (2021) analyzed organization, coherence, and cohesion in 339 first-year undergraduates' English academic term papers at a Hong Kong university. They used a diagnostic checklist to assess the organization, coherence, and cohesion with descriptors derived from the interviews with raters, such as the overall organization of a text, cross-paragraph transition, sequence of paragraphs, and within-paragraph unity. While the descriptors reflected what raters evaluated when assessing the textual features, the descriptors lacked the theoretical foundations.

Thus, to provide a comprehensive specification of discourse competence, Wang and Xie (2022) developed a fine-grained discourse competence scale based on theories and literature, identifying five components of discourse competence in academic writing: (1) topic building, (2)

global coherence, (3) local coherence, (4) logical connectives, and (5) reader-writer interactions. First, topic building refers to the appropriate focus on the topic from the writing prompt. Second, global coherence concerns the paragraph development in the introduction, body paragraphs, and conclusion. The introduction should provide sufficient background and context of the topic; the body paragraphs should effectively and logically advance the main argument; the conclusion should briefly summarize the main points. Third, local coherence refers to the theme-rheme (given-new) information structure, which interconnects the supporting details. Fourth, local connectives include the complexity and accuracy of connectives, which are the linking words with logical or semantic connections. Finally, reader-writer interactions are analyzed with the complexity of hedges and boosters. In the rating rubrics, the five components corresponded to ten discourse features, and the authors used the rubrics to assess the strengths and weaknesses of 108 essays written by Chinese EFL undergraduate students. They found that the students generally performed well in utilizing logical connectives but needed improvements and further practice in topic building, global coherence, and reader-writer interaction.

2.3 Research question

Since the analysis of discourse competence in Wang and Xie's study (2022) focused on problem-solving essays in a classroom setting, this study is to replicate their study design, validate their fine-grained rating scale with detailed sub-constructs of discourse competence, and examine the discourse competence in argumentative essays in a large-scale, high-stakes placement test setting where discourse competence was rated holistically. The present study adapted the discourse competence rating scale from Wang and Xie (2022) with five components and nine features, and the research questions are:

1. How does the individual discourse competence feature differ across proficiency levels and profiles in argumentative essays?
2. Can discourse competence features predict an argumentative essay's proficiency level?

CHAPTER 3: METHODOLOGY

3.1 Context of the study

The argumentative essays analyzed in the study are from the English Placement Test (EPT) at a Midwest university in the United States. Newly admitted international students are required to take the test if their pre-arrival English proficiency test scores are below the cutoff scores, such as a total of 103 in the TOEFL iBT with 27 writing and 23 speaking sub-scores, 7.5 in the IELTS with 7 in writing and 6.5 in speaking, and 135 in the Duolingo English Test. The test contains a writing test and a speaking test. In the writing test, test takers first watch a mini-lecture on a given topic in ten minutes. Then, they read six short articles about the given topic, three for pros and three for cons, and a series of guiding questions to help them structure their essay. Test takers have 90 minutes to read the materials, take notes, and compose an argumentative essay following academic writing conventions.

The argumentative essays were rated by certified raters with a holistic rating scale. The rating scale consists of two scores, argument development and lexico-grammatical features. Each score ranges from 1 to 3. Based on the scores in argument development and lexico-grammatical features, essays were then categorized into six profiles: A, B1, B2, C1, C2, and D. A-level essays are strong in both argument development and lexico-grammatical features. They have a well-developed argument, well-structured body paragraphs with sufficient, interconnected supporting details, and precise and sophisticated lexico-grammar. B-level essays are generally strong but noticeably weak in either lexical-grammatical ability (B2) or argumentation (B1). C-level essays are mediocre and have issues in both argumentation and lexico-grammar. C1 essays have more severe issues in lexico-grammar, and C2 essays are weaker in argumentation. D-level essays

have noticeable critical issues in both argument development and lexico-grammar, often impeding reader comprehensibility.

The six profiles correspond to four placement levels: 4 (A), 3 (B1/B2), 2 (C1/C2), and 1 (D). Levels 1 and 2 indicate the test takers need two-semester ESL writing courses. Test takers with placement Level 3 have to take a 1-semester ESL writing course. Since Level 4 shows that test takers demonstrate a certain degree of excellence in argumentation and lexico-grammar, they are exempt from ESL writing courses.

Table 1. Description of the written scores on the test website.

Level	Description
4	Essay provides a developed argument based on a controlling idea, with well-connected ideas that advance the central claim using relevant evidence from the sources. Generous readers will have little issue following the organization of the essay. Essay consistently displays academic lexico-grammar with few noticeable errors.
3	Essay provides a sufficiently-developed argument composed of (often) interconnected, sufficiently-explained evidences supporting a single central claim. Essay displays complex lexico-grammar with academic discourse but sometimes with noticeable errors.
2	Essay provides an argument that supports a single central claim but it sometimes relies on list-like organization of evidences and/or evidences are not clearly connected. Essay may attempt complex lexico-grammar but sometimes with noticeable errors causing comprehension difficulty for the reader.
1	Essay may not provide an argument that supports a single central claim. Evidence poorly support the argument and ideas are not interconnected and/or underdeveloped. Essay may attempt complex lexico-grammar but with frequent, noticeable errors that are often inaccurate and cause comprehension difficulty.

3.2 Data collection

The collected 120 argumentative essays were written by undergraduate or graduate EPT test takers from 2021 to 2022. Two prompts were given for the dataset: green roofs and remote learning. Due to insufficient essays for placement 1 (profile D) in the database, only essays for placements 2, 3, and 4 were collected; 40 essays per placement and 60 essays per prompt. The

average word count of the 120 essays is 551 words; it is close to the length of the essays collected in Wang and Xie's study (2022), which were around 500 words.

Table 2. Level and profile distributions of the dataset.

Placement Level	**4**	**3**	**3**	**2**	**2**	**Number**
Profile	**A**	**B1**	**B2**	**C1**	**C2**	**of essays**
Green roofs	20	11	9	11	9	60
Remote learning	20	12	8	6	14	60

Since disagreements between the scores of argumentation (ARG) and lexico-grammar (LEX) frequently existed in the original ratings, no final profile was given to the essays in the original EPT database. To analyze essays that are more representative of the corresponding levels and profiles, I mostly selected the benchmark essays with two identical ratings in the database. Some essays in Level 4 received a score of 2 in either argument or lexical grammar, which made the average score 2.5. Yet their overall score was still higher than the essays in other levels, so they are marked as Profile A in this study. Only eight essays in Level 2 had different ratings, such as 2-1 and 1-2 in argument and lexical grammar, which made the average score 1.5 in two categories. In these cases, I used the Many-facet Rasch measurement to convert the initial ratings considering the rater effects. According to the converted scores in the output of Rasch, I labeled the eight essays with C1 or C2 as their profiles.

3.3 Data analysis

3.3.1 Rating for discourse competence

While the 10 feature discourse rubrics from Wang and Xie's study (2022) were theory-driven, I adapted the rating rubrics in this study due to three reasons. First, the discourse rubric was used for diagnostic purposes on problem-solving essays in classroom contexts, which is

different from the context of the EPT essays in this study. Second, the first three features, F1 Topic/ Focus, F2 Thesis statement, and F3 Controlling idea, shared overlapping discourse features, and the descriptors looked similar. To better fit the testing context, EPT rating scale, and argumentative essay genre, and to reduce possible redundancy of ratings for discourse competence, I combined the F2 Thesis statement and the F3 Controlling idea into one rating question. I also rewrote the descriptors for the F1 Topic/ Focus, focusing on the unity of the stance and the consistency of the development of argumentation. Third, the original rubrics in Wang and Xie's study (2022) assessed the ten features on a 5-point scale. Since some descriptors were very similar on the 5-point scale, the rating rubrics were revised to a 3-point scale to ensure more distinctive differences between the three scales and to make the rating process easier for the raters. Therefore, the rating rubrics for discourse competence in the current study (see Table 3 and Appendix 1) include nine features with an ordinal scale, ranging from 1 (poor), 2 (fair), and 3 (excellent).

Table 3. Features of the rating scale and rating phases in the current study

Components of discourse competence	**Features**	**Rating phases**
Topic building	F1 Topic/ Focus	1
	F2 Thesis statement	
Global coherence	F3 Introduction	
	F4 Body paragraphs	2
	F5 Conclusion	
Local coherence	F6 Theme-rheme development	
Logical connectives	F7 Connective complexity	3
	F8 Connective accuracy	
Reader-writer interaction	F9 Complexity of hedges and boosters	4

To ensure the consistency of the rating and reduce rater fatigue, the rating process consists of four phases. First, the raters read the introduction paragraphs and rate F1 Topic/Focus, F2 Thesis statement, and F3 Introduction. Second, the raters read the body paragraphs and the conclusion and rate F4 Body parts, F5 Conclusion, and F6 Theme-rheme development. Third, the raters identify the connectives used in the essays and evaluate and rate F7 Connective complexity and F8 Connective accuracy. Finally, the raters focus on the use of hedges and boosters in the argumentative essay and rate F9 Complexity of hedges and boosters. To make the last rating phase smoother, the connectives, hedges, and boosters in the essays were highlighted in different colors (connectives – red, hedges – blue, boosters – green) based on Appendix 4 in Wang and Xie's study (2022), and the types and ratios (per word) of the three linguistic features were calculated before the rating.

One content expert with experience as an instructor and course leader of the ESL academic writing courses who was also a certified EPT essay rater assisted in revising the rating rubrics. After I adapted Wang and Xie's discourse competence rubrics to 9 feature rubrics on a 3-point scale, I explained the rubrics and the purpose of the study to the content expert, and then we rated 12 essays (10% of the data) with the new rubrics individually. The selected 12 essays were written on the same prompt and consisted of four essays from each placement level. In total, we had three face-to-face discussions, including one pre-rating meeting and two post-rating meetings, and each discussion lasted for about one hour. In the post-rating meetings, we compared our ratings and revised the rubrics to better reflect the discourse competence features in the argumentative essays. When there was any disagreement about the interpretation of the rubrics and ratings, we exchanged our thoughts and observations until we reached a consensus. The final inter-rater reliability across the categories measured with Cohen's Kappa was 0.83,

which was almost perfect (Landis & Koch, 1977), and the agreement between the two raters was 0.89, exceeding 0.8. After the rating scale was refined, I rated the rest of the 108 essays.

Several descriptors of the rating scale were modified in the meetings with the content expert. F1 Topic/Focus evaluates not only the unity of the stance but also the extent to which the body paragraphs support the controlling idea. F3 Introduction included descriptors in 2 (Fair) about the degree of irrelevant or unnecessary details to make the three-point descriptors more consistent. F4 Body paragraphs' descriptors had more explicit descriptions about the amount and the effectiveness of the supporting details. In the original F6 Theme-rheme development descriptors, if the essay has two to three coherence breaks, it is considered 2 (Fair). Since the descriptor of 3 (Excellent) stressed no coherence break detected, the descriptor of 2 (Fair) was revised to include one to three coherence breaks. F7 Connective complexity specified the number of types of connectives, from one to three, in the original scale. However, after calculating the counts and types of connectives in the essays, the average type of connectives used was 14, and the average count of connectives was 19, which did not align with the numbers in the original scale. Thus, the revised descriptors of 2 (Fair) in F7 Connective complexity focused on whether some simple connectives are overused. If an essay had fewer than five types of connectives, it would be considered 1 (Poor). Finally, the original descriptors of 2 (Fair) in F8 Connective accuracy mentioned certain connectives are misused, overused, or underused. Since the overuse or underuse of connectives is more about the complexity of connectives, the F8 Connective accuracy descriptors were revised to focus only on the misuses. If there are one or two cases of misuse of connectives, the essay would be considered 2 (Fair). If there are more than three misuse cases, the essay would be considered 1 (Poor). It should be noted that the misuses only

include the connectives that were not used precisely or did not logically fit in the context. The spelling mistakes were not counted as misuses of connectives.

3.3.2 Statistical analysis

The discourse competence ratings are treated as independent variables. The ratings of the nine features are ordinal. For F7 Connective complexity, the type-token ratio of the connectives used in the essays was calculated by the total number of connectives divided by the total number of words (tokens) in the essay. For F8 Connective accuracy, the accuracy rate was measured by the percentage of correct usage of connectives (the number of correct usage of connectives divided by the total number of connectives). For F9 Complexity of hedges and boosters, the numbers and types of hedges and boosters, as well as their type-token ratio were calculated respectively. The ratios of the usages of the connectives, hedges, and boosters will be normalized into frequencies per 1000 words, which are interval data. Since there was no misuse of hedges or boosters in the essays, the accuracy of hedges and boosters was not calculated.

Table 4. Formulas for normalized frequencies of connectives, hedges, and boosters.

	Formulas for frequencies per 1000 words
Types of connectives	(Total types of the connectives in the essay ÷ The essay's word count) ×1000
Counts of correct connectives	(Counts of the total connectives – Counts of the misused connectives) ÷ The essay's word count ×1000
Counts of hedges	(Counts of the hedges in the essay ÷ The essay's word count) ×1000
Counts of boosters	(Counts of the boosters in the essay ÷ The essay's word count) ×1000

The dependent variables are the original EPT placement and profile types. Placement levels 4, 3, and 2 are ordinal data, indicating the test takers' general writing proficiency levels.

Profile types A, B1, B2, C1, and C2 are nominal data, and the scores of argument development and lexico-grammar in the original EPT ratings are ordinal data.

I used R via RStudio to conduct the statistical analysis. First, descriptive statistics was performed on the discourse competence ratings, and the numbers and types of connectives, hedges, and boosters. Second, power analyses were performed to estimate the power of the sample size. The priori power analysis showed that the sample size for correlation should be at least 57 essays ($\alpha = 0.05$; power = 0.8). The obtained sample size of this study is 120 essays, exceeding the requirements ($\alpha = 0.05$; power = 0.986). Third, to answer the research question, I examined the assumptions of normality, linearity, and multicollinearity, and then used Spearman's rank correlation coefficient to check the correlations between the dependent variables (placement and profile) and the independent variables (discourse competence ratings and the normalized ratios of connectives complexity, accuracy, hedges, and boosters). Finally, since some discourse competence features have strong correlations with the EPT placements, ordinal logistic regression ($u=3$, $\alpha = 0.001$; power = 0.99) was performed to further investigate the possibility of using discourse competence features to predict an essay's EPT placements.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Descriptive statistics

Descriptive statistics (see Table 5 and Table 6) show that most essays demonstrated excellent performances in F7 Connective complexity, F8 Connective accuracy, and F9 Complexity of hedges and boosters. The essays also generally performed well in F1 Topic/Focus and F6 Theme-rheme development, but the standard deviations show slightly higher individual variabilities. F3 Introduction and F5 Conclusion have the lowest means, indicating that several essays did not have well-developed introduction or conclusion paragraphs. Since the EPT essays were timed, some test takers might not be able to finish the essays or were only able to write a brief introduction or conclusion and focus on developing the body paragraphs. Some test takers also used irrelevant or unnecessary information in the introduction or conclusion, which lowered the average scores in the two features.

Based on the descriptive statistics, most essays accurately used multiple connectives, hedges, and boosters. On average, the connectives are used more than the hedges and boosters, and the hedges are used more than the boosters, indicating that most test takers utilized effective discourse strategies to compose coherent and formal essays. However, the standard deviations of the four items are noticeably high, suggesting significant individual variabilities.

Table 5. Descriptive statistics for discourse competence ratings, counts, and types of connectives, hedges, and boosters.

	n	Min	Max	Median	Mean	SD	Skewness	Kurtosis
F1 Topic/Focus	120	1	3	3	2.38	0.72	-0.71	-0.81
F2 Thesis statement	120	1	3	2	2.17	0.48	0.46	0.42
F3 Introduction	120	1	3	2	1.99	0.65	0.01	-0.69
F4 Body paragraph	120	1	3	2	2.19	0.65	-0.21	-0.74
F5 Conclusion	120	1	3	1	1.56	0.68	0.81	-0.55
F6 Theme-rheme development	120	1	3	3	2.33	0.77	-0.62	-1.07
F7 Connective complexity	120	1	3	3	2.67	0.54	-1.37	0.89
F8 Connective accuracy	120	1	3	3	2.69	0.55	-1.56	1.46
F9 Complexity of hedges and boosters	120	1	3	3	2.52	0.59	-0.78	-0.4
F1-F9 Sum	120	13	26	21	20.51	2.96	-0.25	-0.52
Word count	120	228	1258	532.5	551.83	148.99	1.09	3.22
Types of connectives (per 1000 words)	120	6.98	56.87	27.47	27.37	8.86	0.29	0.17
Correct connectives (per 1000 words)	120	13.96	58.93	33.76	35.3	9.94	0.18	-0.6
Counts of hedges (per 1000 words)	120	0	42.3	13.96	16.73	9.78	0.64	-0.35
Counts of boosters (per 1000 words)	120	0	57.09	12.65	14.23	8.6	1.59	4.6

Table 6. Descriptive statistics for essays' EPT placement levels and scores for argument (ARG) and lexico-grammar (LEX).

	n	Min	Max	Median	Mean	SD	Skewness	Kurtosis
EPT Placement Level	120	2	4	3	3	0.82	0	-1.52
Average ARG	120	1	3	2.5	2.32	0.73	-0.56	-1.07
Average LEX	120	1	3	2	2.26	0.66	-0.44	-0.85

4.2 Spearman's rank correlation coefficients

The relationship between the discourse competence features and the essay's EPT placements was investigated by Spearman's rank correlation coefficients. Table 7 presents the results of the analysis. According to Dancey and Reidy (2007), correlation coefficients ranging from 0.7 to 0.9 indicate strong correlations. Moderate correlation exists if the coefficients are 0.4 to 0.6, and if the coefficients range from 0.1 to 0.3, the correlation is regarded as weak. The following interpretation of Spearman's correlation coefficients is based on the suggested guidelines.

The EPT placements have strong positive relationships with the average EPT argumentation ($r = 0.79$, $p < .01$), lexico-grammar scores ($r = 0.76$, $p < .01$), F1 Topic ($r = 0.71$, $p < .01$), and the sum of nine discourse competence feature ratings (F1-F9 Sum) ($r = 0.74$, $p < .01$). F4 Body paragraphs ($r = 0.65$, $p < .01$), F5 Conclusion ($r = 0.43$, $p < .01$), and F6 Theme-rheme development ($r = 0.6$, $p < .01$) have moderate correlations with the EPT placements, and the EPT placements have weak correlations with F2 Thesis statement ($r = 0.37$, $p < .01$) and F3 Introduction ($r = 0.25$, $p < .01$).

The EPT argumentation scores have a strong positive correlation with F1 Topic/Focus (r = 0.76, p <.01) and a moderate correlation with F4 Body paragraphs (r = 0.59, p <.01), F6 Theme-rheme development (r = 0.58, p <.01), and F1-F9 Sum (r = 0.66, p <.01), but have weak relationships with F2 Thesis statement (r = 0.27, p <.01), F3 Introduction (r = 0.20, p <.01), F5 Conclusion (r = 0.37, p <.01), and F7 Connective complexity (r = 0.21, p <.01). The rest of the features, F8 Complexity accuracy (r = -0.09, p <.01) and F9 Complexity of hedges and boosters (r = -0.04, p <.01), demonstrate very weak negative correlations with the EPT argumentation scores.

The EPT lexico-grammar scores have moderate correlations with F1 Topic/Focus (r = 0.4, p <.01), F4 Body paragraphs (r = 0.47, p <.01), F6 Theme-rheme development (r = 0.4, p <.01), and F1-F9 Sum (r = 0.53, p <.01). However, the other features only have weak or very weak relationships with the EPT lexico-grammar scores.

Among the nine discourse features, one strong correlation coefficient and two moderate coefficients are shown in the matrix, including F1 Topic/Focus and F4 Body paragraphs (r = 0.71, p <.01), F1 Topic/Focus and F6 Theme-rheme development (r = 0.66, p <.01), and F4 Body paragraphs and F6 Theme-rheme development (r = 0.63, p <.01). Yet the rest of the features only have weak or very weak relationships with one another.

Table 7. Correlation Matrix.

	EPT Placement Level	Average ARG	Average LEX	F1 Topic/Focus	F2 Thesis statement	F3 Introduction	F4 Body paragraph	F5 Conclusion	F6 Theme-rheme development	F7 Connective complexity	F8 Connective accuracy	F9 Complexity of hedges and boosters	F1-F9 Sum	Ratio of types of connectives (per 1000 words)	Ratio of correct connectives (per 1000 words)	Ratio of hedges (per 1000 words)	Ratio of boosters (per 1000 words)
EPT Placement Level	1																
Average ARG	0.79	1															
Average LEX	0.76	0.32	1														
F1 Topic/Focus	0.71	0.76	0.4	1													
F2 Thesis statement	0.37	0.27	0.32	0.22	1												
F3 Introduction	0.25	0.2	0.22	0.23	0.29	1											
F4 Body paragraph	0.65	0.59	0.47	0.71	0.13	0.19	1										
F5 Conclusion	0.43	0.37	0.33	0.34	0.23	0.19	0.28	1									
F6 Theme-rheme development	0.6	0.58	0.4	0.66	0.21	0.21	0.63	0.26	1								
F7 Connective complexity	0.18	0.21	0.07	0.21	0.02	0.04	0.34	0.04	0.32	1							
F8 Connective accuracy	0.07	-0.09	0.17	-0.02	0.02	0.05	0.07	-0.1	0.08	0.05	1						
F9 Complexity of hedges and boosters	0.01	-0.04	0.06	-0.06	-0.09	-0.04	0.02	0.04	-0.06	0.02	0.11	1					
F1-F9 Sum	0.74	0.66	0.53	0.75	0.39	0.46	0.75	0.53	0.75	0.4	0.22	0.2	1				
Ratio of types of connectives (per 1000 words)	-0.02	0.01	-0.02	0.01	0.13	-0.05	0.02	-0.04	0.22	0.47	-0.06	0.02	0.14	1			
Ratio of correct connectives (per 1000 words)	0.01	0.03	-0.02	0.09	0.11	-0.06	0.13	-0.04	0.23	0.36	0.09	0.05	0.21	0.81	1		
Ratio of hedges (per 1000 words)	-0.01	0	0	-0.01	-0.1	-0.08	0.01	0.01	0.03	0	0.07	0.78	0.17	0.01	0.09	1	
Ratio of boosters (per 1000 words)	-0.06	0.05	-0.19	0.01	-0.14	-0.05	-0.04	0.01	0.08	-0.09	-0.03	-0.18	-0.02	0.07	0.18	0.13	1

Note:
The underlined coefficients are insignificant ($p > .05$).
**The rest of the correlation coefficients are significant at $p < .01$.

4.3 Discussion

Research Question 1: How does the individual discourse competence feature differ across proficiency levels and profiles in argumentative essays?

4.3.1 Discourse competence features differ across EPT proficiency levels

As Table 8 and Chart 1 show, essays with higher EPT proficiency levels tended to demonstrate better discourse competence and scored higher in the discourse competence feature ratings. The nine features, except F8 Complexity accuracy, and the F1-F9 Sum, shared this tendency.

Since most essays had various accurately used connectives, hedges, and boosters, the mean scores of F7 Connective complexity, F8 Connective accuracy, and F9 Complexity of hedges and boosters were close to 3 (Excellent) with little difference. In F8 Connective accuracy, Level 3 essays had a lower mean score than Level 2 essays, and the average count of connective mistakes in Level 3 essays is 0.53, higher than the Level 2 essay's 0.38 (Table 9). One possible explanation is that Level 2 essays were generally shorter and used fewer connectives than Level 3 essays. Level 2 essays also relied more on simple connectives than Level 3 essays, which might lower the chance of misusing connectives.

Level 4 and Level 3 essays consistently performed well in F1, F2, F4, and F6, with mean scores higher than 2, while Level 2 essays' average scores were below 2 in the first six features. F1 Topic/Focus, F2 Thesis statement, F4 Body paragraphs, and F6 Theme-rheme development seem to be effective features to distinguish the qualities of the essays across the three placement levels with clear score differences.

Compared with the other features, essays scored lower in F3 Introduction and F5 Conclusion. Specifically, all the essays performed especially less ideally in F5 Conclusion, in

which the mean scores were lower than 2 (Fair). While the majority of the test takers could provide some background information in the introduction, they seemed to struggle with developing complete introduction and conclusion paragraphs in a timed test setting. In the dataset, it is frequent to see some essays only have about one to two sentences in the introduction but only have one sentence in the conclusion. Some could not even finish the last body paragraph, and their conclusion paragraph was missing.

Table 8. Mean scores of discourse competence feature ratings across EPT proficiency levels.

Level	F1	F2	F3	F4	F5	F6	F7	F8	F9	F1-F9 SUM
4	2.90	2.40	2.25	2.65	1.88	2.80	2.80	2.78	2.53	22.98
3	2.58	2.15	1.88	2.30	1.63	2.50	2.65	2.60	2.53	20.80
2	1.68	1.98	1.85	1.63	1.18	1.68	2.58	2.70	2.50	17.75

Chart 1. Line chart of the mean scores of discourse competence feature ratings across EPT proficiency levels.

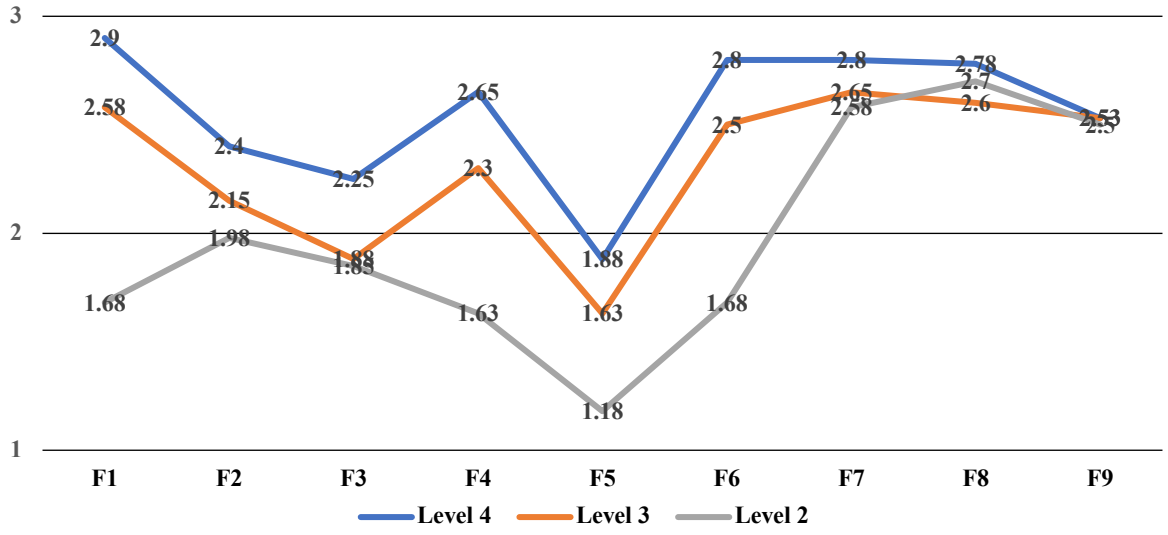


Table 9 presents a similar trend to Table 8, in which essays with higher EPT proficiency levels tended to be longer essays, use more connectives, and utilize more types of connectives. Level 4 essays also had more hedges and boosters than Level 3 and Level 2 essays, and the hedges were used more than boosters. However, some items have slightly different trends than others. For instance, Level 3 essays have the most connective mistakes, the most types of hedges, and the most types of hedges and boosters on average. Level 2 essays also used more types of boosters than Level 4 and Level 3 essays. Yet these slight differences look more like individual variances, not strong indicators of the essay's discourse competence and essay qualities.

Furthermore, contrary to Table 9, Level 3 essays had the most normalized counts of connectives, hedges, and boosters in Table 10. Since Level 4 essays also used grammatical structures, such as participle clauses, and pronouns, such as which, to ensure the coherence of the essays, the differences in the normalized counts seem to be reasonable.

Table 9. Average counts of words, connectives, hedges, and boosters across EPT proficiency levels.

Level	Word Count	Connective Count	Connective Types	Mistakes of Connectives	Hedges Count	Hedges Types	Boosters Count	Boosters Types	Types of hedges + boosters	Accuracy of connectives
4	587.55	21.15	15.43	0.35	9.83	5.13	8.53	4.20	9.33	0.98
3	554.13	19.63	14.95	0.53	9.45	5.28	7.93	4.23	9.50	0.97
2	513.83	18.08	13.30	0.38	8.30	4.38	7.30	4.43	8.80	0.98

Table 10. Average normalized counts of connectives, hedges, and boosters across EPT proficiency levels.

Level	Types of connectives (per 1000 words)	Correct Connectives (per 1000 words)	Count of Hedges (per 1000 words)	Count of Boosters (per 1000 words)
4	26.82	35.30	16.69	14.14
3	28.22	35.51	17.22	14.30
2	27.08	35.08	16.28	14.27

Chart 2. Line chart of normalized counts of connectives, hedges, and boosters across EPT proficiency levels.

4.3.2 Discourse competence features differ across EPT profiles

In Table 11 and Chart 3, discourse competence features do not have specific patterns across the EPT profiles. The only three discourse competence features that aligned with the overall proficiency represented by the EPT profiles are F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development.

In the F2 Thesis statement, B2 essays had higher mean scores than B1 essays, indicating that the more B2 essays in the dataset had a complete thesis statement, consisting of a clear stance and a preview of major points, so more B2 essays received 3 (Excellent) in the discourse competence ratings. The average scores of other B1, C1, and C2 essays in F2 were all close to 2 (Fair), showing that they had a clear stance in the thesis statement, but it could be clearer if they included an effective preview of the major points.

Compared with other discourse features, the essays had weaker performances in F3 Introduction and F5 Conclusion. Their performances in the introduction were slightly better than

those in the conclusion, which might be a result of the timed test setting. Test takers might be able to write a simple introduction paragraph at the beginning of the test, but did not have enough time to compose a complete conclusion at the end. Thus, those incomplete conclusions tended to receive 2 or 1, having only a limited summary of the main points, or failing to restate the main idea with relevant information.

All the essays had comparably better performances in F7 Connective complexity, F8 Connective accuracy, and F9 Complexity of hedges and boosters. The average scores of the three features were all above 2, but no clear and consistent pattern related to the EPT profiles was found in the scores. Since most of the essays successfully used multiple connectives, hedges, and boosters, without many mistakes, the mean scores of the three features were not strong indicators of essays' discourse competence proficiency across the EPT profiles. It should be noted that several B1 essays misused the connectives, on the one hand, and on the other hand, which led to a lower average score in F8 Connective accuracy.

Table 11. Mean scores of discourse competence feature ratings across EPT profiles.

Profile	F1	F2	F3	F4	F5	F6	F7	F8	F9	F1-F9 SUM
A	2.90	2.40	2.25	2.65	1.88	2.80	2.80	2.78	2.53	22.98
B1	2.83	2.09	1.78	2.35	1.61	2.61	2.74	2.39	2.48	20.87
B2	2.24	2.24	2.00	2.24	1.65	2.35	2.53	2.88	2.59	20.71
C1	2.06	2.06	2.12	1.88	1.24	1.82	2.71	2.76	2.47	19.12
C2	1.39	1.91	1.65	1.43	1.13	1.57	2.48	2.65	2.52	16.74

Chart 3. Line chart of the mean scores of discourse competence feature ratings across EPT profiles.

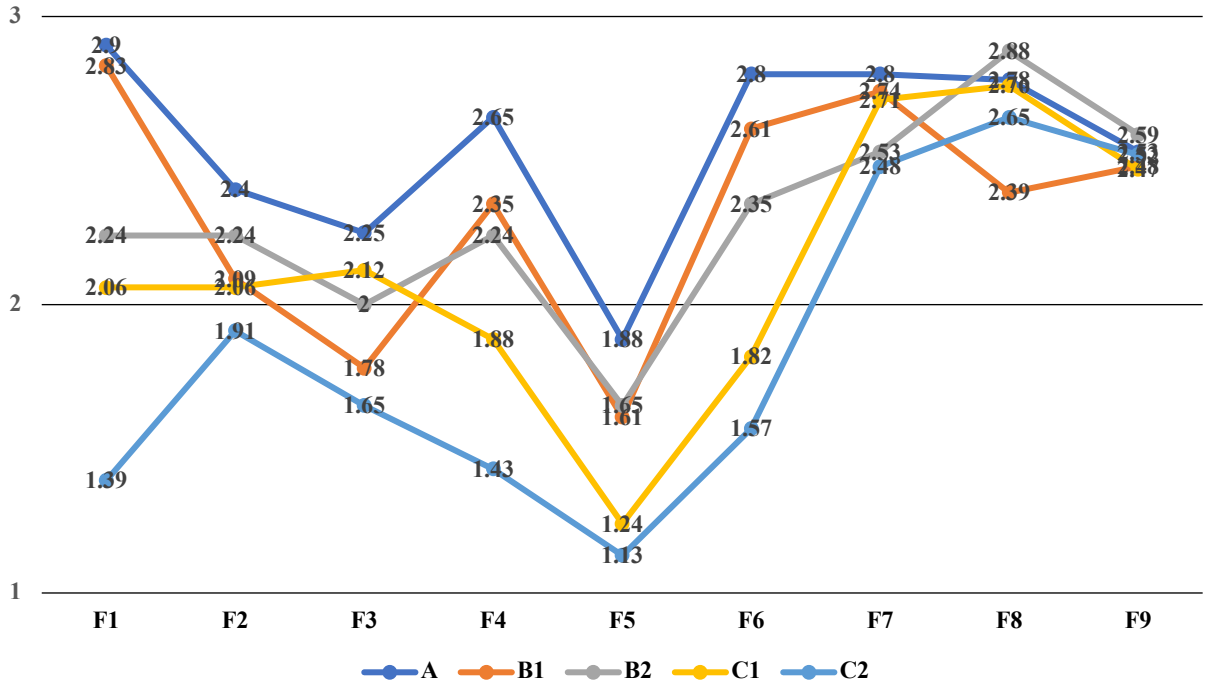


Table 12 presents the average counts of words, connectives, hedges, and boosters across the EPT profiles, and Table 13 shows the normalized counts. Unlike Table 9, where the average counts of words, connectives, hedges, and boosters partially aligned with the EPT placements, no clear pattern between the average counts and the EPT profiles was found in both Tables 12 and 13. In general, all the essays used a variety of connectives, hedges, and boosters, and the hedges were used more than the boosters. The C1 essays had the highest average word count, indicating that the length of the essay was not necessarily an indicator of the essay's quality.

In Table 13 and Chart 4, B1 essays had the highest normalized types of connectives and counts of connectives, hedges, and boosters, yet with little difference from the scores of other profiles.

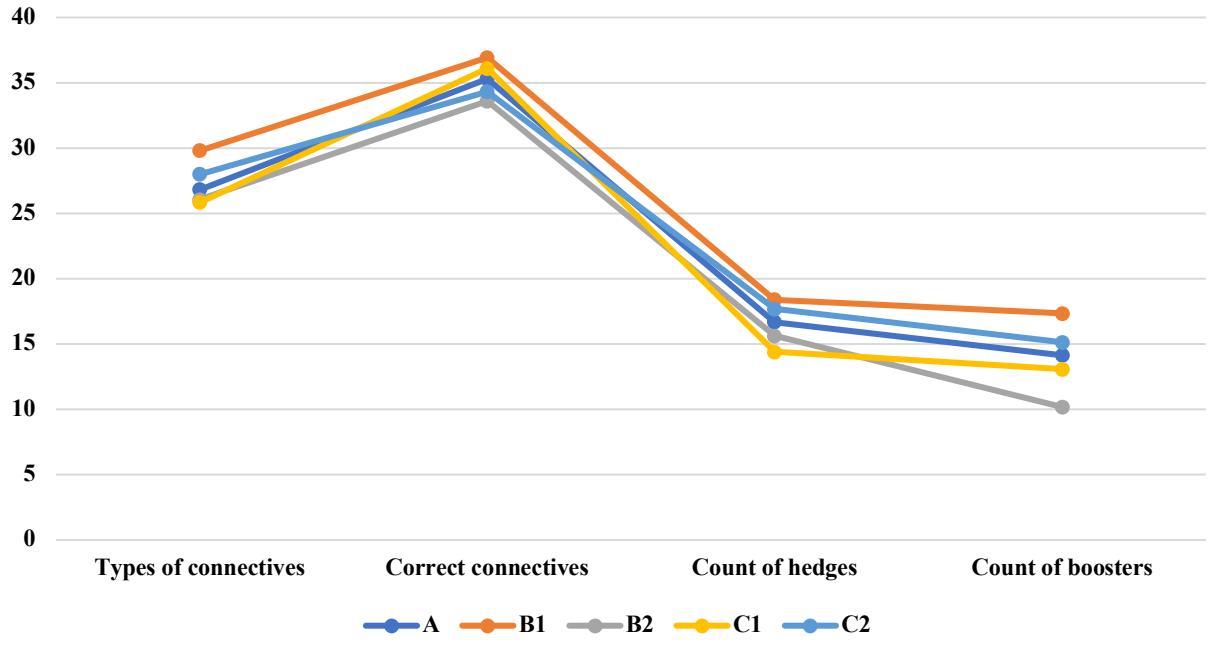
Table 12. Average counts of words, connectives, hedges, and boosters across EPT profiles.

Profile	Word Count	Connective Count	Connective Types	Mistakes of Connectives	Hedges Count	Hedges Types	Boosters Count	Boosters Types	Types of hedges + boosters	Accuracy of Connectives
A	587.55	21.15	15.43	0.35	9.83	5.13	8.53	4.20	9.33	0.98
B1	573.30	21.48	16.35	0.83	10.48	5.04	10.00	4.48	9.52	0.96
B2	528.18	17.12	13.06	0.12	8.06	5.59	5.12	3.88	9.47	0.99
C1	614.94	21.94	15.35	0.29	8.71	4.24	7.94	4.76	9.00	0.99
C2	439.09	15.22	11.78	0.43	8.00	4.48	6.83	4.17	8.65	0.98

Table 13. Normalized counts of connectives, hedges, and boosters across EPT profiles.

Profile	Types of connectives (per 1000 words)	Correct Connectives (per 1000 words)	Count of Hedges (per 1000 words)	Count of Boosters (per 1000 words)
A	26.82	35.30	16.69	14.14
B1	29.81	36.93	18.40	17.35
B2	26.06	33.60	15.63	10.18
C1	25.85	36.10	14.39	13.08
C2	27.99	34.32	17.68	15.14

Chart 4. Line chart of normalized counts of connectives, hedges, and boosters across EPT profiles.



4.3.3 Spearman's rank correlation matrix

To analyze the relationships between the essays' EPT placements and profiles and the discourse competence features, Spearman's rank correlation was conducted. While some holistic discourse competence features had strong or moderate correlations with the EPT placements or profiles, most single linguistic features did not yield significant correlations with the other features and the EPT placements and profiles (Table 7 and Table 14).

Table 14. Discourse features with strong and moderate correlations.

Coefficient Interval	Correlation	Features
0.7-0.9	Strong	EPT Placement – Average ARG ($r = 0.79$, $p < .01$) EPT Placement – Average LEX ($r = 0.76$, $p < .01$) EPT Placement – F1 Topic/Focus ($r = 0.71$, $p < .01$) EPT Placement – F1-9 Sum ($r = 0.74$, $p < .01$) Average ARG – F1 Topic/Focus ($r = 0.76$, $p < .01$) F1 Topic/Focus – F4 Body paragraphs ($r = 0.71$, $p < .01$) F1 Topic/ Focus – F1-F9 Sum ($r = 0.75$, $p < .01$) F4 Body paragraphs – F1-9 Sum ($r = 0.75$, $p < .01$)
0.4-0.6	Moderate	EPT Placement – F4 Body paragraphs ($r = 0.65$, $p < .01$) EPT Placement – F5 Conclusion ($r = 0.43$, $p < .01$) EPT Placement – F6 Theme-rheme development ($r = 0.60$, $p < .01$) Average ARG – F4 Body paragraphs ($r = 0.59$, $p < .01$) Average ARG – F6 Theme-rheme development ($r = 0.58$, $p < .01$) Average ARG – F1-9 Sum ($r = 0.66$, $p < .01$) Average LEX – F1 Topic/Focus ($r = 0.40$ $p < .01$) Average LEX – F4 Body paragraphs ($r = 0.47$, $p < .01$) Average LEX – F6 Theme-rheme development ($r = 0.40$, $p < .01$) Average LEX – F1-9 Sum ($r = 0.53$, $p < .01$) F1 Topic/Focus – F6 Theme-rheme development ($r = 0.66$, $p < .01$) F3 Introduction – F1-9 Sum ($r = 0.46$, $p < .01$) F4 Body paragraphs – F6 Theme-rheme development ($r = 0.63$, $p < .01$) F5 Conclusion – F1-9 Sum ($r = 0.53$, $p < .01$) F6 Theme-rheme development – F1-9 Sum ($r = 0.75$, $p < .01$) F7 Connective complexity – F1-9 Sum ($r = 0.40$, $p < .01$)

Based on the correlation matrix, the EPT placements have strong positive correlations with the average argumentation and lexico-grammar scores, and F1 Topic/Focus.

Since the EPT placements were decided by the EPT argumentation (r=0.79, R^2 = 0.62) and lexico-grammar scores (r=0.76, R^2 = 0.57), strong correlations are expected. Yet the moderate positive relationships between the EPT placements and F1, F4, and F6 indicate that the quality of topic/focus (r=0.71, R^2 = 0.50), body paragraphs (r=0.65, R^2 = 0.42), and theme-rheme development (r=0.60, R^2 = 0.36) might be strong indicators of an essay's overall quality and might affect the essay's final placement scores. The result suggests if an essay is competent in the theme-rheme development, its local coherence is ensured, which could help the readers understand its body paragraphs' arguments, and if the well-structured body paragraphs consistently and effectively support the controlling idea, the general coherence is achieved, and the topic is well-built.

Average ARG has a strong positive correlation with F1 Topic/Focus (r=0.76, R^2 = 0.57), and moderate correlations with F4 Body paragraphs (r=0.59, R^2 = 0.35) and F6 Theme-rheme development (r=0.58, R^2 = 0.33). Similarly, Average LEX has moderate correlations with F1 Topic/Focus (r=0.4, R^2 = 0.16), F4 Body paragraphs (r=0.47, R^2 = 0.22), and F6 Theme-rheme development (r=0.4, R^2 = 0.16). Considering F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development representing the consistency and effectiveness of the overall argumentation, the general coherence, and the local coherence, it is reasonable that the three features have stronger correlations with Average ARG than Average LEX. Although Average LEX is more about linguistic competencies, its moderate correlations with F1, F4, and F6 might suggest that clear language delivery could help effectively develop the three discourse features to some extent. However, the specific linguistic features, such as F7 Connective complexity (r=0.07), F8 Connective accuracy (r=0.17), and F9 Complexity of hedges and boosters (r=0.06), do not have significant correlations with Average LEX.

F1-9 Sum had strong or moderate correlations with the EPT placements, scores, and discourse competence features. Since F1-9 Sum reflects the general quality of the discourse competence, it is reasonable that it has strong correlations with the EPT placements (r=0.74, R^2 = 0.54), and moderate correlations with Average ARG (r=0.66, R^2 = 0.44) and Average LEX (r=0.53, R^2 = 0.28). Similarly, F1-9 Sum also has strong correlations with F1 Topic/Focus (r=0.75, R^2 = 0.56), F4 Body paragraphs (r=0.75, R^2 = 0.56), and F6 Theme-rheme development (r=0.75, R^2 = 0.56). As for the rest of the narrower discourse competence features, F1-F9 Sum has moderate correlations with F3 Introduction (r=0.46, R^2 = 0.21), F5 Conclusion (r=0.53, R^2 = 0.28), and weak correlations with F2 Thesis statement (r=0.39), F7 Connective complexity (r=0.4), and F8 Connective accuracy (r=0.22). Therefore, it could be inferred that among the nine discourse features, the order of the correlations with the overall discourse competence ranging from the strongest to the weakest is F1, F4, F6, F5, F3, F7, F2, F8, and F9.

Among the nine discourse features, F7 Connective complexity has weak correlations with some features, but F8 Connective accuracy and F9 Complexity of hedges and boosters have very weak and insignificant correlations with other variables. One possible reason is that the single linguistic feature only represents a small portion of the general linguistic competence, which makes it difficult to have a significant and observable correlation with the general discourse features or essay qualities. Another reason might be the design of the discourse competence rating rubrics. Since the essays in the dataset mostly demonstrated competent and accurate usages of various connectives, hedges, and boosters, most essays received 3 (Excellent) in the three features. Without much difference in the rating scores, it is difficult to observe any significant variations of the three features across the EPT proficiency levels and profiles. This means that the descriptors of the three discourse competence features should be modified to

better fit the level of the essays or that it would be easier to observe significant variations if the rubrics are used to analyze essays written by test takers with a wider range of English competencies.

Research Question 2: Can discourse competence features predict an argumentative essay's proficiency level?

4.3.4 Ordinal logistic regression

Due to the strong correlations between the EPT placements and F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development, to investigate if it is possible to predict an argumentative essay's placement with the three strongly correlated discourse features, ordinal logistic regression was performed. The predictor variables, F1, F4, and F6, were found to contribute to the ordinal logistic regression model (Table 15). The estimated models are:

$$logit\ (\hat{P}\ (\text{Placement} \leq 2)) = 7.71 - 1.68^{*}\ \text{F1} - 1.36^{*}\ \text{F4} - 0.78^{*}\text{F6}$$

$$logit\ (\hat{P}\ (\text{Placement} \leq 3)) = 10.42 - 1.68^{*}\ \text{F1} - 1.36^{*}\ \text{F4} - 0.78^{*}\text{F6}$$

Table 15. Coefficients and intercepts of the ordinal logistic regression.

	Value	**Std. Error**	**t value**	**p value**
F1 Topic/Focus	1.68	0.46	3.66	2.49E-04
F4 Body paragraphs	1.36	0.48	2.87	4.16E-03
F6 Theme-rheme development	0.78	0.35	2.20	2.72E-02
2|3	7.71	1.21	6.35	2.20E-10
3|4	10.42	1.46	7.15	8.42E-13

Table 16. Proportional odds ratios.

	OR	**2.50%**	**97.50%**
F1 Topic/Focus	5.39	2.29	14.14
F4 Body paragraphs	3.90	1.56	10.18
F6 Theme-rheme development	2.18	1.11	4.46

The result of multivariable ordinal logistic regression indicated that F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development were statistically significant at a 2.5% level of significance ($p < 0.01$). The odds ratio of F1 Topic/Focus (OR = 5.39, 97.5% CI, 2.29-14.14) showed that the increase of one unit in F1 Topic/Focus might make an essay 5.39 times more likely to be placed in Level 3 or Level than Level 2 if the scores in F4 and F6 are controlled. If holding F1 and F6 variables constant, the odds of an essay being placed in Level 3 or Level 4 increase 3.9 times if it receives a one-unit increase in F4 (OR = 3.9, 97.5% CI, 1.56-10.18) and if F6 increases one unit, an essay would be 2.18 time more likely to be placed as Level 3 or Level 4 than Level 2 (OR = 2.18, 97.5% CI, 1.11-4.46), holding the other variables constant.

The result of ordinal logistic regression was aligned with the correlation coefficients in Spearman's rank correlation matrix (Table 7). F1 Topic/Focus had the strongest correlation coefficient with the EPT placement ($r = 0.71$, $p < .01$) and had the largest odds ratio in the regression model (OR = 5.39). F4 Body paragraphs had a moderate correlation with the EPT placement ($r = 0.65$, $p < .01$) and had the second-largest odds ratio in the estimated model (OR = 3.9). F6 Theme-rheme development also had a moderate correlation with the EPT placement, yet the correlation coefficient was slightly smaller than the F1's and F4's ($r = 0.60$, $p < .01$), and the odds ratio was the smallest of the three (OR = 2.18). This showed that the three discourse competence features could predict an essay's final EPT placement scores to an extent. In addition, as F6 Theme-rheme development scores represented the essay's quality of local coherence, F4 Body paragraphs scores reflected an essay's general coherence, and F1Topic/Focus centered on the overall effectiveness, unity, and consistency of the argumentation, the order of the predicting power also corresponded to the composition of coherence and cohesion. If an essay's sentences are mostly logically connected without many coherence break, it ensures the theme-rheme development and local coherence, but it does not necessarily mean the essay has global

coherence and an effective overall topic building. However, if an essay successfully presents a well-developed overall argumentation centering on the controlling idea, it usually has well-structured body paragraphs and has both general and local coherence, which allows readers to follow the flow of argument easily and the essay could receive a higher placement score from the EPT raters. Thus, it is reasonable that F1 Topic/Focus has the largest odds ratio and F6 has the smallest odds ratio in this estimated ordinal logistic regression model, and F1 Topic/Focus might be the strongest predictor of an essay's EPT placement.

4.3.5 Findings of the study

This study analyzed the discourse features in argumentative essays and had three major findings: (1) F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development had the strongest correlations with the EPT placements. (2) Essays with higher EPT placement demonstrated better discourse competencies, and the mean scores of each discourse feature showed a similar tendency, but no clear pattern was found across the EPT profiles. (3) The complexity and accuracy of connectives, hedges, and boosters did not show significant strong correlations with the essay's placement, argumentation, and lexico-grammar scores.

To begin with, Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development are the three features that had the strongest correlations with the placement, which might be due to two reasons. First, the three discourse features are interconnected and critical to the overall development of argumentation. If an essay has too many coherence breaks, the isolated or irrelevant ideas would not only affect the local coherence but also weaken the clarity of the argumentation in the body paragraphs. In this case, as it is not easy to follow the argumentation in the body paragraphs, the general coherence, and the controlling idea would not be effectively supported in the whole essay. Secondly, since EPT raters usually focused on the quality of the

body paragraphs when using the holistic profile-based EPT rating rubrics (Appendix B), the quality of the essays' introduction and conclusion was not a decisive factor in the final placement scores. In addition, although it is important to have a clear thesis statement in an essay, it was not specified and emphasized in the holistic profile-based EPT rating rubrics. As long as the EPT essays had a controlling idea with a clear stance, EPT raters did not deduct any points for the missing preview of major points in the thesis statement. All these factors might lead to weaker correlations between the EPT placement and the discourse features other than F1, F4, and F6. This result is similar to the findings in Plakans and Gebril (2017) and aligned with the previous studies' suggestion that coherence, as a complex feature, might not increase linearly with writing proficiency, and some higher-level essays utilized more inferred cohesion strategies, not necessarily explicit cohesive devices (Graesser et al., 2004).

Secondly, the average scores of the discourse features were aligned with the EPT proficiency levels but not with the EPT profiles. This result is different from Wang and Xie's study (2022), where the essays of higher proficiency levels consistently performed better than the middle and lower-proficiency-level essays in the discourse features and used a greater variety of connectives. One possible reason is that the participants in Wang and Xie's study (2022) were beginner learners in English academic writing classes. Although they had nine to ten years of English learning experience, none had experience writing an English essay longer than 150 words. In contrast, the EPT test takers were required to take the EPT, for their English proficiency test scores only provided limited status admission, such as 79 to 102 as TOEFL total scores. The difference in English proficiency levels and writing experience might lead to the disparities observed across the features. For instance, almost all the essays in this study had various connectives, hedges, and boosters, which led to poor discrimination between the

proficiency levels in the F7, F8, and F9. Another possible reason is that the essays' EPT profiles were based on the original rating scores. While Rasch analysis was performed to help decide the profiles of the eight essays, which had identical average scores in argumentation and lexico-grammar, the rest of the essays' profile scores were given by the original raters. Since the rater effects in the original ratings were not completely considered, the EPT profiles labeled in the study might not accurately reflect the quality of argumentation and lexico-grammar in an essay, which might contribute to the challenges of observing the differences in the discourse feature rating scores across the EPT profiles.

Finally, F7 Connective complexity, F8 Connective accuracy, and F9 Complexity of hedges and boosters did not have significant correlations with the EPT placements and profiles, and the differences between the levels and profiles were small. As most of the essays in the study demonstrated a good command of connectives, hedges, and boosters, the scores of the three discourse features did not yield much meaningful differentiation. Furthermore, because Level 4 essays used other strategies to ensure local coherence, such as pronouns, which, and participle clauses, and these strategies were not connectives, the F7 scores only reflect part of the test takers' discourse competence. In addition, hedges were used more than boosters in most of the essays in this study, which was different from Wang and Xie's study (2022). This might be due to the different English proficiency levels of the test takers in the two studies. Since the EPT test takers had higher proficiency levels and might have more English writing experience than the participants in Wang and Xie's study (2022), most of the essays in the study had more hedges than the boosters.

CHAPTER 5: IMPLICATIONS, LIMITATIONS, AND CONCLUSION

5.1 Implications of the study

While only F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development were found to have strong correlations with the EPT placement and profiles in the study, it did not mean that the other discourse features were not important. Instead, the result reflected the test takers' performances and discourse competencies in the specific timed testing context and in the specific essay genre argumentative. Since EPT served as a placement test for the following ESL writing courses, this result could provide implications for the ESL instructors to identify the incoming ESL students' strengths and weaknesses in discourse competence across different proficiency levels. For instance, writing strategies for composing effective introductions and conclusions could be emphasized in the writing courses across the levels, for most essays received lower scores in F3 Introduction and F5 Conclusion. The ESL curriculum development team could also re-evaluate the current course materials and practices about F6 Theme-rheme development. Based on the result, being able to maintain local coherence and prevent having coherence breaks in the body paragraphs is essential to build up general coherence and strengthen the overall quality of argumentation. Since the students with EPT placement level 2 generally had some issues with local coherence, it would be better to provide them with further instructions and practices to prevent coherence breaks. Finally, the discourse rating rubrics could be a reference for the ESL instructors to evaluate the student's discourse competence in different essay genres and provide more detailed feedback on the discourse features for students' writing.

5.2 Limitations and future research

Several limitations of the study are hoped to be addressed by future studies. First, the testing contexts between this study and Wang and Xie's study (2022) are different. In Wang and

Xie's study (2022), test takers received the prompt about business dilemmas a week before and wrote a problem-solving essay in two hours in class. In the current study, test takers read the prompt and provided materials on the spot and had to finish an argumentative essay in 90 minutes. The time pressure affected their composition performance. Future studies could analyze discourse competence in ESL classroom settings where students finish essays of different genres untimed and observe the improvements from drafts to final essays along with the instructors' feedback.

Another limitation lies in the rating of discourse competence. Although the rating scale was adapted according to the current study's testing context and the original scale was developed by Wang and Xie (2022) with theories and content experts' feedback, the rating scale could be further refined. For instance, attitude markers were not included in the rating scales despite their non-neglectable effect on the reader-writer relationship. Second, the current scale contained some descriptors of language fluency, complexity, and clarity but did not consider the lexico-grammatical mistakes, which might influence the raters' evaluation of the discourse competence with the current scale. Third, the study only had one rater's rating of the 120 essays, and it is unknown how different raters would interpret the discourse feature rating rubrics, and how it would affect raters' performances. Future studies could include more raters in the rating process and develop a fine-grained rating scale that not only meets the requirements of the testing contexts but also separates the effects of lexico-grammar in the rating process.

Finally, the current study used the scores of argument development in the original EPT ratings as the criteria for the essays' argumentation quality. However, it would be better if a content expert specifically rated the essays' quality of argumentation. In addition, although Rasch analysis was performed in the study, the prompt effects and rater effects were not

significant because most of the essays were benchmark essays. Future studies could include more essays, not only the benchmark essays, to investigate if there are any prompt effects on the discourse features, and the scores of the argument development generated by Rasch could be more comparable. Since argument structure and logicality are essential to essay quality, future studies could also further investigate the relationships among argument structure, logicality, and discourse competence to enrich the theoretical frameworks.

5.3 Conclusion

In sum, this study analyzed the discourse competence features in argumentative essays and found some differences across the EPT placement levels and profiles. Among the nine discourse features, F1 Topic/Focus, F4 Body paragraphs, and F6 Theme-rheme development had the strongest correlations with the essays' placement levels. While the rest of the features only had moderate or weak correlations with the EPT placement levels and profiles, the result suggests the test takers' strengths and weaknesses in discourse competence and provides implications for the ESL instructors to develop relevant materials to improve students' discourse competence in English academic writing. The study's discourse competence rating rubrics could also be used as a reference for the ESL instructors to provide more detailed feedback on individual discourse features for students' essays. Further research could be conducted to investigate the discourse competence features in different essay genres, students' levels, and contexts.

REFERENCES

Basturkmen, H., & von Randow, J. (2014). Guiding the reader (or not) to re-create coherence: Observations on postgraduate student writing in an academic argumentative writing task. *Journal of English for Academic Purposes, 16*, 14-22. https://doi.org/10.1016/j.jeap.2014.07.005

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1–47. https://doi.org/10.1093/applin/I.1.1

Connor, U., (1984). A study of cohesion and coherence in English as a second language students' writing. *Research on Language & Social Interaction, 17*(3), 301-316. https://doi.org/10.1080/08351818409389208

Crossley, S., & McNamara, D. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32*. https://escholarship.org/uc/item/6n5908qx

Crossley, S., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 1236–1241). Austin, TX: Cognitive Science Society.

Crossley, S., Kyle, K., & McNamara, D. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing, 32,* 1-16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S., & McNamara, D. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research, 7*(3), 351-370.

https://doi.org/10.17239/jowr-2016.07.03.02

Crossley, S., Kyle, K., & Dascalu, M. (2018). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods, 51*, 14-27. https://doi.org/10.3758/s13428-018-1142-4

Dancey, C., & Reidy, J. (2007). *Statistics without math for psychology*. Pearson.

Gleason, M. M. (1999). The role of evidence in argumentative writing. *Reading & Writing Quarterly, 14*, 81–106. https://doi.org/10.1080/105735699278305

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Zhiqiang, C. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computer, 36*, 193–202. https://doi.org/10.3758/BF03195564

Hu, C., & Li, Y. (2015). Discourse connectives in L1 and L2 argumentative writing. *Higher Education Studies, 5*(4), 30–41. https://doi.org/10.5539/hes.v5n4p30

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Mortensen, L., Smith-Lock, K., & Nickels, L. (2009). Text structure and patterns of cohesion in narrative texts written by adults with a history of language impairment. *Reading and Writing: An Interdisciplinary Journal, 22*(6), 735–752. https://doi.org/10.1007/s11145-008-9150-x

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing, 31*, 98-112. https://doi.org/10.1016/j.asw.2016.08.005

Tabari, M., Johnson, M., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing, 58*. https://doi.org/10.1016/j.asw.2023.100783

Tian, Y., Kim, M., Crossley, S., & Wan, Q. (2024). Cohesive devices as an indicator of L2 students' writing fluency. *Reading and Writing, 37*, 419-441. https://doi.org/10.1007/s11145-021-10229-3

Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing, 53*. https://doi.org/10.1016/j.asw.2022.100641

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, *32*(2), 189–204. https://doi.org/10.2307/356693

Xie, Q., & Lei, Y. (2021). Diagnostic assessment of L2 academic writing product, process and self-regulation strategy use with a comparative dimension. Language Assessment Quarterly. https://doi.org/10.1080/15434303.2021.1903470

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education, 23*(1), 31-48. https://doi.org/10.1016/j.linged.2011.09.004

APPENDIX A: DISCOURSE COMPETENCE RATING SCALE

Scale	Feature	3 Excellent	2 Fair	1 Poor
Topic building	F1 Topic/Focus	The essay centers on a particular stance and presents a well-developed argument. The controlling idea is consistently and effectively supported throughout the essay.	The essay might have a clear stance, but the argument needs further development. The paragraphs support the controlling idea to a limited extent.	The essay does not have a clear stance or lacks unity. The paragraphs poorly support the argument or do not advance the controlling idea. The essay lacks necessary supporting details.
	F2 Thesis statement	The thesis statement is specific, appears at a proper place and is clearly presented. It clearly shows a stance and a preview of major points.	There is a thesis statement clearly stated; it appears at a proper place, but it is somewhat general. It does not clearly show a stance or a preview of major points.	The thesis statement appears in an improper place and lacks the stance or a preview of major points. Or there is no identifiable thesis statement.
General coherence	F3 Introduction	The opening paragraph(s) provide(s) an effective background or context to orient the reader to the topic.	There is some background or context information to orient the reader, but the idea is provided only by briefly suggesting the context or directly announcing the topic. Some irrelevant or unnecessary details are in the introduction.	The opening paragraph(s) provide(s) no background information or heavily relies on irrelevant or unnecessary details to orient the reader.
	F4 Body paragraphs	The body paragraphs have clear and effective topic sentences to organize the text and provide a logical and coherent series of arguments with ample effective supporting details. (0 topic sentence is missing)	The body paragraphs begin with an identifiable topic sentence, but some topic sentences are not effective. (1~2 missing) Some supporting details do not support the topic adequately. More effective evidence or analysis is needed.	Topic sentences are often missing or they digress from the theme. Lack of necessary evidence and analysis. The majority of supporting details are irrelevant or do not support the argument.
	F5 Conclusion	The conclusion rounds off the essay effectively with a clear (re)statement of the writer's position, a summary of the main points, and future implications.	The conclusion provides a limited summary of the main points. Or some irrelevant or redundant information is included in the conclusion.	The conclusion fails to (re)state the main points or does not create a sense of closure. Or a large proportion of new, redundant, or irrelevant information is in the conclusion.
Local coherence	F6 Theme-rheme development	The theme-rheme development is competent. No coherence breaks and unrelated idea progressions are detected.	Some local coherence breaks and unrelated idea progressions are detected in paragraphs, such as misplaced or irrelevant information or isolated ideas. (1~3 breaks)	There are frequent unrelated thematic progressions or coherence breaks within paragraphs. (≥4 breaks)
Logical connectives	F7 Connective complexity	A variety of connectives are used accurately and appropriately.	**Some simple connectives are frequently used, and they may sound mechanical.** The range of connectives used is limited.	A very limited range of connectives are used or connectives are rarely used. (Types≤5)
	F8 Connective accuracy	The connectives are used precisely and concisely, fitting the grammatical and semantic contexts.	Certain connectives are misused. (1~2 cases) *Grammar and spelling errors are not counted.	Connectives are often absent. Or there are frequent misuses of connectives. (≥3 cases)
Reader-writer interaction	F9 Complexity of hedges and boosters	There is a rich variety of hedges and boosters. Hedges and boosters are used in a proper balance overall, conveying a convincing and engaging stance. (Hedges + boosters >4, and hedges > boosters) *4=types	There is frequent use of simple hedges/boosters, but occasionally with mistakes. Or hedges or boosters sometimes do not fit the context. (Hedges + boosters >4, but hedges ≤ boosters)	Hedges and boosters are rarely used. (Hedges + boosters ≤4)

APPENDIX B: EPT WRITING RATING SCALE (2021-2022)

Writing Performance Profile							
		A Strong in both	**B1** Strong in argumentation	**B2** Strong in lexicogrammar	**C1** Weak in lexicogrammar	**C2** Weak in argumentation	**D** Weak in both
Argument Development	Central Claim	Essay provides **a well-developed argument** with a clear controlling idea and supporting details. The argumentation often shows **originality**.	Essay provides a **Sufficiently developed argument** with a clear controlling idea and supporting details.	Essay provides **a clear argument** overall with a controlling idea and supporting details.	Essay provides **a clear argument** overall with a controlling idea and supporting details.	Essay may provide a clear argument or controlling idea, but it may not be effectively supported.	Essay may provide a clear argument or controlling idea, but it may not be effectively supported.
	Paragraph-level Argumentation	- Paragraph-level argumentation has a well-developed TEA or PIE structure with **clear topic sentences** and **ample and relevant supporting details**. - The supporting details are **interconnected** to each other.	- Paragraph-level argumentation has a well-developed TEA or PIE structure with **ample and relevant supporting details**. -The supporting details are **interconnected** to each other.	- Paragraph-level argumentation has a TEA or PIE structure with **relevant supporting details but often needs further development**. - The supporting details are not consistently interconnected to each other.	- Paragraph-level argumentation has a TEA or PIE structure with **relevant supporting details but often needs further development**. - The supporting details are not consistently interconnected to each other.	- Paragraph-level argumentation is composed of **an underdeveloped TEA or PIE structure that often lacks necessary supporting details** (e.g. list-like organization of evidence). - The supporting details are not interconnected to each other.	- Paragraph-level argumentation **severely lacks supporting details**. Even if they are present, they poorly support the point. - The supporting details are not interconnected to each other and **often do not advance the controlling idea**.
Lexico-Grammatical Features		- Essay displays **sophisticated Lexico-grammar that is characteristic of academic written discourse** and conveys **finer shades of meaning precisely**. - Essay has **minor errors that are not noticeable**.	- Essay displays **command of an intermediate range of lexico-grammatical structures** that convey meaning clearly overall. - Essay has **noticeable errors** that sometimes cause lapse of clarity or precision.	- Essay displays **sophisticated Lexico-grammar that is characteristic of academic written discourse.** - Essay has **minor errors that are not noticeable**.	- Essay **lacks complex lexicogrammar** and is sometimes vague and/or unclear, but the ideas are easy to understand overall. - Essay has **noticeable errors that sometimes cause processing difficulty**.	- Essay displays **command of an intermediate range of lexico-grammatical structures** that convey meaning clearly overall. - Essay has **noticeable errors** that sometimes cause lapse of clarity or precision.	- Essay **lacks complex lexicogrammar** and is sometimes vague and/or unclear, and ideas may be obscured by language issues. - Essay has **noticeable errors that sometimes cause processing difficulty**.

APPENDIX C: COLOR-CODED ESSAYS AND RATINGS

Essay No. 4

In a world ingrained in the values of productivity, industrialization and modernization; it is unsurprising for humans to forget their roots with the environment. Jungles have been cut down only to be replaced with glass and concrete ones. This may appear to be a dire situation as cities and nature are two diametrically opposing forces, but there is a solution that encompasses both elements for our benefit. This solution is the implementation of green roofs within urban areas. Based on the lecture, green roofs are defined as roof structures that are covered by vegetation on growth material like soil and are usually classified into either intensive or extensive green roofs. The former refers to a more natural-looking roof as there are diverse flora coexisting together while the latter has a simpler look as it has less variety. Regardless, any type of green roof is better than none. Despite that, some parties believe that this proposition is not good for them due to the high cost of maintenance, extensive planning, and attraction of pests. Nonetheless, most of these complaints are only short-term problems as green roofs provide advantages that far outweigh the disadvantages. For the mass implementation of green roofs to be a reality, city governments must play a role in executing this project since they are the ones with the power to bring meaningful change. **City governments should be responsible for introducing green roofs in their respective areas as they provide the city protection from natural elements, give the city an aesthetic appeal, and save economic cost in the long run.**

First and foremost, natural setbacks can be easily combatted with the aid of green roofs. As humans, we are very susceptible to naturally occurring events such as rain and heat, hence why we resort to building structures to help protect us. Even though a normal structure can withstand a rainstorm or two, it might not survive the endless amount of problems that mother nature can unleash. That is where green roofs come in. In the lecture, it was mentioned that green roofs have helped humanity fend off from harsh weathers such as rainstorms as they give extra durability to dwellings. Recently, they also protect urban structures from excessive amounts of rain. Tall buildings have a hard time in dealing with this issue as they are easily damaged by the sheer pressure of the rainwater. The water could also drain hazardously to unwanted places. Furthermore, heat waves can cause the 'heat island' effect where buildings trap a lot of heat in between buildings. This happens as the concrete rooftops and tar roads can absorb heat efficiently but release them very slowly, causing the area to be hot. With the information from the lecture and article 1, green roofs can make the water be easily contained through precipitation as the soil absorbs water to be used up by the plants, and can be released back to the atmosphere to cool down through evapotranspiration. Not only does the building have a better drainage system, but the plants also continue to receive their own needs. Even the events of 'heat island' can be minimized as the plant layer disperses the heat it receives like in evapotranspiration. It has also been proven in article 1 that green roof temperatures can be 30-40 degrees F cooler than normal roofs. Hence, green roofs are evidently better than conventional ones in handling these issues.

Secondly, green roofs are also a guaranteed way of beautifying the cityscape. Urban areas have always had a reputation of being dull as every building looks equally the same. With green roofs, splashes of green can be found on top of buildings around the city, making even the most boring structures pop out. In article 6, it is stated that green roofs make use of wasted space. Besides the normal green grass lawn; trees, flowers, vegetables and even fruits can be cultivated on green roofs to make the environment more appealing. This means that adding on to making

the place brighter, the public can also produce their own food just on their rooftops. With this, the multiple costs of buying fruits and vegetables from a grocery store can be cut down easily. However, some trouble may arrive when pests are attracted to these gardens as it is natural for them to find a home in these places such as stated in article 5. Though many people see animals such as cockroaches, termites, and mice to be a problem, pest control should be handled by the government by allocating some funds for this exact purpose. In fact, pests are not the only animals to be attracted as useful ones such as bees and worms can also appear and help with maintaining the beauty of the garden in nature's terms. We can observe that green roofs are able to bring out a wonderful sense of calmness as the public can feel closer to nature while still being in the busy city. Therefore, no space will ever be wasted with the implementation of green roofs while it helps maintain a beautiful layout around the area.

Thirdly, green roofs can help with long-term economic cost. It can be understood for building managers and contractors to shun green roofs as they need a high initial cost in structural and maintenance costs. Based on article 2, green roofs are heavier than normal ones which means they need more structural integrity to support it. Furthermore, they also need a more carefully detailed plan before building a roof as green roofs consist of a meticulous structure with multiple layers such as the plants, soil, drainage, filtering, and waterproofing. Even costs for maintaining the green roofs are higher as the plants require watering, feeding and weeding as opposed to concrete surfaces. Despite all of this, the city government has to review the long-term economical benefits of green roofs. Based on article 3, green roofs actually are better for those willing to pay a higher initial investment. This is because the life cycle of a green roof for 50 years has been proven to be of higher monetary benefits than normal ones as they can cut energy cost and reduced stormwater fees. As discussed in paragraph 2, green roofs are able to absorb and release water and heat a lot more easily than concrete roofs. This will help the building to retain warmer temperatures in cold weather and colder temperatures in warm weather. With this, green roofs are significantly more superior than normal roofs in maintaining a sustainable economic cost.

In conclusion, it is without a doubt that city governments shoul accommodate green roofs for the public as they are able to lower prolonged expenses, liven up the area, and protect urban structures from naturally occurring problems. Just as the lecture tells us, it is no wonder that this practice has been around since antiquity as the environmental, social, and economic benefits of green roofs are very helpful for mankind as a whole. Besides that, it is encouraged that the benefits of this idea should be circulated among the people to help raise awareness of the diminishing effects of over-industrialization and rising global temperatures. When the people are educated, there is a higher chance for us to continue moving forward in building urban landscapes while remembering our origins and maintaining a healthy Earth, for a technologically advanced race on a dying planet is still a doomed race. According to the wise words of Mahatma Gandhi, "To forget how to dig the earth and to tend the soils is to forget ourselves."

Essay No.	F1 Topic/Focus	F2 Thesis statement	F3 Introduction	F4 Body paragraph	F5 Conclusion	F6 Theme-rheme development	F7 Connective compelxity	F8 Connective accuracy	F9 Compelxity of hedges and boosters	F1-F9 Sum
4	3	3	3	3	3	3	3	3	2	26

Connectives count	Types of connectives	Mistakes of connectives	Hedges (blue) count	Hedges (blue) Types	Boosters (green) count	Boosters (green) Types
51	27	0	11	7	18	6

Essay No. 97

With advancements in technology, sources of learning and gaining knowledge have evolved rapidly in the recent past. **This has brought into question the long term impact and ramifications of remote learning among the young population of the world.**

Having access to lessons in school, colleges and other learning platforms from the comfort of one's home has ensured that more number of children have the opportunity to learn. Traditional modes of education have given way to remote platforms which provide the flexibility to learn at one's pace so as to gain a sound understanding of a particular concept. This has led to an increased retention rate among children as compared to learning in an in-person environment. Remote learning has also contributed to children spending more time with their families as they save considerable amount of time that would've otherwise been spent on commuting to school.

While remote learning may have it's set of advantages, the overall development of a child is significantly impacted by the isolation it brings with it. At an age when students should be interacting with their peers, learning how to work in teams and exploring their hobbies, they are compelled to spend long hours in front of a screen. Apart from impeding their personal growth, remote learning has also led to children developing a sense of detachment from what they are being taught. Increased boredom and lack of focus are some other effects of the current learning environment of young students.

Yet another disadvantage of remote learning is how the health of a child is affected, which can be both mental and physical. Increased screen time has led to a large student population developing Myopia or short sightedness. According to a study, approximately 40% of children in Europe and North America are affected by Myopia and by 2030 this number could rise to 40% of the global population i.e 3.5 billion. Additionally, multiple students have reported increased levels of anxiety in social gatherings and fatigue when carrying out day to day household chores.

Lastly, children are not the only ones who have to face the negative consequences of learning in a remote environment. Parents, who play an integral role in a child's upbringing, have had to deal with the compounding responsibilities. Acting as an aide to the teacher, monitoring daily assignments and keeping a track of the overall learning of a child has played it's toll on the moods of parents, with parents showing signs of increased stress, anger and worry.

In conclusion, although remote learning may be a feasible mode of education for the future, it's negative consequences hinder the all round development of a child.

Essay No.	F1 Topic/Focus	F2 Thesis statement	F3 Introduction	F4 Body paragraph	F5 Conclusion	F6 Theme-rheme development	F7 Connective compelxity	F8 Connective accuracy	F9 Compelxity of hedges and boosters	F1-F9 Sum
97	1	1	1	3	1	3	3	3	3	19

Connectives count	Types of connectives	Mistakes of connectives	Hedges (blue) count	Hedges (blue) Types	Boosters (green) count	Boosters (green) Types
11	10	0	6	5	1	1