2019

Validity argument for EPT written argumentative essays

Zaha Alonazi

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Linguistics Commons

Recommended Citation
Alonazi, Zaha, "Validity argument for EPT written argumentative essays" (2019). Graduate Theses and Dissertations. 17386.
https://lib.dr.iastate.edu/etd/17386

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Validity argument for EPT written argumentative essays

by

Zaha Alonazi

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Applied Linguistics and Technology

Program of Study Committee:
Bethany Gray, Major Professor
Carol Chapelle
Abby Dubisar
Gary Ockey
Amy Froelich

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2019

Copyright © Zaha Alonazi, 2019. All rights reserved.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENT</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2. BACKGROUND OF THE STUDY</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 3. LEXICAL COMPLEXITY IN ESL WRITING</td>
<td>41</td>
</tr>
<tr>
<td>CHAPTER 4. SYNTACTIC COMPLEXITY IN ESL WRITING</td>
<td>78</td>
</tr>
<tr>
<td>CHAPTER 5. DISCOURSE ANALYSIS OF TOPICAL DEVELOPMENT</td>
<td>105</td>
</tr>
<tr>
<td>CHAPTER 6. THE IMPACT OF EPT USE ON ESL TEACHERS AND STUDENTS</td>
<td>159</td>
</tr>
<tr>
<td>CHAPTER 7. SUMMARY AND CONCLUSIONS OF FINDINGS</td>
<td>192</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>206</td>
</tr>
<tr>
<td>APPENDIX A. RUBRICS OF THE EPT WRITING</td>
<td>214</td>
</tr>
<tr>
<td>APPENDIX B. ACTFL PROFICIENCY GUIDELINES 2012-WRITING</td>
<td>216</td>
</tr>
<tr>
<td>APPENDIX C. CODING OF TOPICAL DEVELOPMENT AND ISES</td>
<td>217</td>
</tr>
<tr>
<td>APPENDIX D. SURVEY OF ESL TEACHERS</td>
<td>218</td>
</tr>
<tr>
<td>APPENDIX E. PROTOCOL OF ESL TEACHERS’ INTERVIEWS</td>
<td>221</td>
</tr>
<tr>
<td>APPENDIX F. ESL LEARNERS’ QUESTIONNAIRE</td>
<td>222</td>
</tr>
<tr>
<td>APPENDIX G. INTERVIEW PROTOCOL WITH ESL LEARNERS</td>
<td>227</td>
</tr>
<tr>
<td>APPENDIX H. INFORMED CONSENT FOR ESL STUDENTS’ INTERVIEWS</td>
<td>229</td>
</tr>
<tr>
<td>APPENDIX I. LEARNERS’ SUGGESTIONS OF ACTIVITIES FOR IMPROVING WRITING</td>
<td>231</td>
</tr>
<tr>
<td>APPENDIX J. IRB APPROVAL FOR ESL SURVEY AND INTERVIEWS</td>
<td>232</td>
</tr>
<tr>
<td>APPENDIX K. IRB APPROVAL FOR ESL LEARNERS’ SURVEY AND INTERVIEWS</td>
<td>233</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT

Looking back at my PhD journey in the US and in ISU in particular, I feel thankful for this valuable fruitful experience that enriches both my cultural and scientific knowledge. Special thanks and gratitude to the chair of my committee Dr. Bethany Gray and my committee members: Dr. Carol Chapelle, Dr. Gary Ockey, Dr. Amy Froelich and Dr. Abby Dubisar for their patience, cooperation and their constructive feedback that help culminating this project.

Heartful gratitude to my advisor DR. Bethany Gray for all the guidance and valuable comments throughout my pilot study and my final dissertation. It was through discourse analysis and grammar analysis of language which I had with Dr. Bethany that the seeds of this dissertation were first planted. They were further developed and nurtured through the courses I had with Dr. Chapelle, Dr. Ockey and Dr. Froelich. The course on assessments’ validation that I had with Dr. Chapelle and the advanced measurement courses that I had with Dr. Ockey and Dr. Amy not only resonated with questions and concerns that I have always had in mind but also opened my eyes to avenues for pursuing answers in scientifically disciplined manners. Like many ESL teachers, I always have the concern of whether my syllabi goals meet my students’ needs and whether my judgment of my students’ performance is fair enough. I owe my confidence in my ability of pursuing answers to my concerns to the theoretical and empirical research knowledge I gained under ISU roof.

I should also thank ESL teachers and students who agreed to participate in this study. Without their valuable and insightful contributions to the surveys and the interviews, a considerable proportion of this study may not be easily completed.
Achieving sufficient proficiency in academic writing is critical in university level setting. It is not surprising hence, that the admission to English speaking universities is usually conditioned not only by a particular total score from Standardized tests of English proficiency, e.g., TOEFL or IELTs but also a specific band score in writing. To further ensure students’ readiness for university level writing, many US universities use in-house developed assessments. The main goal of in-house placement tests is to identify students who may experience difficulty coping up with the demands of university level writing and hence, place them in ESL writing courses that are tailored to their linguistics needs. Unlike high-stakes tests such as standardized tests of writing in which the validity of score interpretations and use is extensively investigated, the validity of score interpretations or placement decisions of in-house placement tests of writing may not be sufficiently explored. Besides the potential lack of logistic resources, one potential reason for the lack of validity research is that unlike high stakes tests, in-house placement tests often lack a clear definition of the writing construct or the subconstructs being assessed. The extensive research that has been conducted on the construct or the multi-subconstructs of writing reflects not only its complexity but also the need for delineating the boundaries of the construct or the subconstructs being assessed. This becomes even more crucial with placement tests whose main goal is to make placement decisions based on test takers’ linguistic needs.

A major subconstruct that has been used as an anchor in determining development in writing in both placement and standardized tests is linguistic complexity. Although there is no agreed upon definition on the subconstructs of linguistics complexity, there is a general agreement that it is reflected in three main levels: lexical, syntax and discourse.
Despite being used widely in the rating rubrics of writing assessments particularly in placement tests, there is little research on how complexity features with its three major aspects can be used to explore the validity of score use and interpretations in placement tests of writing. The extent to which the variations in learners’ performance as reflected by placement decisions mirror differences in students’ ability of using complex language has not, to my knowledge, been explored with the EPT reading based writing test at Iowa State University. Also, expected to bring benefits to both teachers’ and learners, placement decisions need to be evaluated with regard to their impact on the focus of instructional practices and the development in learners’ written performance. Bearing these concerns in mind, the current study attempts to investigate the validity of score interpretations and the use of the English Placement test of writing (EPT) administered at ISU. To facilitate the process of investigating the assumptions underlying the score interpretations and the use and the type of evidence to be examined, an interpretative /use argument is used.

Two main inferences are investigated: explanation and ramification. The first connects the performance as reflected by the EPT placement decision to the construct or the subconstruct being assessed. In other words, to what extent is the performance in the EPT writing supported - as expected by theory-by learners’ use of complex language. The second inference focuses on the impact of using the EPT test. In other words, to what extent do the stakeholders namely students and teachers benefit as a result of the placement decisions.

To address these concerns, a four -part study is conducted using a mixed method approach in which quantitative results are triangulated by qualitative analysis. The examination of the explanation inference focused on a quantitative and a qualitative analysis of three main aspects of linguistic complexity: lexical, syntax and discourse which were investigated in three
separate studies. A subsample corpus of 554 texts from the EPT written responses was used to analyze lexical and syntactic complexity features. Five indices reflecting three main aspects of lexical complexity were investigated: diversity, sophistication and cohesion. The Multivariate MANOVA shows a statistically significant main effect for lexical complexity on the placement levels. The pairwise comparisons revealed statistically significant differences among the groups. The results also suggest an influence of the education level (graduate vs. undergraduate) on the complex vocabulary use. The findings revealed partial support for the explanation inference in the validity argument for the EPT writing section. The Pass graduates displayed statistically higher use of complex vocabulary in terms of diversity, sophistication and cohesion than level B students. In addition, the undergraduate Pass displayed statistically higher means of cohesive vocabulary and higher means of sophisticated vocabulary (measured by AWL, bigram range) than level C. However, the differences between undergraduate placement levels of C and B were not consistent specifically with using words from the academic word list or AWL (a measure of sophistication). In addition, no statistical differences were found between graduate students in Pass and D levels.

The analysis of frequency use of 12 syntactic features posited by previous research to be characteristics of academic writing pointed to differences in syntactic complexity among the groups. The results of syntactic complexity indicated an effect for the education level (graduate vs. undergraduates) on the frequency of use of syntactically complex structures. The syntactic analysis revealed support for the EPT validity argument. With the exception of nouns as prenominal modifiers, the graduate level students in Pass and D level showed higher frequency use of nominalization and prenominal modifiers i.e. adjectives and post nominal modifiers
(prepositional phrases, non-finite relatives, that clause complements of nouns). Their writings also exhibited relatively less frequency of finite adverbials than their lower undergraduate levels.

Using a subsample of the EPT corpus (92 texts), the examination of the third study or the third aspect of linguistics complexity i.e. discourse level complexity focused on two main dimensions: the differences in the use of topical developmental patterns and the frequency and discourse functions of the initial sentence elements or ISEs. MANOVA test was used to examine if there were statistically significant differences in the use of topical organizational patterns. The results showed that developmental patterns have a main effect with a relatively medium size effect. The between subjects test showed statistically significant differences in the use of sequential progression. The pair wise comparison using LSD post hoc test showed that the graduate Pass utilized statistically more sequential progression than Level B. In addition, the undergraduate Pass and graduate D levels used statistically higher sequential patterns than their lower undergraduate peers in C and B levels. The One-Way ANOVA test revealed statistically main effect for the frequency of ISES on the grouping variable. A Tukey HSD post hoc tests revealed that the use of ISEs by graduate level learners in Pass and D levels was statistically significantly higher than their undergraduate peers. Although the five placement levels displayed similar distribution of ISES forms namely with linking adverbials, prepositional phrases and adverbs, the Pass group-as was shown by the qualitative analysis- exhibited better awareness of the discourse functions of the ISES. The differences in both topical progression and ISEs came in support for the EPT validity argument.

The investigation of the ramification inference in the fourth study utilized surveys and interviews with teachers and students to survey the overall satisfaction with the EPT writing decisions, course materials and to examine: 1) teachers’ beliefs about aspects of good academic
writing, common problems in their ESL students’ writing and the focus of their ESL instruction,
2) ESL students’ self-rated development after ESL writing courses. Results from the survey and
interviews pointed to a partial concurrence between what teachers believe to be important for
quality writing and the linguistic aspects emphasized in ESL writing courses. The assessment of
students’ self-assessed development after ESL writing courses, however, revealed general
satisfaction and a statistically significant self-rated improvement in the overall writing ability
and in their lexical, syntactic and discourse level writing ability.

The findings from the four studies contributed to the validity argument of the EPT
writing. Some findings came with partial support while others fully support the underlying
assumptions of the explanation and ramification inferences. The findings of this dissertation are
intended to direct attention to issues, I believe, worth considering and to avenues for further
research in the EPT context and with placement tests of writing in general.
CHAPTER 1. INTRODUCTION

English Writing Placement Tests

English placement tests are commonly used in-house tests at universities in English speaking countries. These tests usually include several sections such as reading, listening, and writing to measure ESL students’ academic readiness to enroll in university level courses. Even though they are not high-stakes tests like IELTS and TOEFL, their use has substantial impact on ESL students, ESL teachers, content teachers and other stakeholders. Enrolling in university level courses requires a minimum level of writing proficiency without which both students and faculty members may struggle to achieve the course objectives. Despite the importance of placement exams, restricted resources have led to a lack of validation research on score interpretations and the use of placement tests compared to high-stakes tests. In addition to restricted resources, many placement tests may lack specifications of the language construct being examined which in turn poses an additional burden for validation research of in-house placement tests. Having no explicit statements about the construct being measured and the intended consequences of test score uses makes it harder to evaluate the underlying assumptions of the score uses and interpretations.

Similar to many in-house tests, the validity of the score interpretation of the writing section of the English Placement Test for Non-Native Speakers of English (EPT) administered at Iowa State University has not been sufficiently investigated. All students who are accepted at the university and whose first language is not English and do not meet the criteria of exemption are required to take the EPT test. The exemption criteria meant to exclude students who demonstrate high proficiency in English. The exemption hence, applies to 1) students who graduated from US high schools and whose SAT and ACT scores demonstrate high level of English proficiency, 2)
students who received their bachelor and master degrees from English speaking countries, 3) students with a score of at least 600 in paper-based TOEFL or 100 and above in Internet-based TOEFL, 4) students with a score of at least 7.5 in IELTS or a score of 7 and above in Pearson Test of English. Accordingly, the population of examinees of the EPT are within a more restricted proficiency range than the test takers of many standardized English proficiency tests.

The test includes writing, reading and oral communication skills\(^1\). Reading and writing tests are taken in one session whereas oral communication is held in another session. The placement decisions for the writing test reflect whether ESL learners need extra ESL writing courses or whether they have sufficient writing ability to proceed to First Year Communication (undergraduate students) or their normal program of study (graduate students).

Students who do not pass the test take a diagnostic test on the first day of the semester in order to identify those who might be misplaced. In English 101B, which is a lower level writing course, the syllabus focuses on writing skills at the sentence and paragraph level. English 101C and English 101D correspond to higher level courses. The focus of English 101C is on the critical thinking process and involves students in planning, developing, and revising written work for the undergraduates whereas 101D concentrates on research papers or thesis sections for graduate students.

The extent of appropriateness of the EPT score interpretations and the use of scores in placement decisions requires further validation research. In fact, little is known about the validity of the score interpretations for the writing section of the EPT. Particularly, little is known about how placement decisions are related to the linguistic characteristics posited to be indicators of competent academic writing. This is of paramount importance given that aspects related to

---

\(^1\) The reading and listening tests are no longer included in the EPT and they have been replaced with an integrated writing and an interactive oral communication test.
competent use of language and linguistic complexity have received considerable attention in both L1 and L2 research on writing (Bulté & Housen, 2014). This tendency is reflected in consideration of complexity “as a valid and basic descriptor of L2 performance, as an indicator of proficiency and as an index of language development and progress” in both L1 and L2 writing research (Bulté & Housen, 2014, p. 43). Despite its stated importance, as a sub-construct of writing, there is no unified definition for linguistic complexity. There is an overall agreement, however, that linguistic complexity is reflected mainly in the use of appropriate words and phrases, maintenance of grammar conventions and the ability to construct coherent texts with appropriate rhetorical devices (Cumming, Kantor, Powers, Santos & Taylor, 2000).

In addition to the evaluation of the EPT score-based interpretations, the evaluation of the consequences of the EPT score-based decisions has not been sufficiently be addressed. In other words, there is a need to examine the extent to which the intended outcomes of the EPT placement decisions have been achieved. One major desired outcome of using the EPT scores is to place students in courses where they can improve their writing ability.

Bearing in mind that the evaluation of consequences is at the heart of the evaluation process of test score uses and considering the relevance and the importance of complexity features to score interpretation, the current dissertation aims to investigate the validity of the score uses and interpretations of the writing section of the English Placement Test for Non-Native Speakers of English (EPT) at Iowa State university. The validation process will mainly focus on investigation of 1) the complexity of language use in the actual performance of learners at different placement levels and 2) the impact that the EPT use has on ESL teachers and learners.
The above research goals will mainly be achieved through a mixed-method research that utilizes 1) a corpus-based analysis of test takers’ written responses, (2) an examination of ESL teachers’ satisfaction with the EPT placement decisions and how the test use affects ESL teachers’ perception of academic writing, and (3) an investigation of ESL learners’ perceived improvement in academic writing after ESL writing courses. To organize the process of investigating the EPT writing score interpretations, an argument-based approach to test validation was adopted.

An Argument-Based Approach to Validation

Test scores are used to make interpretations of test takers’ performance and make decisions (e.g., placement decisions) based on test takers’ performance in a particular test in a particular time (Kane, 2006, 2013). Unlike traditional views of validity which perceive validity as an inherent attribute of the test, an argument-based approach focuses on evaluating the claims resulting from such interpretations and use. Argument-based approaches to validation provide a framework for evaluating the plausibility of the claims or interpretations and/or use made based on test scores.

Chapelle and Voss (2014) identified the following main characteristics of research using argument-based approaches: (1) the use of an interpretation/use argument which includes claims, assumptions and inferences, and (2) the use of that argument as the basis for developing a validity argument. In the interpretation/use argument, claims and inferences about the observed performance are laid out. Claims are statements or conclusions drawn based on observations of performance or “grounds” using Toulmin’s (2003) term. The link or the inferential relationship between the claims and the grounds or actual performance is referred to as an inference. Each inference then has assumptions that need to be backed by theoretical and/or practical evidence.
In other words, the interpretation/use argument is used to construct a validity argument for linking claims with evidence (Chapelle, Enright & Jamieson, 2008; Kane, 2013).

The most common frameworks that build on the argument-based approach to validate test score uses and interpretations appeared in Bachman and Palmer’s (2010) proposed assessment/use argument and in Kane’s (2006, 2013) interpretative/use argument (IUA). Bachman and Palmer’s (2010) assessment/use argument is centered on how the test scores are used. Kane’s interpretation/use argument provides a more comprehensive and a balanced approach to validation of both the interpretation and use of test scores or placement decisions. In his description of an example of an interpretation argument of a placement test, Kane (2006) discussed five inferences:

- Evaluation or “scoring inference”: the use of the scoring criteria is appropriate and as intended.
- Generalization: the score is reliable and can be used as an estimate for expected scores across parallel tasks and test forms.
- Extrapolation: the test performance is linked to relevant contexts or criteria of language use.
- Explanation: the test performance reflects a theory-based construct.
- Decision making (or using Bachman’s (2005) term utilization): the appropriateness of test scores for making decisions about admissions, placement or curriculum for test takers.

Bachman and Palmer (2010) made a further distinction between utilization and consequences, with utilization being related mainly to decision making processes and consequences referring to the impact of using the test. Chapelle, Cotos & Lee (2015) used the
term ramification inference instead of consequences. To maintain consistency of terminology with previous research (Chapelle, Cotos and Lee, 2015; Li, 2015), the term ramification will be adopted throughout this dissertation.

In addition to the five inferences proposed by Kane, the inference of domain analysis was investigated by Chapelle, Enright & Jamieson (2008) as a part of the validity argument for the TOEFL iBT. The inference of domain analysis connects the knowledge and skills revealed by the observed performance to those needed in the target language domain-based on domain analysis conducted during test development.

Even though the general meaning of the inferences is the same across different interpretation/use arguments, the details of the score interpretations generated from one test in a particular context may not be applicable to another test (Kane, 1990). In other words, each test requires its own IUA. Once developed for a test, the framework of IUA provides clear guidelines for collecting evidence that is most relevant to the interpretation and use of a given test. Providing equal weight for the evaluation of use and test score interpretations renders IUA framework to be adequate for the purpose of this study. Therefore, the IUA framework will be adopted in the present study to guide the process of validating the score uses and interpretations of the EPT writing section at Iowa State University.

**An Overview of the Study**

The validation research in this dissertation utilizes the interpretative/use argument with a specific focus on the explanation and ramification inferences (more discussion on explanation and ramification inferences is provided in the second chapter). The explanation inference is investigated by examining how the linguistic characteristics of written responses at different placement levels reflect different levels of writing ability. The ramification inference requires evidence about the consequences of placement decisions on ESL teachers and ESL students. In
the investigation of the written responses, this dissertation utilizes a corpus-based analysis of the variation of the use of linguistic features in written responses at various placement levels. The corpus-based analysis focused on linguistic complexity by examining three main areas of language use: lexical, syntactic and discourse level complexity. The written responses assigned at higher placement levels are expected to display higher level of linguistic complexity in terms of vocabulary, syntactic structures and discourse level development.

The investigation of the ramification inference focuses on the impact that the EPT use has on ESL teachers on one hand, and on ESL learners on the other hand. For this purpose, surveys and interviews are used to examine ESL teachers’ perceptions of features of academic writing and their overall satisfaction with the EPT placement decisions. In addition, the study investigates how the use of the test influences teachers’ instructional practices. A self-assessment survey of writing ability along with interviews with ESL students will be used to gauge the impact the use of EPT exerts on ESL learners’ writing ability. Specifically, the following questions are pursued to investigate the explanation and ramification inferences associated with the interpretation/use argument of the EPT writing section:

1. To what extent do ESL writers of different placement levels exhibit differences in lexical complexity, including lexical diversity, lexical sophistication, and lexical cohesion? (explanation)

2. To what extent do ESL writers of different placement levels exhibit differences in syntactic complexity features? (explanation)

3. To what extent do ESL writers of different proficiency levels differ in their use of topical progression (parallel, indirect progression, sequential, extended progression, coherence break)? (explanation)
4. To what extent do ESL writers of different proficiency levels differ in their use of initial sentence elements in terms of frequency, linguistic realizations, and semantic functions? (explanation)

5. To what extent do ESL teachers’ instructional practices concur with their beliefs about good academic writing? (ramification)

6. To what extent ESL learners’ performance of linguistic ability improve after ESL writing courses? (ramification)

**Significance of the Study**

The significance of this dissertation stems from the increasing interest in the assessment of ESL writing in general and the need to evaluate the use and interpretations of writing placement decisions in ESL context in particular. With regard to the assessment of writing, the study provides an example of how researchers can overcome the lack of detailed information on the construct of writing ability underlying a particular test. The present study illustrates how rating rubrics and research findings can be utilized to weave an interpretative/use argument. This argument is developed based on the claims underlying the test under question and the type of evidence needed to support the proposed claims. Also, the focus on linguistic complexity as a major subcomponent of writing ability demonstrates the utility of building a validity argument for a narrow sub-construct of writing ability. Unlike test developers who are required to build a validity argument for the proposed construct underlying the test, researchers can narrow their validity argument to a sub-construct of writing ability (Kane, 2013).

In terms of our understanding of the writing construct, the investigation of the explanation inference contributes to our knowledge of the construct of writing ability and how its measurement is influenced by task characteristics. In this sense, the current dissertation directs researchers’ attention to the need to consider the interrelatedness between complexity features
and task demands. The investigation of complexity at the discourse level illustrates the benefits of combining both discourse analysis and corpus linguistics in highlighting the link between form and function in language use, especially when considering the distribution of linguistic features across different writing levels.

The mixed method approach utilized in investigating the ramification of using the EPT test illustrates the importance and the need of using both surveys and interviews in triangulating and interpreting findings. It also underscores the link between placement tests and the relevant ESL classes. The evaluation of the appropriate use of test scores in placement decisions should consider how the placement decisions are reflected in the design and content of ESL writing courses so that they become compatible with learners’ linguistic needs.

The results of the current research provides evidence for the degree to which the validity argument for the EPT writing test is supported. The results may also shed light on the features posited by the literature to be indicators of quality writing and their relatedness to task characteristics. The investigation of how the use of the EPT writing test scores influences ESL instruction and learners’ self-perceived improvement provides informative insights on areas of teaching and learning that call for further attention. Both the design and the findings of this dissertation provide guidelines to other language programs interested in evaluating the use of in-house placement tests in an ESL context.

**An Overview of the Organization of This Dissertation**

This dissertation is organized into seven chapters. The current chapter serves as an introduction. The second chapter is a background of the study containing several sections. The first section provides an overview of theoretical models of writing ability. The second section describes the context of the EPT writing section, the relationship between linguistic complexity and the writing ability construct assessed by the EPT, and how the three dimensions of linguistic
complexity (lexical, syntactic and discourse level features) are used to investigate the assumptions underlying the explanation inference.

The third section reviews the interpretative argument approach to test validation research, followed by a discussion of previous validation research on the writing section of the EPT. The fourth section presents an overview of the EPT test interpretations, uses and consequences with specific focus on explanation and ramification inferences. Chapters 3-5 are each dedicated to one of the following studies addressing the explanation inference: lexical complexity, syntactic complexity and topical development (i.e., discourse complexity). The sixth chapter will address the research questions relevant to the ramification inferences, or the impact of the use of EPT placement decisions on ESL teachers’ instructional practices and ESL learners’ perceived self-improvement in writing. Each study contains sections for literature review, methodology, results and discussion. Relevant implications and avenues of research are also be presented at the end of each study. The seventh and final chapter is a synthesis of the results, bringing together conclusions and recommendations drawn from the investigation of the explanation and ramification inferences of the EPT in chapters 3-6.
CHAPTER 2. BACKGROUND OF THE STUDY

Interest in the assessment and development of writing ability has been, for decades, at the center of second language acquisition (SLA) research. Nevertheless, and despite the plethora of research on writing ability, there is no single definition of the construct of writing ability that applies to all contexts (Hamp-Lyons & Kroll, 1997). This is attributed to the fact that writing is “an act that takes place within a context, that accompanies a particular purpose, and that is appropriately shaped for its intended audience” (Hamp-Lyons & Kroll, 1997, p. 8). What represents good writing in a narrative may not necessarily represent good writing in expository writing, or in essays across a range of different academic communities (Smit, 2004). Despite the various definitions of what constitutes writing ability, linguistic complexity has been posited as an important component of language ability in general (Canale & Swain 1980; Bachman, 1990; Bachman & Palmer, 1996) and of writing ability in particular (Cumming et al. 2000; Conner & Mbeya, 2002; Kaplan & Grabe, 1996). Hence, it has often been used as a major criterion for distinguishing students of different proficiency levels in “a posteriori construct validation” research of language assessments (Rimmer, 2006, p. 497). The assumption is that as language proficiency increases, the complexity of the language used increases as well. This tendency is reflected in the scoring rubrics of standardized tests of writing (e.g., TOEFL iBT) in which aspects of complex language have a considerable scoring weight.

Similar to many language proficiency tests, the frequent use of complex linguistic structures constitutes a major part of the scoring rubric of the writing section of the EPT writing section. To understand the relatedness between linguistic complexity and the writing ability construct as assessed by the EPT, one needs to explore the role of complex language use in the theoretical models proposed for describing writing ability. In addition, contextualizing and
justifying the investigation of linguistic complexity within the framework of an interpretation /use argument is necessary to establish the context and motivation for this dissertation. Serving these purposes, this chapter has three main sections. The first section discusses in three subsections the relationship between writing ability as assessed by the EPT and its relatedness to the construct of linguistic complexity. The three subsections present an overview of writing models, a discussion of the construct of linguistic complexity adopted in this dissertation and a delineation of the relatedness between the EPT writing rubric with test and linguistic complexity. The second section in this chapter is dedicated to the discussion of various approaches to test validation research, funneling down to the benefit of using an interpretation /use argument in the present study. The third and final section will then delineate the structure of the EPT interpretation / use argument with a focus on the explanation and ramification inferences under investigation in this study.

The Writing Ability Construct for the EPT and its Relationship to Linguistic Complexity

Overview of Constructs of Writing Ability

Every test is expected to have an underlying theory of language. Such theory is expected to be reflected in the design, scoring rubrics and the subsequent scoring-based decisions. This is true with any language test including those targeting writing ability. The developments in writing assessments reflect changes in the perceptions of the relatedness of language and context. Writing is not a mere text-based product but a communicative act that is influenced by the requirements of a particular context. The relatedness between context and language ability (including writing ability) is, however, perceived differently by different language theories. Three major approaches demonstrate the interrelatedness of language ability and context: trait focused, task focused and interactionists. From trait perspectives, consistencies in performance are attributed to the consistencies in test takers’ characteristics. In other words, contextual factors
(e.g., prompts, type of task, etc.) are designed to interact with the performance and hence the ability or trait in question is intended to be generalizable regardless of the context (Bachman, 2007; Chapelle, 1998). For example, grammar, along with phonology and lexicon, in Lado’s (1961) “skills and elements” model of language ability was perceived as a discrete component of knowledge that is different from language skills (writing, speaking, listening, and reading).

To the contrary, a task-oriented approach posits the context has an important influence on the consistency of performance. Essential to a test design in this approach is the authenticity and the specificity of the context in which the task is performed. The assumption is that performance in a particular task predicts performance on similar tasks in the future. Overall, both trait and task-based approaches failed to capture the interaction between language and context in their language proficiency models (Chapelle, 1998).

Realizing the interactive nature of language, interactionists’ approach, however, takes a more balanced stance that acknowledges the interplay of both contexts and traits. Performance in an interactionists’ approach is “a sign of underlying traits influenced by the context in which it occurs, and therefore it is a sample of performance in similar contexts” (Chapelle, 1998, p.43). In other words, the dimensions of traits should be defined with reference to the context which in turn should provide a theory that justifies the linguistic choices constrained by the broader sociolinguistic context (Chapelle, 1998). The relatedness of language and context inspired models that emphasize the communicative nature of language performance including that of writing ability. Even though there is no one theory of writing ability, the models reflect attempts to conceptualize the non-static nature of writing ability and the interaction between the three major aspects of writing ability: text, cognitive processes and variables related to participants (e.g., L1, subject knowledge), setting and task characteristics among other factors (Cumming&
Riazi, 1997; Candlin & Hyland, 1999). The next paragraphs will overview some of these proposed models and how they motivated and guided the focus on linguistic complexity in source-based writing.

Canale and Swain (1980) proposed a model of communicative competence of speaking. The model encompasses four competencies: grammar, discourse, sociolinguistic and strategic competence. Although the model is primarily intended for the analysis of speaking ability, Conner and Mbeya (2002) argued that the same competencies are also needed in writing. Building on findings of linguistic and rhetorical research of written texts, Conner and Mbeya (2002) argued that the grammatical competence of writing encompasses mechanisms of spelling and punctuation in addition to vocabulary and structure while discourse competence requires besides cohesion and coherence an awareness of the discourse organization of genre. Sociolinguistic competence in Conner and Mbeya’s perspective is reflected by the appropriateness of the written genre, register and tone. Strategic competence refers to interpersonal relationships between the reader and the writer, which are reflected by the use of linguistic devices such as transitions and metatextual markers.

Even though Conner and Mbeya’s model of communicative competence of writing emphasizes the multiple interrelated competencies of writing ability, the definition of grammatical competence is restricted to the sentence level, which fails to show how grammatical competence interacts with other competencies at a higher level of discourse.

Building on theoretical research in linguistics and empirical findings from applied linguistics and test development studies, Bachman (1990) and later Bachman and Palmer (1996) proposed a model that emphasizes the interaction between the various components of the language competence construct and the context of language use. The structure of the model was
also developed based on Bachman and Palmer’s (1982) findings that grammar (morphology and syntax) and pragmatic competence (vocabulary, cohesion and organization) are related, whereas sociolinguistic competence is distinct. Accordingly, they propose that language ability encompasses two major competencies: organizational and pragmatic competencies. Organizational competence involves both textual (cohesion and organization) and grammatical (vocabulary, morphology, syntax, phonology/graphology) competence. Pragmatic competence involves illocutionary and sociolinguistic competencies. Illocutionary competence, as the name suggests, focuses on the communicative functions of the linguistic forms used, which are influenced by the situation of use. In highly formal language, complex use of language reflects sensitivity to social conventions and/or register variations (Bachman, 1990).

Even though their model is not intended to capture writing ability per se, Bachman and Palmer’s model exhibits the interrelatedness of form and function in language use. Their definition of organizational competence reflects the appropriate use of language at the sentence as well as at the discourse level (the use of markers of semantic relatedness and the adherence to the conventional development of ideas). Moreover, the strong emphasis on the association and contribution of both textual and grammatical knowledge to textual development makes it more applicable to assessment contexts.

Cumming et al.’s (2000) conceptualization of writing ability is compatible with Bachman and Palmer’s model of language competence. Similar to Bachman and Palmer’s organizational competence, they underscore three main linguistic planes of writing ability: the word, sentence, and discourse level. Cumming et al. (2000) argue that L2 writing ability depends on the writer’s “selection of appropriate words and phrases; on facility with the conventions of grammar,
punctuation, and spelling; and on the competent use of logic and rhetorical devices to sustain a reader’s attention and direction” (p. 14).

Drawing on Canale and Swain (1980) and Bachman (1990), as well as L1 research on writing, Grabe and Kaplan (1996) proposed a model that underscores the relatedness of textual features, cognitive processes involved in the writing task, and contextual variables. Contextual factors involve participants, topic, tasks, setting, and textual input and output. Cognitive processes include verbal processing and internal processing of output. Verbal processing is influenced by several elements, including short and long-term memory, language competence, and knowledge of the world. Internal processing is the process of continuous revision of output based on the goals set in an earlier stage of writing. Grabe and Kaplan’s model (1996) of writing represents the multi-dimensionality of the writing ability construct and the interplay of the various variables shaping the written product. In addition, it highlights the necessity of inspecting the textual component of writing through the lens of its surrounding context (e.g., task type, participants’ characteristics).

In short, the discussion of communicative competence models of language in general and of writing ability in particular highlight the need to connect writing ability and consequently linguistic complexity to the context and the purpose of communication. The proposed models emphasize that a definition of writing ability should consider grammatical, discourse and sociolinguistic knowledge. A major common thread, hence, among all attempts of conceptualizing writing ability is that writing ability is a multi-faceted construct of which linguistic complexity constitutes a critical component. There is also general agreement that linguistic complexity goes beyond the appropriateness of words and phrases to include adequate use of syntactic structures and rhetorical discourse functions. The interest in linguistic aspects of
texts is motivated by “the goal of helping students produce better texts, definitions of quality aside” (Polio, 2003, p. 39). Enlightened by the emphasis on the relatedness of context and language use (Bachman & Palmer, 1996; Kaplan & Grabe, 1996) and the various aspects of linguistic complexity in writing (Cumming et al., 2000), the operationalized definition of linguistic complexity in this dissertation considers word, sentence level and discourse level complexity in the academic writing assessment. The next section will further explicate the major components of the linguistic complexity construct adopted in this study.

**The Construct of Linguistic Complexity at the Lexical, Grammatical and Discourse Level**

Linguistic complexity in the current study is defined as the ability to communicate meaning using appropriate organizational patterns, a variety of simple and complex language structures, and a sophisticated level of vocabulary that is typical to the context of academic writing. The definition, hence, replicates what Bachman and Palmer (1996) refer to as organizational competence by considering word, sentence, and discourse level complexity. The operational definitions of each of the subconstructs (i.e. lexical, syntactic, and discourse levels), which are illustrated in chapters 3-5, take into account research findings of the linguistic features prominent in timed reading based argumentative essays (further discussion of the task under investigation is in section 2.1.3).

A considerable number of studies have examined how linguistic features, including aspects of lexical and syntactic complexity, correlate with or predict human ratings of written texts. Generally, three main aspects of lexical complexity have been posited to affect human judgements of writing quality: diversity, sophistication and lexical cohesion (Guo, Crossley & McNamara, 2013). Diversity refers to the range of lexical words in a text (Engber, 1995; Guo, Crossley & McNamara, 2013), while sophistication in vocabulary refers to the use of advanced academic words in a text (Laufer & Nation, 1995). Lexical cohesion is perceived as a device for
creating textual unity (Halliday & Hasan, 1976) and it can be achieved through “the selection of vocabulary by means of reiteration” (Halliday & Hasan, 1976, p. 274). High-level writers are expected to use more diverse, sophisticated, and lexically cohesive vocabulary.

In addition to lexical complexity, syntactic complexity has been designated as an important feature of writing development. The significance of syntactic complexity in writing is reflected in the scoring weight assigned to syntactic complexity in the scoring rubrics of standardized language proficiency tests such as TOEFL and IELTS as well as the huge number of studies that examined the relation between complexity, L2 language proficiency and writing development (Guo, Crossley & McNamara, 2013; Ortega, 2003; Stockwell & Harrington, 2003; Vyatkina, 2013; Wolfe-Quintero, Inagaki, & Kim, 1998; Taguchi, Crawford, & Wetzel, 2013). Syntactic complexity has been generally defined in terms of the ability to use “a wide variety of both basic and sophisticated structures [that] are available and can be accessed quickly, whereas a lack of complexity means that only a narrow range of basic structures are available and can be accessed ” (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 107). Features of syntactic complexity have been associated with linguistic structures commonly used in academic written as opposed to conversational discourse (Biber et al. 1999). In fact, corpus based research has contributed a great deal not only to our understanding of the prominent linguistic structures in academic writing as opposed to spoken language but also to how such differences are reflected at different writing proficiency levels (Biber & Gray, 2013; Biber, Gray & Poonpon, 2011; Biber, Gray & Staples, 2016).

In addition to lexical and syntactic complexity, discourse development has also been signaled as an indicative of quality writing. Nevertheless, compared to lexical and syntactic complexity, fewer studies have focused on topical developments as a feature of discourse
complexity. One reason is that analysis of topical development requires manual coding, which can be time consuming with high numbers of texts. One method that has been commonly used in investigating organizational patterns favored by proficient writers in discourse is topical structure analysis (TSA; Lautamatti, 1987; Schneider & Connor, 1990; Knoch, 2007). TSA focuses on how the topical subject of a sentence is related to a previous topic or comment in a text and what structures are utilized to maintain the flow of ideas throughout the text. Realizing that organizational patterns constitutes an important part of the scoring rubrics of the writing section of the EPT necessitates further investigation of how learners of different writing levels maintain the flow of their ideas in texts.

Overall, the distribution of complexity features, be it lexical, syntax or discourse level features have been found to be influenced by task characteristics, such as timed vs untimed and argumentative vs. narrative (Lu, 2011) and integrated vs. independent (Cumming et al., 2005). Hence, there is a need for more studies that take both task characteristics and corpus-based findings of differences between spoken and written language into consideration. More importantly, research on complexity should also be pedagogically insightful so that stakeholders (ESL teachers, ESL students and material developers) are informed of linguistic features that reflect complex use of language in writing.

Although not part of the linguistic complexity per se, factors of task characteristics and time are being taken into consideration in the investigation of the complexity features in the present dissertation. First, argumentative essays were selected because they were found to trigger the use of more syntactically complex features than narratives for example (Lu, 2011). Second, the selection of lexical, syntactic and discourse level features is based on empirical research findings of the dominant features in argumentative essays. In addition, the discussion of the
distribution of these features will take into consideration the assumption that the scores of linguistic complexity can be influenced or clouded by other contextual factors such as time pressure (Lu, 2011). Research has shown that the frequency use of complex structures tends to be higher in untimed vs. timed tasks (Lu, 2011).

To wrap up, the discussion in this section addressed the general definition of the linguistic complexity adopted in this study and its relatedness to recent conceptualizations of writing ability and empirical research findings. Relying on theoretical justifications is, however, not sufficient when the goal is to validate the use and interpretations of a particular test in a particular context. The operationalized definition of linguistic complexity needs to show relevance to the construct measured by the EPT writing test. Visiting the rating rubrics of the EPT writing becomes hence, significant for they represent “the most concrete statement of the construct being measured” (Weigle, 2002, p. 72). The next section elaborates on the relatedness between the linguistic features included in the construct of linguistic complexity and the construct of writing ability addressed by the EPT test.

**Alignment of the EPT Writing Rubric with the Constructs of Writing Ability and Linguistic Complexity**

Before discussing the relationship between the EPT writing rubric and linguistic complexity as a sub-construct of the writing ability assessed by EPT writing section, an overview of the tasks included, time and the rating process of the EPT writing section is needed.

The EPT writing section is a part of the EPT which includes in addition to writing reading and oral communication tests. The EPT writing section, which was delivered in a paper/pencil format is a type of integrated task. Students have 30 minutes to read two passages of contradicting views, write a summary of about 100-150 words and then an argumentative essay of around 250-350 words to reflect their stance toward the issue discussed in the reading
passages. Based on their performance in the writing test, ESL students are considered Pass (exempted from taking any ESL courses) or are placed in either Level B (the lower writing level for undergraduates and graduates), Level C (a higher writing level for undergraduate students), or Level D (a higher level for graduate students only).

Raters of the EPT written essays are either ESL teachers, faculty members, or graduate students at the English department at the university. The rating process commences with a norming session in which raters are asked to rate samples from writers with different writing ability levels and discuss their justification of the final decision based on performance descriptors for each level. Each essay is rated by two different raters. If there is a disagreement among the raters, the placement level that is consistent with that of the summary task (rated by a third rater) is selected.

The rating rubric include descriptors of Pass, C/D and B levels (see Appendix A). The same rating rubric is used to evaluate both tasks i.e. summary and argumentative essays in the EPT. The level descriptors are based on The American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines 2012 of writing proficiency (see Appendix B) Accordingly, the Pass level in EPT corresponds to Advanced Mid, 10C/101D in EPT corresponds to Advanced Low and 101B in the EPT corresponds to Intermediate High in ACTFL criteria. Throughout this dissertation and to facilitate the intelligibility of discussing differences among various placement levels, ACTFL proficiency levels will be used interchangeably with their corresponding EPT levels.

The EPT rating decisions, hence, result in three placement levels: Pass, C/D and B level. With the exception of minor references to graduate vs. undergraduates in the pass level (see Appendix A), the level descriptors of the EPT do not generally make a distinction between
graduate and undergraduate levels. Given that the results of the pilot study of this dissertation pointed to differences between graduates and undergraduates’ use of complex linguistic structures, the education level (graduate vs. undergraduates) will be considered in comparing the various placement levels of the EPT writing section (see section 3.2.2 in Chapter 2 for further discussion of the EPT sample and subsample).

Even though the EPT writing has no specific definition of writing ability, the scoring rubric provides a glimpse of the facets emphasized during the rating process. The descriptors of placement levels as shown in Table 2.1 corroborate the sub-components of linguistic complexity delineated in the previous discussion—by considering complexity at the vocabulary, syntax and organizational levels and by emphasizing adherence to academic writing conventions. The rating rubric for the EPT writing test includes a general description of the written responses in each level as well as the specific linguistic criteria pertinent to each of the three levels: Pass, C/D and B. The general description for example, indicates that the Pass level is expected to show few grammatical and stylistic errors and be able to compete with native speakers’ writing. The writing of C/D level is described as showing “somehow” effective presentation of formal topics yet is tarnished by redundancy and repetition. The lower level B is, however, expected to have limited ability in communicating formal topics. The rubric of placement levels of writing concentrate on the following main categories: organization, grammar and vocabulary, functional aspect, mechanics and comprehensibility. With respect to organization, higher levels are expected to use more cohesive devices and demonstrate appropriate organization of thoughts across paragraphs with no or little digression. With regard to grammar and vocabulary, higher levels are expected to use a variety of complex syntactic structures as well as an adequate use of vocabulary. Lower levels to the contrary, show more vocabulary and structural repetition. The
functional aspect evaluates the appropriateness of the linguistic features to the rhetorical function and context of use (argumentative academic writing). Hence, the functional aspect considers the use of contrast/compare language supported with clear examples as well as the use of true and untrue conditionals. It also emphasizes the proper use of formal language.

Table 2.1 *The Relatedness between the Construct of Writing in the EPT Writing Section and the Construct of Linguistic Complexity*

<table>
<thead>
<tr>
<th>EPT Levels</th>
<th>Lexical Complexity</th>
<th>Syntactic Complexity</th>
<th>Discourse Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vocabulary</td>
<td>Syntax</td>
<td>Organization</td>
</tr>
<tr>
<td>Pass</td>
<td>• Adequate vocabulary use</td>
<td>• A variety of complex syntactic structures</td>
<td>• Utilize a variety of cohesive devices • Effectively organizing the flow of thoughts in several paragraphs. • Stay on topic with little digression.</td>
</tr>
<tr>
<td>101C (UG)</td>
<td>• Fair use of vocabulary sometimes making circumlocution</td>
<td>• Fairness use of syntactic structures • Moderate repetition of the same sentence structures</td>
<td>• Use limited number of cohesive devices • Somewhat awkward or incoherent flow of thoughts. • Sometimes fail to stay on topic, making some digression</td>
</tr>
<tr>
<td>101D (G)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101B</td>
<td>• Limited use of vocabulary (repetition of the same expressions or redundancy)</td>
<td>• Mostly rely on simple sentence structure • Connect sentences largely relying on coordinate conjunctions (e.g., and, but, or) and common subordinate conjunctions (e.g., because, if, when) • Many grammatical mistakes hindering readers’ comprehension</td>
<td>• Very limited use of cohesive devices • Incoherent flow of thoughts • Lack of development of ideas</td>
</tr>
</tbody>
</table>

The scoring rubrics hence taps into sentence level (vocabulary, syntax) and discourse level complexity (cohesion, appropriate thought development). The level-descriptors that reference syntactic complexity, however, provide no reference to specific syntactic structures
that are considered complex. The only exception is the rubric descriptors for level B which point to an excessive use of subordinators (e.g., while, if, because…etc.) and conjunctions for connecting sentences.

In addition to the rating rubrics, the development of the definition of linguistic complexity and the operational definition of its subcomponents took into consideration the characteristics of the reading -based argumentative essays. Both the length and the rhetorical demands of argumentative writing trigger more use of complex structures, hence, rendering it more appropriate for investigating potential differences in written performance across levels.

In short, validation of the use and interpretation of test scores is tied mainly to the context of use and the underlying theory shaping the construct at hand. Both theoretical perspectives of language ability and the context of the assessment under question provide a valid criterion for evaluating the interpretation and use of a particular language test. Justification of contextualizing the research on linguistic complexity within the framework of interpretation /use argument calls for further discussion of the development of the concept of validity and validation approaches, which will be fleshed out below.

**Approaches to Language Assessment Validation**

Previous research on test validation was mainly concerned with how valid a test is, i.e. whether the test is assessing what it is intended to measure rather than how the scores of a test are used or the subsequent actions based on testing results. Accordingly, validation research (e.g., Wall, Calpham and Alderson, 1994) focused mostly on supporting different types of validity: face validity (the test appears to measure what it is supposed to measure), content validity (the test content is representative of the domain content), concurrent validity (the test scores match other measures of students’ abilities), or construct validity (a test measures what it
is purported to measure). More recent conceptualization of validity, however, have shifted from focusing on tests per se to how the scores are interpreted and utilized in subsequent decision-making processes. Tests are administered with the assumption that they benefit all stakeholders. Validation of the test use and interpretations is then responsible for examining the extent to which these benefits are realized. This is reiterated by the emphasis of the current *Standards for Educational and Psychological Testing*, which state that “it is the interpretations of test scores that are evaluated, not the test itself” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [APA, AERA, & NCME], 1999, p. 9).

This shift is reflected by several frameworks that embrace such a view of validity. Some studies (e.g., Guerrero, 2000) examined the validity of test-score interpretation under Messick’s (1989, 1994) unitary validity conceptualization. Messick’s unified validity framework perceives validity as an overall evaluative judgment that requires both theoretical and empirical evidence of how test scores are interpreted and used. In other words, the soundness of test score interpretation is evaluated based on the theoretical and or empirical evidence supporting the proposed use. The difficulty of adopting Messick’s approach lies in the lack of clear guidelines for researchers to accumulate and link various types of evidence.

Building on Messick’s theoretical conceptualization of validity and with the intent of presenting more practical step-by-step guidelines, Bachman and Palmer (2010) introduced a framework of test usefulness which has also been used in several validation studies (e.g., Doe, 2013). The framework is useful for research intended to focus on the use and consequences of a test. However, it does not provide a comprehensive coverage of other types of evidence related
to the interpretation of scores, the analysis of actual performance, or the generalization of performance to other similar contexts.

A more balanced framework taking both the uses and interpretations of test performance into consideration is the interpretation /use argument (IUA) proposed by Kane (2013). In agreement with recent views of validity (e.g., Messick, 1989, Bachman & Palmer, 2010), Kane (2006, 2013) perceives validity as an evaluation of test use and interpretation rather than a property of a test per se. Hence, to consider interpretations and uses as highly valid, they should be supported by appropriate evidence. Being in line with the recent views of validity, the IUA will be adopted in this dissertation to guide the process of validating the interpretation and use of the writing section of the EPT. Further description of the IUA is presented in the next section.

**An Interpretation/Use Argument Approach to Language Assessment Validation**

An IUA, according to Kane (2013, p. 2), “includes all of the claims based on the test scores (i.e., the network of inferences and assumptions inherent in the proposed interpretation and use)”. In other words, it explicates the proposed interpretation and use of a test by laying out all inferences about the observed performance, decision making processes and conclusions (Kane, 2006). The chain of inferences then indicates the research needed for backing assumptions associated with each inference. Kane’s (2006) framework for an interpretation argument has been used to develop validity arguments for high-stakes tests such as TOEFL iBT (Chapelle, Enright, & Jamieson, 2008).

Thus, the process of validation in the validity argument approach commences with the specification of the interpretation argument, which is achieved by laying out all inferences and their underlying claims or assumptions and warrants about the observed performance (Kane, 2006). Claims refer to conclusions that test makers draw based on learners’ performance. For example, one claim could be that a test taker’s writing level is not adequate to study in an
English-medium university based on his/her performance on a writing placement test. The warrant which is used to license the claim is “a held principle or a rule of thumb” (Chapelle, Enright & Jamieson, 2008, p. 6-7) about the characteristics of test takers at certain ability levels; following the previously stated claim, a potential warrant could be that learners whose writing ability is not adequate for studying in English medium universities exhibit weak abilities in using complex and well-organized language structures in their writing on the test. This claim and its warrant in the interpretation argument of the score meaning and the intended use of test scores is then used as a basis for the subsequent development of a validity argument in which the claims based on test interpretations are evaluated using theory and/or evidence – that is, assumptions underlying each warrant need to be supported by evidence. In linking inferences about the test use with evidence, an IUA aligns with the Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999) definition of validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA & NCME, 1999, p. 9). The next section discusses the interpretation/use argument for the EPT writing test using the terms expressing the six inferences that were examined by Chapelle, Enright and Jamieson (2008) in addition to ramification inference, which was discussed by Chapelle, Cotos and Lee (2015). The claims, warrants, and assumptions, and the type of evidence needed for investigating each inference will also be discussed.

An Interpretation/Use argument for the Writing Section of the EPT

The writing section of the EPT is used to provide indicators of ESL learners’ writing ability and their readiness for university level writing at Iowa State University. The placement decisions are then used to assign students who are not sufficiently prepared in ESL writing courses designed to address their linguistic needs. Explicit statements about the targeted writing ability construct and the score interpretation of the EPT writing are not stated in detail in the
rubric. For example, even though the written response is based on a reading passage, the comprehension of the passage or the adequate use of citations do not appear to be accounted for in the rating process. In addition, the rubric indicates that higher level students use more complex structures, yet, there is no operationalized definition of linguistic complexity in terms of the specific linguistic features to be used. It is then, the test evaluator’s role to examine the hidden assumptions and the possible alternatives for test score interpretations (Kane, 2006). In such cases, the goal is to explicate assumptions and “subject them to scrutiny perhaps by individuals who have different values” (Kane, 2006, p. 26). The placement decisions based on the EPT writing section reflect one holistic score for performance on two types of tasks: summary and argumentative essays. Even though the two tasks are functionally different, both are evaluated using the same rubrics.

Although the rating rubric of the EPT written responses includes no operationalized definition of the writing ability per se, it shows that aspects of complex language use in terms of vocabulary, syntax and discourse is a major component of the writing ability targeted by the EPT tasks. The level descriptors of the EPT state that high level writing is marked by the use of complex vocabulary and grammatical structures, cohesive and well-organized structures. Investigating the validity of the EPT writing use and interpretations (including inferences and decisions) requires a framework that guide the evaluation process. The validity argument approach which guided the development of a validity argument for the score interpretations and use of the TOEFL test (Chapelle, Enright, Jamieson, 2008) is used for this purpose. The interpretation/use argument as shown in Figure 2.1 connects the observed performance to the conclusions drawn about the test taker’s abilities and then to the decision-making process and consequences through a network of inferences about performance, assumptions, and warrants
supporting inferences. The interpretation/use argument then guides the type of evidence needed for the relevant claims.

The arrows in Figure 2.1 show that the inferences move in one direction from identifying skills representative of the targeted domain to the observed performance and finally to the consequences of use. The interpretation/use argument commences then by domain description (the extent to which the abilities, skills and the tasks targeted by the test are representative of the targeted domain), evaluation (appropriate quantification of the scores), generalization (consistency of scores across parallel contexts), explanation (theoretical explanation of performance), extrapolation (extrapolation of the observed performance to other contexts (e.g., classroom) and other test formats), utilization (use of scores to make decisions about placement). The meaning, assumptions and the type of backing needed for each of these inferences are explained in detail in Chapelle, Enright and Jamieson, (2008). The last inference is ramifications which indicates that the use of the test has beneficial consequences (Chapelle, Cotos & Lee, 2015). In the next paragraphs, each of these inferences will be investigated with relevance to the EPT writing along with their relevant warrants, assumptions and the type of backing needed for each inference. A summary of the EPT inferences, assumptions and backing evidence are presented in Table 2.2.
Figure 2.1 An interpretation and Use argument for the EPT Writing (Adapted from Chapelle, Enright & Jamieson, 2008)
Domain analysis inference

The first inference in the framework of interpretation/use argument is the domain analysis, which links observed performance in the EPT writing to the performance in the target domain (Chapelle, Enright & Jamieson, 2008). With respect to the writing section of the EPT, the target domain is university level academic writing at Iowa State University. The warrant is that the observed written performance on the EPT reflects the knowledge and abilities in tasks representative of those needed at the university-level writing. The assumption is then that the tasks in the targeted domain can be modelled and that the critical linguistic knowledge needed to study in the ISU context can be identified. The backing of this warrant can be obtained from 1) a need analysis survey of experts’ perceptions (namely experienced ESL teachers and content instructors) of linguistic abilities, particularly those relevant to the construct of linguistic complexity in academic writing, and 2) research that examine the construction of the assessment tasks and the extent of similarity to those in the target domain.

Evaluation inference

The second inference, i.e. evaluation, is related to the appropriate evaluation of the observed performance (Chapelle, Enright & Jamieson, 2008) or the link between the observed performance on the EPT writing and the placement decisions. It is thus, based on the warrant that the scores obtained from observing or evaluating EPT written responses are reflective of the targeted academic writing ability. The warrant of the evaluation inference can thus entail two assumptions: 1) tasks are administered in a manner appropriate enough for obtaining a score that reflects the targeted writing ability, and 2) that the scoring rubric used in evaluating the written responses are adequately constructed to provide evidence of the targeted writing ability (Chapelle, Enright & Jamieson, 2008). The backing of the first assumptions can be obtained from interviews and or surveys of test takers perceptions of factors relevant to test administration that potentially
affected their performance on the test (e.g., time, test form (paper and pencil), clarity of instructions). Evidence supporting the assumption regarding the appropriateness of scoring rubrics can come from two sources respectively: 1) judgments by experienced teachers, raters and applied linguistics of the appropriateness of the scoring rubrics in tapping aspects of linguistic complexity in argumentative essays and 2) verbal protocols of the raters’ scoring processes. In the current EPT writing section where one holistic rating scale (see section 2.1.3) is used for scoring both summary and argumentative writing tasks, it is important to examine the appropriateness of scoring rubrics in capturing linguistic features relevant to these types of writing. Experts’ judgement is needed for assessing the appropriateness of the scoring rubrics used for the integrated writing tasks. Experts in this context refer not only to ESL teachers, but also to experienced applied linguists (ALs).

The reason for including ALs is that teachers usually have a general sense of what might be appropriate and may in many cases depend on their intuition in judgment. Including applied linguistics who are experienced with register variation (spoken vs. written registers) can enhance the face validity of the scoring rubrics. The second evidence pertaining to the appropriateness of scoring rubrics can be attained from investigating how raters actually use the rubrics when they assess writing texts. Raters with various educational backgrounds, and different teaching and rating experiences may focus on different aspects of writing ability, even when using the same scoring rubrics (Barkaoui, 2007).

**Generalization inference**

The third inference in the IUA is the generalization inference, which focuses on the reliability and consistency of scoring (Kane, 2006). It is built on the warrant that the scores obtained from observing or evaluating written performance can be generalized or used as “estimates” of expected scores in parallel tasks and across different administration conditions, e.g., different raters (Chapelle, Enright & Jamieson, 2008, p. 19). One assumption is that the rating and
task specification of the EPT writing section are specified in a manner that facilitates the development of parallel tasks. The support for this assumption can be provided through: 1) evidence of score reliability; for example, studies of inter-rater reliability (the degree to which different raters give consistent estimates) and intra-rater reliability (the degree of agreement among repeated administrations of a test performed by a single rater); 2) evidence that parallel forms have similar levels of difficulty; and 3) generalizability studies that examine potential sources of errors in assessment measures (e.g., type and number of tasks and raters) with the goal of attaining reliable estimates of test takers’ performance.

**Explanation inference**

As the name suggests, the fourth inference, i.e. explanation, explains how the attributes of performance are linked to a theoretical construct. The explanation inference in this study is built on the warrant that the attributes of performance on the EPT writing test are related as expected theoretically to linguistic complexity - a major sub-construct of writing ability. The assumption underlying this warrant is that aspects of linguistic complexity will vary across placement levels as expected by theory. That is higher levels of writing are expected to have higher frequency use of complex vocabulary, complex syntactic structures and to display more elaborate textual organization. The backing for this assumption builds on an analysis linguistic complexity features at vocabulary, syntax and discourse levels in test takers’ writing.

**Extrapolation inference**

In the fifth inference, i.e. extrapolation, the observed scores are connected to an external criterion. In other words, the scores of a particular test can be compared to those obtained from another test assessing the same or similar construct. The warrant underlying this inference is that the linguistic complexity as a sub-construct of the writing ability assessed by the EPT accounts for the quality of writing in ISU context. One assumption is then that the indicators of using complex
language on the EPT written responses are related to complex language use in other criteria or assessment measures in academic setting. The backing for this assumption can be derived from correlational studies between indicators of linguistic complexity in EPT written responses and the similar indicators in other criteria, particularly: 1) scores of complex language use in standardized tests of English language proficiency (such as TOEFL or IELTS), and/or 2) self-assessments or teacher evaluations of the ability of using complex language. When teachers’ evaluation is used as a correlational criterion, considering views from students or other fellow teachers is recommended as well (Kane, 2006, p. 49).

**Utilization inference**

The sixth inference in IUA is utilization. In the TOEFL frame work of validity. Utilization is based on the warrant that the estimates of the performance quality obtained from the EPT scores are useful for making decisions about placement at ISU ESL writing courses. The assumption for this inference would then entail that the meaning of the EPT writing test scores is clearly interpretable by department administrators, teachers, test takers, and other relevant parties. Support for this assumption can be obtained from evidence of: 1) the availability of materials for score interpretation that are comprehensible to test takers and all relevant parties, i.e. reports that provide useful information on what test takers know and can do; and 2) the availability of appropriate test-preparatory materials for EPT writing test. Both types of backing can be obtained through qualitative or mixed method research where questionnaires and interviews are used.

**Ramification inference**

An interpretation argument not only evaluates the plausibility of test score interpretations, but also their consequences or pedagogical potential, which are encompassed in the ramification inference. In fact, even when the test provides accurate estimates of test takers’ abilities, this does not mean that the test and its outcomes are useful for teachers’ instructional planning (Kane, 2006).
In placement tests such as the writing section of the EPT, “an ideal course” is expected to focus on the type of competencies, skills, and abilities that students have problems with or are not mastered sufficiently (Kane, 2006). The ramification warrant for the EPT writing section is that the consequences of using the test and the decisions that are made based on the test outcomes are beneficial to learners and teachers. Two assumptions underlie this warrant, with two types of backing that utilize qualitative and quantitative methods. Relevant to learners, the first assumption is that test takers will be able to improve their ability in using complex language structures after ESL writing courses. This assumption is backed by a survey of students’ self-assessment of their writing ability with relevance to linguistic complexity before and after ESL writing courses. Pertinent to teachers, the second assumption is that the EPT writing assessment helps promote good instructional practices in ESL writing classes. This warrant is backed by a mixed method study (questionnaires and interviews) that focuses on ESL teachers, in order to evaluate: 1) teachers’ satisfaction with the curriculum objectives and materials, and 2) whether the focus of teachers’ instruction aligns with their beliefs about important features in academic writing on one hand and whether the focus of ESL instruction addresses students’ linguistic needs-as indicated by test score interpretations on the other hand.
### Table 2.2 Summary of Warrants, Assumptions and Backing Associated with each Inference in the EPT Writing Interpretation /Use Argument (adapted from Chapelle, Cotos & Lee, 2015)

<table>
<thead>
<tr>
<th>Inference</th>
<th>Warrant licensing the inference</th>
<th>Assumptions</th>
<th>Backing (numbers correspond to the relevant assumptions)</th>
</tr>
</thead>
</table>
| A. Domain Description   | The observed written performance on the EPT reflects the kind of knowledge and abilities needed for tasks representative of those at the university-level writing | 1. linguistic features characteristic of academic writing can be identified in learners’ written responses.  
2. Assessment tasks representative of those in ISU academic domain can be identified. | 1. Needs analysis survey (of ESL writing teachers and content instructors) to identify linguistic abilities important in academic writing in general and in argumentative writing in particular.  
2. Task modelling                                                                                       |
| B. Evaluation           | The observed performance on the writing section of the EPT is evaluated to provide observed scores that are reflective of the targeted writing ability. | 1. Tasks are administered in a manner appropriate enough for obtaining a score that reflects the targeted writing ability  
2. The scoring rubric used in evaluating the written responses are adequately constructed to provide evidence of the targeted writing ability. | 1.1. Verbal protocols of scoring processes.  
1.2. Expert judgements of the scoring rubrics.  
2. Interviews and/or surveys of test takers perceptions of administrative factors that potentially affected their performance on the writing test, such as time, test form (paper and pencil), clarity of instructions. |
| C. Generalization       | The scores obtained from observing or evaluating written performance are “estimates” of the scores that can be attained from similar tasks and across different administration conditions. | 1. Rating and task specification of the EPT writing section are specified in a satisfactory manner for the development of parallel tasks.  
2. Scores are consistent across tasks, raters, and occasions. | 1.1. Evidence of score reliability (inter- and intra-rater reliability)  
1.2. Evidence that parallel forms have similar levels of difficulty.  
2. Generalizability studies                                                                                     |
Table 2.2 (continued)

<table>
<thead>
<tr>
<th>Inference</th>
<th>Warrant licensing the inference</th>
<th>Assumptions</th>
<th>Backing (numbers correspond to the relevant assumptions)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D. Explanation</strong></td>
<td>The attributes of performance in the EPT writing section are related as expected theoretically to linguistic complexity, a sub-construct of writing ability.</td>
<td>1. Performance on the EPT writing test reflects test takers’ writing ability in terms of linguistic complexity features.</td>
<td>1. Discourse analysis of indicators of linguistic complexity across placement levels assigned by the EPT writing scores.</td>
</tr>
<tr>
<td><strong>E. Extrapolation</strong></td>
<td>The linguistic complexity as a sub-construct of the writing ability assessed by the EPT accounts for the quality of writing in ISU context</td>
<td>1. The observed linguistic complexity in EPT written responses is related to similar complexity indicators on other assessment criteria.</td>
<td>1. Correlational studies of indicators of linguistic complexity in EPT writing with a) indicators of linguistic complexity in test takers’ writing in standardized tests, and b) learners’ self-assessment of linguistic ability.</td>
</tr>
<tr>
<td><strong>F. Utilization</strong></td>
<td>Estimates of the writing quality obtained from EPT scores are useful for making decisions about placement in ISU ESL writing courses.</td>
<td>1. The meaning of scores is clearly interpretable by academic administrators, test takers and teachers.</td>
<td>1. The availability of materials for score interpretation that are comprehensible to test takers and all relevant parties 2) The availability of appropriate test-preparatory materials for the EPT writing test.</td>
</tr>
<tr>
<td><strong>G. Ramification</strong></td>
<td>The consequences of using writing section of the EPT and of the decisions that are subsequently made are beneficial to stakeholders.</td>
<td>1. ESL students’ ability of using complex language structures will improve after ESL courses. 2. The instructional practices in ESL writing courses will improve, and instructors use the scores to plan and organize their instructional activities.</td>
<td>1. Surveys and interviews with students’ self-assessments of their linguistic abilities before and after writing courses. 2. Surveys and interviews to examine: a) ESL writing teachers’ satisfaction with the placement decisions, as well as curriculum objectives and materials for ESL writing courses. b) The alignment between ESL teachers’ perception of good writing and the focus of their instruction. C) The alignment between the focus of teachers’ instruction and ESL students’ linguistic needs.</td>
</tr>
</tbody>
</table>
To sum up, this section discussed the seven inferences of the IUA of the EPT writing section: domain description, evaluation, generalization, extrapolation, extrapolation, utilizations and ramifications -along with their warrants, assumptions and backing evidence. Situating the context and the focus of the current dissertation requires an overview of the validation research on the EPT and the limitations that ignited this work, which is the topic of the next section.

**Extending Validation Research on the EPT**

To date, only one study has been conducted to validate the use and interpretations of the writing section of the EPT test. Li (2015) utilized the interpretation/use argument framework to evaluate the appropriateness of use and interpretation of the EPT scores of the three sections: reading, writing, listening. The focus of Li’s research was on extrapolation and ramification inferences. In his investigation of the extrapolation inference, Li examined the correlation between the EPT scores, TOEFL iBT scores, and test takers’ self-assessments of their language ability. Moderate correlations were found between TOEFL iBT and the EPT scores, and weak correlations were reported between the EPT scores and self-assessments, indicating partial support for the inference of extrapolation in the validity argument of the EPT.

The examination of the ramification inference focused on the consequences of the use of the EPT on teachers and test takers. Interviews with the stakeholders and an examination of the writing progress of ESL students were conducted. The investigation of the ramification inference indicated overall satisfaction for ESL teachers and the students of the ESL writing courses. To examine improvement of writing, Li conducted an analysis of the diagnostic writing test at the beginning and at the end of the semester. With the exception of an increase in lexical density and a decrease in phrasal coordination, the results revealed no statistically significant improvement in lexical and syntactic complexity. The need for further validation research of the writing section of the EPT arises from the fact that (1) Li’s study is based on an older test design, which was
changed effective January 2016, and that (2) the written samples analyzed by Li were taken from the diagnostic tests administered at the beginning and the end of the semester from a relatively small number of participants (18 from 101B and 14 from 101C).

The goal of the present study is to construct an interpretation /use validity argument for the writing section of the EPT, specifically focusing on the explanation and ramification inferences. The present study addresses the limitations of Li’s study by (1) analyzing the linguistic characteristics of the language produced by test takers for the writing section of the EPT, (2) investigating the validity of score interpretations for the new writing section of the EPT which has been used since Spring 2016, (3) using a corpus-based analysis of a large number of texts from different placement levels (described in detail in Chapter 3), and by (4) using self-assessment surveys to examine ESL learners’ improvement in features signified to be characteristics of academic writing. In addition, surveys and interviews with ESL instructors of writing will be conducted to 1) assess their satisfaction with the EPT placement decisions and to 2) investigate the concurrence between instructors’ views of important features in academic writing and their instructional focus.

**Summary and Conclusion**

This chapter provided an orientation to the selection and definition of the construct of writing ability, and the sub-construct of linguistic complexity used in this study. The overview of models of writing ability reflects how the conceptualization of writing has developed from a static, transferable skill to an interactive social act. The discussion of writing models also demonstrates the increasing interest in textual analysis of features posited as indicators of writing ability and the need to expand the construct of linguistic ability beyond vocabulary and grammar to cover discoursal features. In validation research, theoretical justifications for investigating a particular construct need to be compatible with the setting, purpose and the context in which a
particular test is used. The discussion in this section has illustrated that the operationalized
definition of linguistic complexity in this dissertation draws on theoretical and empirical bases,
in addition to the descriptions of linguistic features expected at different placement levels
included in the rating rubrics for the writing section of the EPT. The chapter has also outlined the
interpretation /use argument of the EPT, with a specific focus on the explanation and
ramifications inferences that are the focus of the present study.
CHAPTER 3. LEXICAL COMPLEXITY IN ESL WRITING

Ensuring that the expected scores on the EPT writing task reflect the level of linguistic complexity- a major sub-construct of EPT writing ability- is the main warrant underlying the explanation inference in the IUA of the EPT writing tasks. Linguistic complexity as indicated earlier will be examined at three levels: lexical, syntax and discourse. The focus of this chapter is on lexical complexity. Three main aspects of lexical complexity have been proposed to affect human judgement of writing quality: diversity, sophistication, and lexical cohesion (Guo, Crossley & McNamara, 2013). A considerable amount of research on writing development and test validation has examined how aspects of lexical complexity differ across integrated and independent writing tasks and how lexical complexity correlate with or predict human ratings.

Integrated tasks were found to promote the use of more sophisticated vocabulary. It was also found that the relationship between indicators of lexical complexity and writing proficiency levels may not be similar across both independent and integrated essays (Cumming, Kantor, Baba, Erdosy, Eouanzoui, & James, 2005; Knoch, Macqueen & O'Hagan, 2014; Kyle & Crossley, 2016).

This line of research is, however, still limited in terms of the number and type of tests. In fact, there is little known about aspects of lexical complexity in integrated tasks used for placement purposes. Hence, there is a need for more studies that take task characteristics into consideration. This is of a paramount importance, especially for validation research on test score use and interpretation. Linking task characteristics to corresponding aspects of lexical complexity sheds light on how lexical features included in the rubrics are in line with the task demands. From a pedagogical perspective, research on complexity attracts stakeholders’
attention (ESL teachers, ESL students, and material developers) to linguistic features reflective of complex use of language in writing.

Bearing this in mind, the current research examines aspects of lexical complexity across different levels of EPT written responses. The first section of this chapter provides a review of studies that have addressed the three aspects of lexical complexity along with the measures used in the empirical investigations. In light of the literature recommendations and findings, measures of three components of lexical complexity will then be discussed followed by the research question addressed in this study. The second section discusses the methodology including a description of the EPT corpus and the statistical analysis used for answering the research question. The third section highlights the findings from the analysis of the three main components of lexical complexity: Diversity, sophistication and lexical cohesion. A discussion of the results is presented in the third section followed by a conclusion and a discussion of implications in the fourth and fifth sections respectively.

Components of Lexical Complexity across Proficiency Levels and Task Types

Lexical Diversity

Lexical diversity, which refers to the range of lexical words in a text, has been found to be indicative of the quality of writing (Engber, 1995; Guo, Crossley & McNamara, 2013; Ong & Zhang, 2010). Lexical diversity is calculated by the ratio of different words to the total number of words in a text. That is, the lexical diversity of a text increases with the use of different words and little repetition of the words that have already been used. One common index of lexical diversity is Type/token Ratio, or TTR, which refers to the ratio of different lexical words (types) to the total number of words (tokens) in a text. However, TTR has been criticized for reflecting the length of a text rather than the range of the vocabulary (Engber, 1995). Therefore, several transformations of TTR have been proposed, including the D index, Corrected Type/token Ratio
(CTTR), Root Type/token Ratio (RTTR), and word types squared divided by the total number of words $WT^2/W$ (see Lu, 2012 for more discussion of various formulas). Another modified type/token measure of lexical diversity focuses on the ratio of lexical word types (unique word types) to the number of lexical words in a text (Cumming et al., 2005; Engber, 1995; Lu, 2012). The difference between this measure and the other type/token ratio measures is in the definition of words and the definition of types. Unlike common TTR measures, in which a word includes both content and function words, words in this modified type/token measure follow the definition of Quirk, Greenbaum, Leech, and Svartvik (1985, as cited by Engber, 1995) in referring only to open-class items: “nouns, adjectives, full verbs, and adverbs with an adjectival base, especially those with an -ly suffix” (p. 145). Also, unlike type/token measures in which different inflections of the same word are considered different types (e.g., go, goes, went, going), these inflections are considered one type in this modified type/token measure (Engber, 1995; Lu, 2012). Following Lu (2012) and in order to differentiate this modified TTR measure from common TTR measures, the modified type/token measure will be referred to as “the measure of lexical variation.”

Lexical diversity has been found to have a generally positive correlation with holistic scores of timed argumentative essays. Engber (1995) found that diversity has positive statistically moderate correlation with holistic scoring of timed argumentative essays (independent task) in both ESL intermediate and advanced level students when lexical errors were included in the count and a high correlation when lexical errors were excluded. Similar results were reported by Grant and Ginther (2000), who found that lexical type/token ratio increases with the proficiency level of the students in completing timed, independent argumentative essays.
Research also shows differences in lexical diversity when task types (integrated versus independent) are compared across various proficiency levels. Cumming et al. (2005) found a general increase of diversity as the level of proficiency increases. However, integrated tasks portrayed more diverse vocabulary than independent ones. The researchers attributed this difference to the possibility of the test takers borrowing words verbatim from the source when reading and listening. In their investigation of discourse typical of learners at different proficiency levels in the TOEFL iBT test, Knoch, Macqueen and O'Hagan (2014) found more lexical diversity was produced in high levels of writing in both independent and integrated tasks. However, the results of analyses of the integrated tasks were less revealing than those of the independent ones. Post hoc tests of integrated tasks showed either no statistical differences among any of the five proficiency levels or that the differences were only restricted to comparisons between high and low proficiency levels. Cumming et al. (2005) and Knoch, Macqueen, and O'Hagan (2014) suggest that learners produce more diverse vocabulary in integrated tasks compared to independent tasks. The studies also suggest that differences among proficiency levels, especially in lower levels, might be masked by the nature of the integrated tasks, which require using some words, especially technical ones, from the source-based reading or listening passages.

In addition to task types, task complexity and writing genres (e.g., narrative versus argumentative) have been found to affect the level of lexical diversity in writing. Examining the effect of task complexity in terms of planning time (± planning time) and the supportive content (± supportive content) on lexical complexity, Ong and Zhang (2010) found that free writing (involving no planning) and writing with supportive content produced statistically more lexical diversity. Contrary to Ong and Zhang’s (2010) findings, Kormos (2011) found no difference
between tasks of different difficulty levels (picture only/picture and supportive content) on lexical diversity. Kormos attributed the discrepancy between her results and those of Ong and Zhang (2010) to the use of narratives in Kormos study as opposed to the use of argumentative essays in Ong and Zhang’s research.

Overall, the research on lexical diversity points to a positive correlation between diversity and holistic scores in timed argumentative essays in both dependent and independent tasks. It further suggests that integrated tasks trigger more diverse vocabulary compared to independent ones. The differences in diverse use of vocabulary among groups of various proficiency levels seem, nevertheless, less apparent in integrated tasks compared to independent ones, particularly among lower level writers. The aforementioned studies also indicate that compared to narrative writing, argumentative essays spark enhanced use of diverse vocabulary.

As for the adequate measures of diversity, the findings of Engber (1995) and Cumming et al. (2005) suggest that the definition of TTR as the unique word types to the number of lexical words in a text, which was adopted by Engber (1995), is more appropriate, especially with argumentative essays in high-level writing. Another merit of restricting TTR to content words is that the use of content words was found to be a statistically significant predictor of holistic scores in ESL students’ writing (Leo, 2012). Even though the use of diverse vocabulary better reflects lexical knowledge and, hence, higher quality writing, it is not a sufficient indicator of one’s ability to use complex vocabulary. The use of advanced academic words, as will be discussed below, is important, especially in academic settings.

**Lexical Sophistication**

Lexical sophistication, another aspect of lexical complexity, refers to the use of advanced academic words in a text (Laufer & Nation, 1995). Some researchers (e.g., Knoch, Macqueen & O'Hagan, 2014) equate advanced words with academic words from the Academic Word List.
(AWL, Coxhead, 2002) and calculate the percentage of words in a text that are AWL words. Unlike the General Service List (West, 1953, cited in Bauman & Culligan, 1995) which organizes the most common words in English in K1 (first thousand common words) and K2 (second thousand common words), the AWL which was developed by Coxhead (2002) focused on words that are frequent across academic texts. The words in AWL which has 570-word families are organized in terms of their frequency in 10 lists with the first list reflecting the most frequent academic words and the last or the tenth list represents the least frequent ones. In addition, advanced academic words can also be defined in terms of their range in a large reference corpus (i.e., the number of texts in a reference corpus in which a word occurs; Guo, Crossley & McNamara, 2013; Kyle & Crossley, 2016). Facilitated by computational tools, sophisticated vocabulary use has been examined across different proficiency levels and across different tasks.

Equating advanced words with academic ones, Knoch, Macqueen, and O'Hagan (2014) used two measures for identifying lexical sophistication as a part of their TOEFL validation research of integrated versus independent writing tasks. The first measure calculates the tokens from AWL and off-word list/total number of content words. The second computes the percentage of AWL words used in a text. The results show that learners produced a larger number of sophisticated words in integrated tasks than the independent across all proficiency levels. The analysis also shows that generally higher levels produced significantly more sophisticated words than lower levels, but no statistical differences were reported among the adjacent lower levels.

In addition to the frequency of academic words in a text, the use of sophisticated vocabulary has been examined in terms of the number of text categories in a reference corpus in which a word occurs (Gries, 2008). High-range words are words that are more contextually
diverse (i.e., they can be used in various contexts and in larger number of texts) whereas narrow-range vocabulary are register-specific and are used in a smaller number of texts in a reference corpus (Kyle & Crossley, 2016, p. 17). More proficient writers are expected to use more words with a relatively low range in a reference corpus. Kyle and Crossley (2015) found that words with average range scores were negatively correlated with scores of lexical proficiency, suggesting that the use of high-range vocabulary or general-purpose vocabulary is indicative of lower lexical proficiency. In their investigation of differences in lexical sophistication between integrated and independent writing tasks of TOEFL, Kyle and Crossley (2016) found low range words to be a stronger predictor of writing quality independent tasks than in integrated tasks. The discrepancy between Kyle and Crossley (2015, 2016) findings might be attributed to the criteria of proficiency used or the type of writing examined. The results of word range were compared to scores of lexical proficiency in Kyle and Crossley’s (2015) study whereas in Kyle and Crossley (2016), the range scores were compared to holistic writing scores. Also, the type of writing tasks used in these different studies might be an influential factor. The corpus in the first study was mainly free type of writing whereas the other was timed writing.

With advances in computational tools, measures of sophisticated vocabulary covered not only individual words, but also multiword units such as bigrams and trigrams (i.e., co-occurring two and three-word sequences; Kyle & Crossley, 2016). Proficient writers are expected to adhere to multi-word phrases that are frequently used in academic language. Kyle and Crossley (2015) found that the use of frequent n-grams or multi-word expressions correlate positively with scores of lexical proficiency, indicating that the use of highly frequent n-grams signals better collocational knowledge and, therefore, greater lexical proficiency. By the same account, it is expected that multiwords that occur in larger number of texts to correlate positively with writing
quality. In addition to lexical diversity and sophistication, lexical cohesion has also been examined in lexical complexity research, as will be discussed below.

**Lexical Cohesion**

Lexical cohesion as a component of lexical complexity has also been signified as important to the quality of writing. Lexical cohesion is defined by Halliday and Hasan (1976) as “the cohesive effect achieved by the selection of vocabulary by means of reiteration or by collocation” (p. 274). Even though collocation and reiterations are suggested to be indicative of lexical cohesion, the current research will consider reiteration as a cohesion measure, while collocation will be kept under the sophisticated use of vocabulary. This decision is justified by the fact that: 1) cohesion is mainly reflected by semantic relationships between ideas, and that 2) collocational knowledge of words reflect and complement the knowledge of using advanced academic words rather than a cohesive connection of ideas per se. Repetition, synonyms (words of the same meaning) and hypernyms/hyponyms and polysemy are proposed as reiteration strategies (Halliday & Hasan, 1976). Hypernyms/hyponyms reflect hierarchical semantic relationships. Hypernym refers to a subtype or a component of a larger semantic category (e.g., a *desk* is a hypernym of *furniture*). On the other hand, *furniture* is considered a hyponym or a superordinate word. Words with more hypernyms or semantic subtypes reflect more direct and less ambiguous meaning in a text (Guo, Crossley & McNamara, 2013). Polysemous words refer to those that convey more than one sense; for example, the word *eye* has more than one sense, including a literal meaning as in *my eye hurts*, or a metaphorical meaning, as in *keep an eye on my stuff*.

Research findings have pointed to a correlation between holistic scores of writing and lexical cohesion (Guo, Crossley & McNamara, 2013; Kyle & Crossley, 2016; Meisuo, 2000). Guo, Crossley, and McNamara (2013) found a positive correlation between the hypernymy of
nouns and holistic scores in both integrated and independent writing tasks. The hypernymy score is defined as “the mean number of subordinate terms words in a text have” (Kyle & Crossley, 2016, p. 17). Similarly, using six measures of lexical complexity, Kyle and Crossley (2016) concluded that hypernym scores are statistically significant predictors of TOEFL-integrated and independent tasks. However, similar to indices of diversity and sophistication, it was found that lexical cohesion does not have the same effect across both independent and integrated task types. Hypernym indices were stronger indicators of writing quality in integrated tasks. As for polysemous scores, Gue, Crossley and McNamara (2013) found that words with low Polysemous scores (i.e., with fewer potential meanings) correlate positively with holistic scores of the independent tasks, though no relationship was found between polysemous scores and integrated writing tasks. One conclusion that may be reached here is that the three aspects of lexical complexity should be interpreted in light of the demands of task types.

**Purpose of the Study and Research Questions**

Considering the importance of lexical complexity in reflecting the development of ESL writing and, hence, differences among different placement levels, the current study examines how ESL learners at various placement levels differ in their writing with respect to aspects of lexical complexity. The following question is investigated: To what extent do ESL writers of different placement levels exhibit differences in lexical complexity, including lexical diversity, lexical sophistication, and lexical cohesion?

**Methodology**

**The EPT Corpus**

The EPT corpus (ISU EPT Corpus of Learner Writing, Release 2.0) includes the written responses for all test takers from January 2016 through January 2017 (N = 991). The corpus contains two essays for each test taker, the “Read and Summarize” task and the “Read and
Argue” task. The corpus thus, contains a total of 1982 texts and 363,144 words (ISU EPT Corpus of Learner Writing, Release 2.0). The hand written texts were transcribed verbatim and were corrected for spelling errors that may affect the reliability of corpus automatic tools (e.g., taggers or concordances). Grammatical errors, however, were not corrected. All corpus files have a header of test takers’ demographic information namely placement level, year, student status i.e. graduates vs. undergraduates, topic code indicating the prompt/topic of the written essay, L1, college, Gender and initials of the coders who spelled check the file. Two versions of the corpus are available: tagged version using Biber tagger and untagged corpus. For the purpose of lexical analysis, the untagged corpus is used. The “Read and Argue” sub-corpus, which is summarized in Table 3.1. is the focus of this study.

Table 3.1 Summary Information for the EPT Writing Task (Read and Argue Task.) Source: ISU EPT Corpus of Learner Writing (Release 2.0).

<table>
<thead>
<tr>
<th>Level</th>
<th>Number of test takers</th>
<th>Total number of Words</th>
<th>Mean Text Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>101B</td>
<td>202</td>
<td>40,676</td>
<td>202.37</td>
</tr>
<tr>
<td>101C</td>
<td>298</td>
<td>69,255</td>
<td>232.40</td>
</tr>
<tr>
<td>101D</td>
<td>157</td>
<td>37,201</td>
<td>236.95</td>
</tr>
<tr>
<td>Pass</td>
<td>334</td>
<td>86,796</td>
<td>260.65</td>
</tr>
<tr>
<td>Total</td>
<td>991</td>
<td>233,928</td>
<td>233.09</td>
</tr>
</tbody>
</table>

Corpus Sample

The EPT corpus, as shown in table 3.1, is not balanced in terms of the number of test takers in each level across semesters and also in terms of the educational level (graduate vs. undergraduates). Given that the findings from the pilot study of this dissertation suggested the need of making a distinction between placement levels occupied by graduates vs. undergraduates and the importance of using a balanced sample in statistical analysis, only a subsample of the EPT corpus is used for the analysis of lexical and grammatical complexity (see Table 3.2). The
sub sample corpus was obtained by selecting a relatively equal number of texts from the five placement levels considered in this dissertation. The file numbers of the corpus texts were first arranged in an ascendant order in an Excel sheet, then the first 100-150 file numbers - depending on the number of texts in each level-were selected from each level. The length of written responses found to impact holistic scores (Ferris, 1994). Therefore, and to minimize the potential confounding effect of length on the assigned holistic scores, only texts with 100 words and above were retained for the analysis. Headings were manually removed from all text files in the corpus sample before uploading the files to corpus analytical tools.

Although three placement levels are yielded by the EPT rating process (Pass, C/D and B levels), the grouping in this dissertation (refer to section 2.1.3 for further discussion) takes into consideration the potential influence of the education status (i.e., graduate vs. undergraduate) on the use of complex language structures. Therefore, the subsample in this dissertation is divided into five groups: Pass (graduates) Pass (undergraduates), Level D (graduates), Level C (undergraduates), and the Low-Level B (undergraduates). The mean text length of the groups ranges from 230-273 words, as shown in Table 3.2. Given that the Pass and B levels include graduate and undergraduates’ students, a comparable number of texts from both graduates and undergraduates in the Pass Level is used. However, that was not possible with Level B, because of the small number of graduate students assigned to this level (only 27 students). Therefore, only texts from undergraduates are used for Level B for the final analyses. The subsample corpus has a total of 554 texts with 138,308 total number of words.
Table 3.2 Summary Information from the Subsample from the EPT Writing Section, “Read and Argue” Essay

<table>
<thead>
<tr>
<th>Level</th>
<th>Number of Texts</th>
<th>Mean Text Length</th>
<th>Total Words</th>
<th>Number of Texts</th>
<th>Mean Text Length</th>
<th>Total Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>101</td>
<td>273</td>
<td>27530</td>
<td>101</td>
<td>263</td>
<td>26558</td>
</tr>
<tr>
<td>Level C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>101</td>
<td>230</td>
<td>22509</td>
</tr>
<tr>
<td>Level D</td>
<td>98</td>
<td>244</td>
<td>24664</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Level B</td>
<td>-</td>
<td>-</td>
<td>153</td>
<td>242</td>
<td>37047</td>
<td></td>
</tr>
</tbody>
</table>

Measures of Lexical Complexity

In the current study and in light of the literature’s recommendations, lexical complexity is considered in terms of the three components outlined above (lexical diversity, sophistication, and cohesion). Higher placement levels are expected to produce more diverse, sophisticated and cohesive lexical expressions. Indices from TALLES (Crossley, Kyle & McNamara, 2016) and the Lexical Complexity Analyzer (Lu, 2012) (free tools available online) that reflect these three dimensions, as shown in Table 3.3, are used. The lexical variation index is provided by the Lexical Complexity Analyzer while all other indices are derived from the TALLES software.

Because some researchers have pointed to positive correlations between content words and lexical complexity (Leo, 2012), this study will adopt the measure of lexical variation (LV), or “the ratio of the number of lexical word types to the total number of lexical words in a text” (Lu, 2012, p. 6), in the Lexical Complexity Analyzer. The higher the values of LV, the more diverse vocabulary are used. The excerpts (each with 32 tokens) below represent two texts with high and low LV respectively. Even though both excerpts have the same number of words, the excerpt of high diversity includes more unique words (16 vs. 13) suggesting more variation and less repetition of words than the text with low diversity value:
(1) Spending so much time staring at a iPad also deteriorates the eyesight greatly. This would bring an extra cost as well, visiting the optometrist in order to get contacts or glasses prescribed…. (P U, LV score of .9)

(2) Using robots for all kind of work eliminating the abilities, talents, and creative thinking of people. I believe that we should use our abilities, talents and creative thinking to do our task. (B Level, LV score of .4)

Sophistication is operationalized as the frequent use of AWL words, the use of high-range academic bigrams, and the average number of low-range academic words used in a text. The use of the AWL will be investigated using the index of the normed use of academic words (AWL). The AWL index reflects the use of academic words from all 10 lists developed by Coxhead (2000). Texts with higher AWL values reflect more frequent use of AWL words and hence, have more sophisticated vocabulary than those of lower values. Excerpts 3 (AWL score of .11) and 4 (AWL score of .03) below illustrate the differences between texts with high and low AWL values (AWL words are in bold):

(3) The use of iPads should be fully implemented and utilized within the heart of the youngsters these days. The very first evidence that fully support my views are the surveys conducted by…(C level, AWL score = .11)

(4) Research from many university show that students work better in class with pen and paper. It's also cost you more money to have tablet or Ipad with your phone and your laptop. (BU, AWL score = .03)

Contextual diversity or the range of academic words and bigrams -the other components of lexical sophistication-will be examined using indices of the range of academic words or AW and academic bigrams. The index of the academic words range indicates the number of texts in
the reference corpus (the Corpus of Contemporary American English or COCA) in which words from a particular text occur. Words with high range values mean that they are more common and hence, are used in a wider number of academic texts whereas words with lower range values (e.g., antifungal; Kyle & Crossley, 2016, p. 14) are less common and hence, are used in a smaller number of written texts. Accordingly, the lower values of academic words or AW range are indicative of using less common words and consequently more sophisticated vocabulary. Examples 5 and 6 (AW words are in bold) highlight the differences between two texts from Pass Level with lower AW range score and B level with relatively higher AW range score. The examples show that the words in B level are mostly common words (e.g., people, things, stand, spare, spend). Hence, they occur in a larger number of texts in the reference corpus (COCA). The words from Pass Level, however, are less common mainly from sub-list 4 (Words from AWL are in bold). In addition to AWL words, the excerpt also includes specific academic words (proliferation, unheard, exponential) with low range (averaging .02) indicating that they appear in a smaller number of texts.

(5) **Machine labor** has been integrated into the workforce for decades, and with the advancement in technology and innovations that enhances machines capabilities, the proliferation of machine is unheard of exponential over the last few years (PU, range score of .57)

(6) Some people stand by the opinion that they can have more spare time since they do not have to work so hard. As a result, people can spend more time on their hobbies and other things (B Level, range score of .67)

To the contrary of word range, commonly used academic bigrams reflect better knowledge of word combination. Therefore, they are expected to be used more in high level
writing. Examples of bigrams in Table 3.3 illustrate the differences in academic bigram knowledge between a student from B Level and Pass undergraduate with bigram range scores of .11 and .18 respectively. To the contrary of bigrams in the Pass Level, some bigrams used by B level students have zero occurrence in the academic reference corpus (bigrams with no occurrence are in bold). Others have relatively low range scores (help you, other things) indicating that they are not commonly used in the reference Academic corpus.

Table 3.3 Examples of Bigrams in Pass and B level

<table>
<thead>
<tr>
<th>Level B</th>
<th>Level PU</th>
</tr>
</thead>
<tbody>
<tr>
<td>more encourage</td>
<td>said to</td>
</tr>
<tr>
<td>match other</td>
<td>be beneficial</td>
</tr>
<tr>
<td>Choose whatever</td>
<td>plenty of</td>
</tr>
<tr>
<td>help you</td>
<td>Of labor</td>
</tr>
<tr>
<td>Other things</td>
<td>Critical thinking</td>
</tr>
</tbody>
</table>

Lexical cohesion is reflected in using semantically related words. Therefore, indices reflecting lexical cohesion (e.g., polysemy and hypernym) are all examined. The polysemy score refers to the “mean number of senses contained in content words” (Kyle & Crossley, 2016, p. 17). Due to their violation of normality assumption, polysemy indices were excluded (further discussion of assumption testing is in the coming paragraphs).

Hyponymy nouns refers to the mean number of superordinate terms a word has in a text, (Kyle & Crossley, 2016, p. 17). In calculating the hyponymy scores, indices in TAALES differentiate between the score of all senses or all meanings versus the most frequent senses or
meaning a word contains. For example, the noun *eye* has two major meanings: either referring to a part of the body or to the ability for judging items or situations. The first meaning or sense (e.g., *my eyes need further care*) is more frequent than the second sense (e.g., *William was a man of discernment, with an eye for quality*) (Collins Online Dictionary). Given that both indices tap into similar concepts, only the index of hypernymy nouns (all senses) will be used. Higher scores of hypernymy nouns indicate higher mean of subordinates in the text and hence more cohesive use of vocabulary.

Examples 7 and 8 demonstrate differences between Pass and B Level writing with hypernymy scores of 7.5 and 5.2 respectively. Unlike the excerpt from B level which is featured by repetition and lack of specific examples, the use of superordinate terms in Pass Level reflects how the clarity of nouns is promoted by examples and further elaboration (subordinate terms are in bold). For example, the superordinate term “areas” is further explicated using examples of “judgements creativity, empathy, and crafts”. Similarly, jobs that are handled by robots are further elaborated through examples of “expeditions or clearing minefields “.

(7) **Robots may excel in some areas but they're not superior to humans.**

**Humans are superior to robots; our judgements creativity, empathy, and even crafts are all better than robots… Robots should take **jobs** that humans can't do. For example: Exploring expeditions in Mars done by NASA or clearing minefields (PU, Hypernymy Noun score of 7.5)

(8) **people can adapt their own self-directed learning independence by meiming the contexts or document online. When young people face some problems academically or non-academically. The first thing up**
to their mind is searching documents online. As we known, everyone will meet any unknown problems anywhere anytime. So as time goes by people will have a good habit of self-deriet learning independence under unawareness. (B Level, Hypernymy Noun score of 5.2)

The measures used to examine the three aspects of lexical complexity are summarized in Table 3.4 along with explanations of how each index is calculated. The results of descriptive and inferential statistics of lexical complexity aspects across the EPT placement levels is the topic of the next sections.

Table 3.4 The List of Measures of Lexical Complexity Used in the Current Study

<table>
<thead>
<tr>
<th>Component of lexical complexity</th>
<th>Measure</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity</td>
<td>Lexical variation (LV)</td>
<td>The ratio of the lexical word types divided by the total number of lexical words.</td>
</tr>
<tr>
<td>Sophistication</td>
<td>Percentage of AWL words</td>
<td>Total number of AWL words in a text divided by the total number of words in a text (TAALES index guide)</td>
</tr>
<tr>
<td>Sophistication</td>
<td>Range of academic words</td>
<td>Sum of range scores (i.e., number of documents in a reference corpus-academic COCA that a word occurs in) / number of words in text with range score (TAALES index guide)</td>
</tr>
<tr>
<td>Sophistication</td>
<td>Range of academic bigrams</td>
<td>Sum of bigram range scores / number of bigrams in text with range score. (TAALES index guide). Note: The range score of each bigram is obtained from calculating the number of documents in a reference corpus (academic COCA) in which a bigram occurs.</td>
</tr>
<tr>
<td>Lexical cohesion</td>
<td>Hypernymy of nouns (all senses)</td>
<td>Sum of hypernymy scores of nouns / total number of nouns with hypernymy scores in a text (TAALES index guide).</td>
</tr>
</tbody>
</table>

Statistical Analysis

Indices reflecting lexical diversity (measured by LV), sophisticated vocabulary (measured by percentage of AWL words, the range of academic words and the range of
academic bigrams) as well as lexical cohesion (measured by hypernymy of nouns) were examined. The normality assumption was first checked using the Shapiro-Wilk test of normality and histograms. Only variables that met the normality assumption (p>.05) across all placement levels were retained. Therefore, indices of polysemy were excluded. Multicollinearity was then examined using a Pearson correlation matrix. The correlation ranged from .1 to .6, hence, multicollinearity was not an issue. Multivariate normality was examined using Mahalanobis distance (MD) at a .001 significance level. Two texts (1 from level C and 1 from B level) found to have MD<.001. However, the analysis with and without these cases show no difference; hence, it was decided to keep them. Descriptive statistics will first be used to examine trends and potential differences across the placement levels. To examine if differences are statistically significant, the five indices will then be entered in MANOVA as dependent variables with the placement level as the independent variable.

**Lexical Complexity Results**

**Lexical Diversity**

The descriptive statistics (as visualized in Figures 3.1 to 3.5) suggest some differences among the groups in their vocabulary knowledge. The use of diverse vocabulary appears to increase in connection with the writing placement level across graduates and undergraduates (Figure 3.1). However, there seem to be minor difference in means of lexical variation between graduates in Pass (M = .725, SD = .1) and D (M = .708, SD = .1) Levels. Larger differences are witnessed between the Pass and other lower undergraduate levels. The undergraduates in the Pass Level displayed higher mean of diverse vocabulary (M = .729, SD = .1) than their fellows in C (M = .693, SD = .11) and B (M = .683, SD = .13).
Lexical Sophistication

Lexical sophistication is measured by the use of AWL words, range of academic words and the range of academic bigrams. With respect to AWL use as shown in Figure 3.2, the graduates in the Pass Level slightly outperformed their counterparts in D level ($M = .82$, $SD = .02$ vs. $M = .76$, $SD = .03$). The pattern of increase does not seem to be as consistent with undergraduate levels. The undergraduate Pass level is shown to have higher means of AWL ($M = .067$, $SD = .03$) than both their fellows in C and B Levels. However, and contrary to the theoretical expectations, B level tends to have higher means of AWL ($M = .052$, $SD = .02$) than their C Level fellows ($M = .061$, $SD = .03$).
As for the second index of lexical sophistication, the range of academic words or the use of context specific words as illustrated in Figure 3.3 shows relatively consistent decrease in academic words’ range as the writing level increases indicating that higher level writers use less common words than other groups. Negligible differences are noticed between the graduates in Pass (M = .609, SD = .03) and D level (M = .614, SD = .03). As for undergraduate levels, the undergraduate Pass level displayed slightly lower range of academic words than his fellows in both B and C levels. No differences in the mean use of low range academic words is noticed between C and B levels as both have the same mean (M = .622, SD = .03).

The results from the third component of lexical sophistication revealed as shown in Figure 3.4 a steady increase in the mean of bigram range scores as the placement level increases.
indicating that higher level students in both graduate and undergraduate levels demonstrate better knowledge of two-word combination.

*Figure 3.4* Estimated Marginal Means of Academic Bigram Range across EPT Placements Levels

**Lexical Cohesion**

The descriptive statistics of lexical cohesion measure show relatively similar trend to the use of diverse and sophisticated vocabulary. Slight variation in using hypernymy nouns are witnessed between the graduates in Pass (M = 6.75, SD = .28) and D level (M = 6.65, SD = .30). A more noticeable difference is between the undergraduate Pass level and their counterparts in C and B levels with the former using higher mean of hypernymy nouns (M = .672, SD = .30) compared to C (M = 6.63, SD = .31) and B (M = 6.46, SD = .33)
Figure 3.5 Estimated Marginal Means of Hypernymy Nouns

Overall, the pattern of using complex vocabulary across graduate and undergraduate levels suggests small differences between the graduate level students in favor of Pass Level. More significant variations among placement levels is noticed in undergraduate level. The Pass level outperformed other groups in using diverse, sophisticated and cohesive lexicon. To verify the statistical significance of these variations, the results of statistical analysis will be discussed below.

Statistical Differences by Placement Level

To examine if the lexical complexity indices as a group have statistically significant effect on the placement levels of the EPT written responses, MANOVA analysis was conducted. Examination of the effect of the composite of lexical complexity indices on the placement level has shown to be statistically significant: Pillai’s trace = .309, F (20, 2192) = 9.190, p< .05, partial $\eta^2 = .077$, indicating that approximately 7.7% of multivariate variance of the lexical complexity indices, which is a small size effect, is associated with the placement or grouping factor. The subsequent univariate tests, as shown in Table 3.5, revealed a statistically significant effect (p < .05) for the placement level on all indices. Simple post hoc contrasts (as shown in Table 3.6) were conducted to understand the differences among groups. The results of pair-wise
comparisons will be presented in the following order: graduate levels, undergraduate levels and finally graduate vs. undergraduate levels.

Table 3.5 Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Lexical Variation</td>
<td>.174</td>
<td>4</td>
<td>.044</td>
<td>3.488</td>
<td>.008</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>AWL Words</td>
<td>.056</td>
<td>4</td>
<td>.014</td>
<td>22.904</td>
<td>.000</td>
<td>.143</td>
</tr>
<tr>
<td></td>
<td>Range of academic</td>
<td>.014</td>
<td>4</td>
<td>.004</td>
<td>4.632</td>
<td>.001</td>
<td>.033</td>
</tr>
<tr>
<td></td>
<td>Bigrams</td>
<td>.016</td>
<td>4</td>
<td>.004</td>
<td>8.285</td>
<td>.000</td>
<td>.057</td>
</tr>
<tr>
<td></td>
<td>Hypernym nouns (all senses)</td>
<td>6.434</td>
<td>4</td>
<td>1.608</td>
<td>16.996</td>
<td>.000</td>
<td>.110</td>
</tr>
</tbody>
</table>

Bonferroni post hoc tests revealed that, with regard to the diversity of lexical words, no statistically significant differences were found between graduate students in Pass and D levels. As for undergraduates, Pass Level shows statistically higher mean (p<.05) of diverse vocabulary than their peers in B Level. When comparing graduate to undergraduate levels, the graduate Pass level has outperformed B Level in the use of varied vocabulary with statistically significant difference (p = .04 < .05). With respect to sophisticated vocabulary (measured by AWL words, range of Academic words and range of academic bigrams), the results show and similar to those of lexical diversity that the differences between graduate level students are not statistically significant. Accordingly, the reporting of results will focus on undergraduate levels and
differences between graduates and undergraduate levels in using sophisticated vocabulary. The use of AWL by the Pass undergraduates is shown to be statistically higher than their undergraduates peers in C Level (P = 0.0, P < .05). The differences in the use of AWL words are, however, not as consistent between adjacent lower level undergraduates as B group is found to use significantly more AWL words (P =0.03, P< .05) than C Level. The results with regard to the use of AWL words revealed also differences between graduate and undergraduate levels. Both the graduates in Pass and D level outperformed the undergraduates in C and B Levels in the mean use of AWL words and the differences were statistically significant (p = 0.00).

Contrary to the results from AWL use, the results from the range of academic words revealed that the differences between placement levels in undergraduate levels were not statistically significant (P>.05). Nevertheless, the pairwise comparisons of graduate vs. undergraduate levels disclosed statistically significant differences. Graduate students in Pass Level used academic words with statistically lower range score than their undergraduate peers in C Level (p = .01, p <.05) and B Level (p = .01, p <.05) with a mean difference of .13 and .012, respectively.

The findings of bigram range further support the variation in lexical sophistication between the undergraduates. The undergraduate writers in Pass Level are shown to use academic bigrams of statistically higher range (p = .00 <.05) than B level with a mean difference of .012. In addition, the undergrads in C level are shown to statistically outperformed their fellows in B Level in using bigrams of high range score (p = .01<.05). Statistically significant differences were also found between graduate vs. undergraduate levels. Graduate Pass are shown to use academic bigrams of statistically higher range (p = .00, p<.05) than B level.
The findings of lexical cohesion from the index of hypernymy nouns (all senses) revealed also similar results to those of lexical diversity and sophistication indices. The undergraduate Pass used statistically higher means of hypernymy nouns ($p = 0.00, p < .05$) than B level students. In addition, C Level writers displayed statistically higher means ($p = 0.00, p < .05$) of hypernymy nouns than the B level students. When considering differences between graduate vs. undergraduate levels, both graduate Pass and D levels used statistically higher means of hypernymy nouns ($p = 0.00, p < .05$) than students in B level. Further elaboration on the findings of lexical complexity analysis and how they are demonstrated in actual students’ writing will be fleshed out in the next section.

Table 3.6 Post hoc Tests of Lexical complexity

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(I) Level</th>
<th>(J) Level</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV</td>
<td>Bonferroni</td>
<td>Pass(G)</td>
<td>Pass(UG)</td>
<td>-.004</td>
<td>.016</td>
<td>1.000</td>
<td>-.048</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>.017</td>
<td>.016</td>
<td>1.000</td>
<td>-.028</td>
<td>-.061</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.027</td>
<td>.016</td>
<td>.928</td>
<td>-.018</td>
<td>-.071</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.042*</td>
<td>.014</td>
<td>.037</td>
<td>.001</td>
<td>.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pass(UG)</td>
<td>D</td>
<td>.020</td>
<td>.016</td>
<td>1.000</td>
<td>-.024</td>
<td>-.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.030</td>
<td>.016</td>
<td>.561</td>
<td>-.014</td>
<td>-.075</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.046*</td>
<td>.014</td>
<td>.016</td>
<td>.005</td>
<td>.086</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>C</td>
<td>.010</td>
<td>.016</td>
<td>1.000</td>
<td>-.035</td>
<td>-.055</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.025</td>
<td>.014</td>
<td>.815</td>
<td>-.015</td>
<td>-.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>B</td>
<td>.015</td>
<td>.014</td>
<td>1.000</td>
<td>-.026</td>
<td>-.056</td>
<td></td>
</tr>
<tr>
<td>AW Words</td>
<td>Bonferroni</td>
<td>Pass(G)</td>
<td>Pass(UG)</td>
<td>.015*</td>
<td>.003</td>
<td>.000</td>
<td>.005</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>.006</td>
<td>.003</td>
<td>.982</td>
<td>-.004</td>
<td>-.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.030*</td>
<td>.004</td>
<td>.000</td>
<td>.020</td>
<td>.040</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.020*</td>
<td>.003</td>
<td>.000</td>
<td>.011</td>
<td>.029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pass(UG)</td>
<td>D</td>
<td>-.009</td>
<td>.003</td>
<td>.100</td>
<td>-.019</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.015*</td>
<td>.004</td>
<td>.000</td>
<td>.005</td>
<td>.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.005</td>
<td>.003</td>
<td>.952</td>
<td>-.004</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>C</td>
<td>.024*</td>
<td>.004</td>
<td>.000</td>
<td>.014</td>
<td>.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.014*</td>
<td>.003</td>
<td>.000</td>
<td>.005</td>
<td>.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>B</td>
<td>-.010*</td>
<td>.003</td>
<td>.030</td>
<td>-.019</td>
<td>-.001</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.7 (continued)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(I) Level</th>
<th>(J) Level</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of academic words</td>
<td>Bonferroni</td>
<td>Pass(G)</td>
<td>Pass(UG)</td>
<td>-0.04</td>
<td>0.04</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>-0.05</td>
<td>0.04</td>
<td>1.000</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>-0.013*</td>
<td>0.04</td>
<td>0.013</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>-0.012*</td>
<td>0.04</td>
<td>0.006</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pass (UG)</td>
<td>D</td>
<td>-0.01</td>
<td>0.04</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>-0.09</td>
<td>0.04</td>
<td>0.206</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>-0.09</td>
<td>0.04</td>
<td>0.148</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>-0.08</td>
<td>0.04</td>
<td>0.371</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>-0.08</td>
<td>0.04</td>
<td>0.291</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>0.00</td>
<td>0.04</td>
<td>1.000</td>
<td>-0.010</td>
</tr>
<tr>
<td>Range of academic Bigrams</td>
<td>Bonferroni</td>
<td>Pass(G)</td>
<td>Pass(UG)</td>
<td>0.01</td>
<td>0.03</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>0.09</td>
<td>0.03</td>
<td>0.057</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>0.005</td>
<td>0.03</td>
<td>1.000</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.014*</td>
<td>0.03</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pass(UG)</td>
<td>D</td>
<td>0.007</td>
<td>0.03</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>0.003</td>
<td>0.03</td>
<td>1.000</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.012*</td>
<td>0.03</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>-0.004</td>
<td>0.03</td>
<td>1.000</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>0.005</td>
<td>0.03</td>
<td>0.596</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.009*</td>
<td>0.03</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>Hypernym nouns (all senses)</td>
<td>Bonferroni</td>
<td>Pass(G)</td>
<td>Pass(UG)</td>
<td>.026</td>
<td>0.043</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>.097</td>
<td>0.043</td>
<td>0.248</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.114</td>
<td>0.044</td>
<td>0.093</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.282*</td>
<td>0.039</td>
<td>0.000</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pass(UG)</td>
<td>D</td>
<td>.071</td>
<td>0.043</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.088</td>
<td>0.044</td>
<td>0.453</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.256</td>
<td>0.039</td>
<td>0.000</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>.017</td>
<td>0.044</td>
<td>1.000</td>
<td>-1.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>.185*</td>
<td>0.039</td>
<td>0.000</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>.169</td>
<td>0.040</td>
<td>0.000</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Discussion

Lexical Diversity

For lexical diversity, the results were in line with previous research, which shows that the use of diverse vocabulary is indicative of quality writing (Engber, 1995; Knoch, Macqueen & O’Hagan, 2014; Ong & Zhang, 2010). The descriptive statistics revealed that the higher-level writers in Pass Level produced more diverse texts than their lower level counterparts in C and B levels. However, the difference was only statistically significant between the Pass and B levels. Examples 9 and 10 below illustrate the difference in using diverse lexical words between the Pass and B Level students. (Note: unique words are in bold). Although the first excerpt contains
less tokens (39 tokens) than the one from Level B (53 tokens), it has more diverse words. The
lack of diversity in Level B and as demonstrated by example 10 is featured by the frequent
repetitions of words and phrases. The lack of statistical difference between writers at the adjacent
graduate levels and adjacent undergraduate levels might be attributed to the nature of reading-
based integrated tasks where the distinction in the use of diverse vocabulary - as was pointed out
by findings from previous research (Knoch, Macqueen & O'Hagan, 2014) - becomes less
apparent.

(9) The past few decades have witnessed a dramatic change in education tools with
the advent of iPads and laptops. Some people hold the notion that the technical
devices will distract student from studying, because of games and web
applications. (PU, LV score of .9).

(10) Robots take over some of jobs today, and one day, they maybe will take over all
works from human, on that time, no people going to work, robots will do
everything for human, and be fat without exercises, and then, one day all human
are dead, robots will take over the whole earth. (B, LV score of .4)

In sum, the findings provide further evidence that high level writers utilize wider range of
vocabulary. The descriptive statistics of lexical diversity pointed to a consistent increase in the
use of diverse vocabulary as the placement level increases. However, similar to Knoch,
Macqueen & O'Hagan’s (2014) findings, the only statistically significant difference was between
the higher-level writers in Pass and B Levels. The consistency of the findings of the EPT writing
with previous research including research on standardized proficiency tests such as TOEFL
(Knoch, Macqueen & O'Hagan, 2014) indicate that the relationship between lexical complexity
and writing level holds across different testing contexts.
Lexical Sophistication

The results of sophistication indices also provide further evidence that proficient writers tend to use more sophisticated vocabulary. They tend to use more academic words, a greater number of less common words, and more contextually diverse academic bigrams. The analysis of sophisticated vocabulary disclosed a wider range differences between undergraduates writing levels. The results from the first index of sophistication that is AWL words indicated that the Pass undergraduates demonstrate statistically significant knowledge of AWL words than their undergraduate counterparts in C level. The differences in the use of AWL words between the undergraduates in the adjacent levels of C and B were, nevertheless, not as consistent. Level B use of AWL words is shown to be statistically higher than that in C Level. The lack of statistical significance between Pass and B levels on one hand and the discrepancy between C and B level in using AWL words from the other hand, might be attributed to the tendency of repeating words in B Level. This assumption is supported by the qualitative analysis of sample texts from both C and B (five texts from each). The close inspection of how AWL words are used suggests that both C and B level tend to have low repertoire of AWL and that the higher AWL score in B level might not be attributed to their knowledge of AWL per se but to the excessive repetition of a small range of AWL words. Also, the use of AWL words in level B is often inappropriate reflecting a lack of vocabulary knowledge. The following excerpt is from Level B with a relatively high AWL score of .13. The analysis of AWL words revealed repetition of words such as devices (11 times), challenges and technology. It also illustrates lack of an appropriate use of AWL words (AWL words that are inappropriately used either semantically or grammatically are bold and underlined).

(11) society are **conflicts** with technology abuse has already for a long time. With great developing on smart phone and convenience devices, most of the universities are
facing into **challenges** by controlling and **benefiting** electronic devices. Such as iPad, iPad could be both **benefits** and disadvantages. Talk about **benefits** on smart **devices**. ..still able to take all the study required as a whole into a smart **devices**. ..Technical **devices** have definitely solved a lot of study problems. however it also brings more **challenge** as save time.... education providers are facing **challenge** within smart **devices** usages.....Higher cost and hot popularity makes the rest of someone who are not able to purchase the smart **devices** fall into a situation by disgroup with others. In the education provider side. cheating through smart **devices**, and copying .... are much more often due to the developing on smart **devices**. ....smart **devices** does **achieved** a lot of **beneficial assignment** for human’s living, and make the nations closer in general thinking. However, the smart **devices** are making human to become more lazy ....I do **achieve** the great scientist Einstein had said many years before... (B Level, AWL score of .13)

To the contrary, the use of AWL words in C level is featured by less repetition and more adequate use of words. With the exception of the word “text”, no redundancy in using AWL can be noticed (AWL words are in bold) in C level writing with relatively high AWL score of .12:

(12) **The acceleration of high-tech develop contributes** to a variety of **method** for student to learn in **texts**, author show that iPad can play an **positive** role among study of student. In **contrast**, **text** 2 author think that it is not comfortable for student use iPad to learn In my opinion, The idea of **texts** is reasonable.

To begin with, as we can seen from **texts**. Professor of project management said that most **percentage** of student show great interest in using iPad to learn their **academic** course and even other course. additionally, student prefer to use iPad
that print the articles. It is more environmental friendly what's more, using iPad can release the weight of their backpack, which is more benefit for the health of student. In deed, there are several drawbeck of using iPad in school. For instance, a large amount of games and movie in iPad which can be highly addicted by student. It can damage student mentally and physically…. It is difficult to promote the usage of iPad among student last but definitely not least, it may damage the eyesight of student who spend so much time on the computer screen. Overall, I think the possible benefits of iPad maybe outweight the potential risk of it. (C Level, AWL score of .12)

The use of AWL suggests that that the factor of graduate vs. undergraduate seems to play a role in learners’ use of academic words. The graduate level learners in both Pass and D levels demonstrated more knowledge of AWL words from their undergraduate counterparts in C and B levels. Examples 13-14 illustrate differences in the use of AWL (AWL words are in bold) between the Pass (Graduates) and B Level. The Pass Level, as illustrated by Example 13 has a high score = 1.2. whereas Example 14 has a low AWL score of 0.01.

(13) Firstly, with machines doing those simple jobs, we can focus on more difficult jobs like designing, researching, and other jobs that require a high level of intelligence. (PG, AWL score of 1.2)

(14) Robots can replace people from some special situation for sure. they can be rebuild or repair, but they dont have one thing but humans do - emotion. Robots maybe smart, but they only obey the orders that people give them. they wont their
own mind. For this point, something only humans can do. (Level B, AWL score of .01)

The effect of graduate status on the use of sophisticated vocabulary can also be noticed in the findings from the range of academic words. Even though the results revealed no statistically significant differences among the groups with graduate or undergraduate status, the range of academic words used by graduate Pass was statistically lower than that of undergraduates in C and B Levels indicating that the use of less common words is indicative of high-level writing. Excerpts 15 and 16 are from pass (undergraduates) and C level with academic word range scores of .55 and .65 respectively. The examples illustrate how proficient writers demonstrate their vocabulary knowledge through using expressions from less frequent academic words. In addition to academic words from AWL (words in bold are words from AWL sub-list 2 and above), the excerpt from Pass Level includes academic words such as advancement, enriched, pros, cons, overshadow with relatively low range score (averaging .02).

(15) the advancement of the technologies, the humans lives have become much more easier and faster. Use of computers, machines, electronic, have enriched human life. Talking about the pros & cons of these technologies, pros always overshadow the cons, hence technologies are always welcomed by human

(16) iPad is easy to carry. You can bring it to anywhere you will to go. And because of the light weight. It can avoid your trouble to bring lots of heavy books from a building to another. As Woyke mentioned it can help a lot of you needn't print out all of the information. You can easily to study with this
technology's wonderful products. In the other hand, we can easily find the net work. (C Level)

The lack of statistical differences between undergraduate levels in using low range words partially support Kyle & McNamara (2016) findings that the use of low range words is not a strong predictor of holistic scores in integrated tasks. However, the variations between graduate and undergraduate levels in using low range words are in line with the previous literature’s conclusions that the use of words that are more restricted in their use correlate with higher holistic scores on writing tests (Guo, Crossley & McNamara, 2013; Knoch, Macqueen & O'Hagan, 2014; Kyle & Crossley 2015).

The results from the bigram index lend support to the conclusions of previous research regarding the positive relationship between collocational knowledge and writing quality (Kyle & Crossley, 2015). The findings also disclosed more variations among groups with undergraduate status. The writing of undergraduates in Pass displayed statistically better knowledge of multiword-collocations than B Level. In addition, the writers in C level used statistically more common academic bigrams than their B Level counterparts.

In addition, similar to the findings from the AWL and range indices, the use of academically more common bigrams is statistically higher in graduate Pass than their undergraduate counterparts in C and B levels. The list of bigrams in Table 3.7. illustrates the level of lexical knowledge in terms of bigrams between Pass (undergraduates; bigram range score of .21) and B Level (bigram range score of .07). The bigrams used by the Pass level student display knowledge of the structure of noun phrases (known fact, the fact, fact that), adjectival phrases (e.g., social ethical) and phrasal verbs (bring down). In contrast, the bigrams in Level B reflect features common in spoken rather than academic style, e.g., using present
participle tense (e.g., *doing, going*), common nouns (e.g., *things, thing*) and personal pronouns, mainly first singular pronoun *I* (Biber et al., 1999). The use of bigrams in B level writing is also featured by grammatical errors (e.g., *i going*) and atypical or non-sensical clusters (e.g., *machines recess, college bus*).

Table 3.7 *Examples of Bigrams from Low Level B and Pass (undergraduates)*

<table>
<thead>
<tr>
<th>P (undergraduate)</th>
<th>B Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>robots replacing</td>
<td>keep doing</td>
</tr>
<tr>
<td>social ethical</td>
<td>things in</td>
</tr>
<tr>
<td>human jobs</td>
<td>bad thing</td>
</tr>
<tr>
<td>known fact</td>
<td>because there</td>
</tr>
<tr>
<td>large extent</td>
<td>college bus</td>
</tr>
<tr>
<td>various sources</td>
<td>do things</td>
</tr>
<tr>
<td>vast amount</td>
<td>doing it</td>
</tr>
<tr>
<td>the fact</td>
<td>draw the</td>
</tr>
<tr>
<td>bring down</td>
<td>i going</td>
</tr>
<tr>
<td>a known</td>
<td>i learned</td>
</tr>
<tr>
<td>a lot</td>
<td>i love</td>
</tr>
<tr>
<td>in fact</td>
<td>i was</td>
</tr>
<tr>
<td>in the</td>
<td>i would</td>
</tr>
<tr>
<td>big disadvantage</td>
<td>machines recess</td>
</tr>
</tbody>
</table>

Overall, the findings of lexical complexity in terms of sophisticated vocabulary revealed that high level writers use more academic words, and higher number of contextually restricted words. They also display better knowledge of academic word combinations. The results from the analysis of lexical sophistication namely the use of AWL words suggest that graduate students have better knowledge of academic words than their undergraduate peers. It also suggests that the use of AWL words particularly in adjacent levels of undergraduates may not be indicative enough of the writing level in reading based argumentative essays. Accordingly, a more comprehensive understanding of lexical sophistication across placement levels of graduates and undergraduate students can better be achieved by considering the range of AWL and bigrams along with measures of AWL word knowledge.
**Lexical Cohesion**

Lexical cohesion can be constructed through the use of various reiteration strategies including the use of hypernymy. The hypernymy score, as discussed earlier, refers to the mean number of subordinate terms a word has in a text. High hypernymy scores indicate a large number of superordinate terms (Kyle & Crossley, 2016). The results from this study show that the use of nouns with hierarchical semantic relationships is a feature of high-level writing. The use of hypernym nouns was found to statistically distinguish between undergraduate Pass and C levels from the lower B Level. In addition, the index of hypernymy differentiated between graduates and undergraduate levels. The graduate Pass and D level demonstrated statistically significant higher means of hypernym nouns than B level.

The use of more superordinate terms, as shown in Example 17 from the Pass Level (hypernymy score = 7.7) reflects how the clarity of the nouns is promoted by examples and further elaboration. On the other hand, and as illustrated by Example 18 from Level B (hypernymy score = 6.5), the lack of hierarchical semantic relationships in low levels, particularly in Level B, contributes to the ambiguity of the nouns used, which in turn affects readers’ ability to follow the writer’s argument.

(17) *On the other hand when we gaze at the jobs which include human interactions and maintaining the human relations, such as psychological consultancy or nursing and patient care, we see that the robots are never an option. Similarly, the tasks such as painting, sculpting, drawing, craftsmanship, dance, singing which need uniqueness in every artists, cannot be done by robots. (PG)*

(18) *Robots are the new technology in these 30 years. It help people with a lot of things. But it also let people start losing their jobs. Some people doesn't agree*
with using robots but for me Roberts giving a lot of benefit to human beings. (B Level)

Overall and compared to indices targeting the use of AWL and lexical variation, the measure of hypernym nouns tends to provide a more transparent view of differences in lexical complexity among ESL writers in integrated tasks. Targeting vocabulary of specific semantic functions render them less influenced by the characteristics of integrated writing tasks. The variations among placement levels in using hypernymy nouns support previous research conclusions on the positive correlation between hypernymy noun scores and holistic scores in writing particularly in reading based argumentative essays (Guo, Crossley & McNamara, 2013; Kyle & Crossley, 2016; Meisuo, 2000).

Conclusion

The results of lexical complexity in the current study partially support the explanation inference in the validity argument that the linguistic complexity of writing, namely lexical complexity, will vary across placement levels. First, no statistically significant differences were found between graduate students in their ability of using complex vocabulary. Second, while the undergraduates in Pass Level displayed a greater use of diverse and sophisticated words and a better collocational knowledge than their undergraduate peers, B Level has outperformed its undergraduate peers in C Level in using AWL words. In general, lexical complexity features seem to be a distinguishing factor among placement levels of undergraduate status rather than graduate levels. The use of complex vocabulary is also shown to discriminate between graduate vs. undergraduate level students. Overall, indices targeting multi-word collocations and hypernymy nouns tend to reveal a more informative picture of variations between placement levels of undergraduates than the indices of AWL and range of academic words.
Implications for Pedagogy and Future Research

The findings on the whole support the argument for the significance of productive vocabulary knowledge in academic writing. Lexical complexity found to account for 7.7% of the variations in placement levels. Although the effect size is considered small from a statistical perspective, it is still remarkable given the various factors that affect raters’ decisions. Even though the results regarding sophisticated vocabulary use, specifically involving AWL use in lower level writing were in contrast with theoretical expectation, the small-scale qualitative analysis of AWL use in B and C levels align with the EPT rubric description of B Level. Students in B level and as delineated by EPT rubrics demonstrate redundant and repetitive use of small range of vocabulary. In addition, the results from graduate levels of Pass and D indicate that both levels portray similar ability of using complex vocabulary. Accordingly, the study recommends that the EPT rubrics point out differences in lexical knowledge between undergraduate and graduate levels particularly between C/D levels in their use academic vocabulary. In addition, the results from hypernymy noun scores and the analysis of excerpts with high and low hypernymy score provide support for the EPT descriptors of idea development in low and high-level writing. The results from this study has also some implications for ESL writing courses. Considering the importance of lexical complexity in discriminating high and low-level writers and the relative low repertoire of academic vocabulary in B and C levels, ESL writing materials are recommended to emphasize the development of lexical knowledge among undergraduate placement levels.

The limitations and findings of this study suggest several avenues for future research. First, lexical complexity was examined with only two adjacent levels of graduate students. Future research in the EPT context is recommended to investigate differences between Pass, D and B levels. Second, the qualitative differences found between adjacent lower level of writing
i.e. C and B in using AWL words were based on a small number of texts, researchers are invited to further explore differences between adjacent low-level writers in the use of AWL in reading-based tasks. In addition, the discrepancy between word range findings in this dissertation and that of Kyle & McNamara (2016) suggests that the index of word range is more appropriate for discriminating between heterogeneous groups of different writing levels, e.g., graduate vs. undergraduates.

Accordingly, future research particularly in the EPT context is invited to further examine how the use of low range words distinguish different writing levels within and across graduate vs. undergraduate ESL learners in both integrated argumentative and summary tasks. Overall, the variations reported in this study among placement levels are relevant to reading based argumentative essays. It would be more informative to examine if the differences hold also in integrated summary and in independent tasks.
CHAPTER 4. SYNTACTIC COMPLEXITY IN ESL WRITING

Syntactic complexity as a feature of development in writing ability has drawn a great deal of interest in both L1 and L2 research. It is defined as “the range of forms that surface in language production and the degree of sophistication of such forms” (Ortega, 2003, p. 492). Different measures have been used to gauge syntactic complexity. These measures can be categorized into general and more specific (Vyatkina, 2013). General measures focus on the average length of the produced syntactic unit at the clause or t-unit which is defined by Hunt (1970) as “one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it” (p.4). Commonly used measures of length (averaged by words) include mean length of clause, mean length of t-unit, clauses per T-unit (Boulte` and Housen, 2014). More specific measures calculate the ratio of specific structures in relation to another larger syntactic unit, e.g., coordinate structures per clause, complex nominals per clause, complex nominals per T-unit (Lu, 2011).

The assumption is that the larger the measurement unit is, e.g., the mean length of the T-unit or the clauses per T-unit, the more complex language it reflects. However, the problem with these measures is that they lack specificity in defining syntactic complexity and its relatedness to the context or register (spoken vs. written) in which the langue is produced. This lack of definition contributed to the inconsistency of findings regarding the relationship between syntactic complexity and text quality. For example, Beers and Nagy (2009) found that measures of word per T-unit and words per clause produce contradictory conclusions on the relatedness between syntactic complexity and writing level in narratives vs. argumentative essays. However, there is no explanation of what led to the discrepancy in terms of specific syntactic structures. In her synthesis analysis of six commonly used T-unit based syntactic measures (measures tapping
the length, subordination and coordination) across twenty-five L2 research, Ortega (2003) found that the results from these “global” measures vary considerably depending on the instructional setting ESL vs. EFL and how proficiency level is defined (e.g., based on program level or holistic scores). Global measures of syntactic complexity were also criticized for masking the actual features contributing to complex structures and for failing to reflect the structures characteristic of written vs. spoken discourse.

T-unit and clausal subordination measures have come under criticism, particularly from corpus-based research on register variation (Biber & Gray, 2010; Biber, Gray & Poonpon, 2011; Biber, Gray & Staples, 2016). Biber and his colleagues have characterized T-unit based measures as ‘holistic’ because they fail to unveil the specific features contributing to complexity, instead collapsing many different grammatical constructions into a single measure (e.g., clauses per T-unit, or mean length of T-unit, in which no distinction is made in terms of what types of clauses are used or what grammatical structures result in a longer T-unit). In addition, Biber and colleagues argue that the definitions used to operationalize T-unit based measures fail to account for grammatical structures that are most characteristic of the nominal style of academic writing. For example, various linguistic structures can be used in extending T-unit length (e.g., phrasal complexity or clausal embedding). Phrasal complexity was found to be characteristic of academic written discourse, whereas clausal embedding is more prevalent in the spoken mode (Biber, Johansson, Leech, Conrad & Finegan, 1999). The examples below (taken from Biber, Gray & Poonpon, 2011, p. 14) illustrate this argument, as both sentences contain the same T-unit length (20 words).

(1)  
Well, [since he got so upset, I just did not think [we would want [to wait for Tina [to come back]]]]
This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes].

Biber, Gray & Poonpon (2011) argued that based on the MLTU or the mean length of the T-unit measure, both sentences above are considered complex as they both have equal number of words. In fact, if complexity is defined in term of the number of dependent clauses per T-unit, the first sentence, obtained from a natural conversation, would be considered more structurally complex than the second (in written form from a text book). The first sentence has four dependent clauses whereas the second has no dependent clauses. Both measures fail to unveil that complexity in the excerpt obtained from conversation stems mainly from the dependent clauses whereas syntactic complexity in the written form constructed mainly through noun modifiers particularly adjectives and prepositional phrases.

Biber, Gray & Poonpon (2011) concluded, hence, that use of T-unit measures provide not only conflicting but also misleading conclusions about the construct of complexity in written vs. spoken registers. Accordingly, some recent researchers resorted to use more specific T-unit related measures that tap into noun phrase complexity e.g., mean length of noun phrase, ratio of complex nominals per clause and per T-unit (Lu, 2011, Kyle and Crossley, 2018) number of modifiers per noun phrase (Crossley and McNamara, 2014). The findings of these fine-grained measures support corpus-based research on the relatedness of noun phrase complexity to higher writing levels. Nevertheless, these more specific measures of noun phrase complexity are still “holistic” in the sense that they fail to pinpoint the actual linguistic features contributing to noun phrase complexity. Therefore, the analysis of syntactic complexity in terms of the specific linguistic features posited by corpus-based research as characteristic of written register provides
a more appropriate and transparent picture of the relationship between syntactic complexity and writing development (Biber, Gray & Poonpon, 2011).

Corpus based research on register variations revealed substantial differences between spoken and written registers. Complexity in written register constructed mainly through complex nominals whereas spoken language relies heavily on dependent clauses (Biber, 1988; Biber et al., 1999). These findings as will be discussed in the next section have triggered a wave of research on syntactic complexity based on differences between spoken and written registers, with the premise that researchers interested in writing development should use measures that (1) are capable of capturing the grammatical features which result in complexity in writing, and (2) that account for the particular communicative functions of the range of grammatical structures that result in syntactically complex discourse. Some studies have used L1 texts as a reference corpus (Biber & Gray, 2010; Biber, Gray & Poonpon, 2011) or compared spoken versus written discourse in ESL context (Biber, Gray & Staples, 2016). Other researchers (e.g., Parkinson & Musgrave, 2014; Taguchi, Crawford & Wetzel, 2013) examined corpus research findings across L2 writing levels.

Given the research evident differences between spoken and written registers and the failure of T-unit based measures in capturing features characteristics of written register and in spelling out the specific linguistic structures contributing to syntactic complexity, corpus-based register approach will be utilized in this study to investigate syntactic complexity features in the EPT written responses. Accordingly, the literature review in the next section will focus on corpus- based research findings of differences between spoken and written registers. This will be followed by a review of the findings of corpus- based studies in which complexity was examined based on linguistic features common in academic written discourse. Building on the findings of
corpus-based studies, a description of my measures of complexity and the procedures of statistical analysis will then be discussed. The chapter concludes with a discussion of the findings and implications for future research.

**Literature Review**

**Corpus-based Studies on Written vs. Spoken Registers’ Variation**

Corpus-based research on the differences between spoken and written registers has pointed to the prevalence of noun phrase complexity in written discourse. Multi-dimensional analyses of large spoken versus written corpus conducted by Biber (1985, 1986, 1988) revealed that the use of dependent clauses is common in spoken language, whereas noun phrase modifications are typical of academic written discourse. Similar findings were reported by relevant subsequent studies. Comparing 10 major grammatical structures in research articles (Science/ Medicine, Education, Psychology and History) versus American English conversation, Biber and Gray (2010) confirmed Biber’s (1985, 1986) and Biber et al.’s (1999) conclusions that written discourse complexity is reflected by noun phrase complexity, whereas conversation complexity is manifested by clausal subordination. Adverbial clauses (e.g., *because, if* clauses), complement clauses (finite and nonfinite) found to be more frequent in conversation than in academic writing. Noun phrase modifiers (finite and non-finite relative clauses, prepositional phrases, attributive adjectives, pre-modifying nouns) were typical of academic discourse.

In a similar study conducted by Biber, Gray and Poonpon (2011), twenty-eight grammatical features were investigated. The distribution of these features shows that they can be analyzed in terms of the structural (finite dependent clauses versus dependent phrases) and syntactic functions (constituents of a clause versus constituents of a phrase) favored in spoken versus written communication. In other words, spoken discourse relies heavily on clausal subordination, whereas written discourse relies on phrasal features. ANOVA tests with the
features as dependent variables and register as an independent variable were used. Considering both structure and syntactic functions, it was found that some features were strongly related to conversational style (e.g., finite adverbials, finite complement clauses of verbs), others were typical of academic writing (non-finite adverbial clauses), and a third category (e.g., finite and non-finite relative clauses) were less frequent in both registers and, hence, register differences are less categorical compared to other features. At the clause level, the following features were particularly favored by written discourse: non-finite adverbial clause, non-finite complement clauses of adjectives and nouns (adjectives + to clause, adjectives +ing clause, noun+ing clause, noun + to clause) and non-finite relative clauses. At the phrasal level (constituents of a noun phrase), the results were consistent with previous corpus research findings in that prepositional phrases, attributive adjectives, and nouns as prenominal modifiers were statistically favored by the written discourse. Based on the variations found between spoken and written registers, Biber, Gray, and Poonpon (2011) proposed hypothesized developmental stages of writing (rather than a single model that disregards register) that move from the use of finite clauses acting as constituents in a clause to an intermediate stage of using non-finite clauses as constituents in a clause to finally using dependent phrases as constituents of a noun phrase.

Corpus -Based Studies on Syntactic Complexity

The prevalence of noun phrase complexity in written discourse, as opposed to oral conversation, triggered further investigation of how these differences predict differences in ESL proficiency levels or placement levels of writing. Biber, Gray, and Staples (2016) examined the distribution of 23 grammatical features in the TOEFL iBT spoken vs written tasks (integrated versus independent). Consistent with written/spoken register variation results (Biber, Gray & Poonpon, 2011), these authors found that non-finite dependent clauses modifying nouns (noun + ed- relative clause and noun + to- clause) are more common in written integrated tasks, whereas
adverbial clauses, adverbs and verb + to-clauses are more frequent in speech and independent tasks. General linear models with scoring levels, tasks, modes, and test takers as independent variables were used. The post hoc tests for statistically significant features revealed that adverbs, finite adverbials, are common in speech and in lower rated integrated tasks. Nominalization and several structures embedded in noun phrase (attributive adjectives, pronoun modifiers, of phrases, that–clauses as noun complements, to-clauses as noun complements, and ed-relative clauses) were associated with writing and integrated tasks. All these features were found to be statistically significant predictors of task and mode. However, apart from attributive adjectives and that clauses as verb complements, none of these features were found to be a statistically significant predictor of score level.

A factor analysis was then used to detect potential differences based on the co-occurrence of these linguistic features along with some other features. Four factors were identified: literate versus oral, text versus personal experience, abstract versus concrete description, and the fourth factor was personal narration. Mode and task were found to be statistically significant predictors of the four dimensions. Only the dimension of literate/oral was a statistically significant predictor of scoring levels. The positive features in the literate/oral dimension are mainly dependent phrases, such as nouns, and phrasal noun modifiers (of phrase, adjectives), as opposed to clausal features, such as finite adverbials and that clauses controlled by likelihood verbs. The literate/oral dimension was found to have a stronger correlation with the score level in independent writing tasks ($r^2 = 15.2$) than the integrated ones ($r^2 = 1.7$). Biber, Gray, and Staples’ (2016) conclusions corroborate the general findings regarding the differences between spoken and written registers. Phrasal complexity is associated with written discourse, whereas features of clausal subordination are characteristic of spoken discourse. With regard to task type, the
study highlights that the distribution of complexity features is influenced by the task type (integrated versus independent). Integrated tasks stimulate the use of more features of grammatical complexity than independent tasks. Another notable finding is that even though some features, namely attributive adjectives, verb + that clauses are more common in spoken language, they are associated with high rated essays.

In an attempt to examine the relationship between placement levels and features of clausal and phrasal complexity, Taguchi, Crawford and Wetzel (2013) compared high (n = 30) and low rated (n = 24) essays of ESL college students. The placement test was an online integrated writing task in which students were asked to read two texts presenting contrastive views and write an essay comparing and arguing for one view. Building on corpus research findings and using Biber’s automatic grammatical POS tagger, they compared clausal-level complexity measures (subordinate conjunctions, that verb complement, that noun complement, that adjective complement, and relative clauses of nouns) as well as phrasal-level pre- and post-noun modifiers. Similar to findings regarding spoken versus written discourse in Biber (1985, 1986), Biber and Gray (2010) and Biber et al. (1999), Taguchi, Crawford and Wetzel (2013) found that, with the exception of the that clause verb complement, clausal measures mainly as subordinating conjunctions, adjective complements (that complement) and that relative clauses are higher in low rated essays. As for the phrasal level, it was found that the frequency of attributive adjectives and post-noun modifying prepositional phrases was higher in high rated essays compared to low rated ones. The findings of Taguchi, Crawford, and Wetzel conform with the assumption of corpus-based research that finite subordinate clauses that are typical of spoken language are more common in low rated essays, whereas phrasal complexity is typical of written discourse and, therefore, is more frequent in high rated essays.
Building on a subset of the hypothesized developmental progression proposed by Biber, Gray, and Poonpon (2011), Parkinson and Musgrave (2014) examined features of noun phrase complexity across different ESL writing levels. The participants were ESL students studying English for academic purposes (EAP) and MA students enrolled in a TESOL program. The written samples were obtained from students’ course assignments. The 21 EAP samples consisted of untimed argumentative essays (independent task), whereas the 16 MA written samples were part of the students’ course assignments (extended written responses to one question). In support of Biber et al.’s (2011) suggestion that attributive adjectives are acquired at an earlier stage of writing development, it was found that the proportion of attributive adjectives in EAP writing was significantly higher than that in MA writing. The pre-modifying nouns and total prepositional phrases with abstract meaning, suggested by Biber et al. (2011) to be acquired later, were significantly higher in MA writing. The results regarding using attributive adjectives in high writing levels run contrary to Taguchi, Crawford, and Wetzel’s (2013) findings that attributive adjectives are more common in low rated essays. This discrepancy, in the author’s view, is due to the differences in task types (timed versus untimed). With untimed tasks, learners are required to read more academic research, which is reflected in their writing, whereas with timed tasks, they depend mainly on easily accessible features, particularly that of attributive adjectives.

This review of the corpus-based studies on the distribution of syntactic complexity features across different writing levels (Biber, Gray & Staples, 2016; Parkinson and Musgrave, 2014; Taguchi, Crawford & Wetzel, 2013) suggests that certain linguistic features tend to be more common in higher writing level essays, particularly in terms of nouns, nominalization, pre-noun modifiers and prepositional phrases modifiers of nouns. The above review also suggests
that when considering differences in syntactic complexity, it is more informative and interpretable to examine linguistic features in terms of both structure and syntactic functions. For example, both relative clauses and finite adverbials are structurally dependent finite clauses, yet with syntactically different functions, the former is a constituent of a noun phrase, whereas the latter is a constituent of a clause. Such a distinction allows for interpreting the distribution of linguistic features across different registers and, consequently, across different proficiency levels.

Taking the task type into consideration when reviewing syntactic complexity findings, it seems that integrated tasks tend to elicit more nominal complexity. Hence, it would be insightful to examine how syntactic complexity features, at both phrasal and clausal levels, are reflected in ESL integrated writing tasks. Keeping in mind the importance of syntactic complexity as a major component of the rating rubrics in the EPT writing test, it is necessary to recognize how features of syntactic complexity discriminates groups of different writing levels and whether source-based writing tasks influence the extent of variation among the groups. The current research considers syntactic complexity in terms of structure and syntactic functions. In other words, a syntactic structure like finite clauses vs. nonfinite clauses will be examined in terms of their syntactic function, i.e. clausal constituents as opposed to phrasal constituents. The assumption is that higher level texts will show more resemblance to written discourse in using more nonfinite structures and more reliance on compressed noun phrases (Biber & Gray, 2010; Biber et al., 2011; Taguchi et al. 2013), whereas lower level texts will tend to include more dependent clauses particularly finite adverbial structures.
Purpose of the Study and Research Questions

Syntactic complexity has been posited as a main component of linguistic complexity and hence, has been used as an anchor for discriminating students of different writing levels. Therefore, this study will investigate how features of syntactic complexity vary across placement levels of writing in EPT. In light of the findings of corpus-based studies on differences in features of syntactic complexity between spoken and written language and the findings in L2 corpus-based research, the following question is investigated: To what extent do ESL writers of different placement levels exhibit differences in syntactic complexity features?

Based on the literature review on syntactic complexity in spoken vs. written and in ESL timed writing, it is assumed that features characteristics of written discourse and of high rated integrated essays will be more frequent in higher level writing. Specifically, it is expected that higher placement levels will utilize 1) more frequent nominal modifiers both phrasal and clausal 2) less finite subordinate adverbials and more frequency of non-finite adverbials and 3) more frequent use of That verb complement.

Methodology

The Corpus and the Linguistic Features Examined

The corpus used in the current analysis of syntactic complexity is the tagged version of the same sub-sample of the EPT corpus used in the lexical analysis in Chapter 3 (554 texts; see section 3.2.2). The texts were tagged using the Biber tagger (Biber, 1988), which is the same Part of Speech (POS) tagger employed in several corpus-based studies of grammatical complexity (Biber & Gray, 2010; Biber, Gray & Poonpon, 2011). A program to identify and count instances of the complexity features developed for Biber and Gray (2010) was used to calculate the raw and normalized counts of specific syntactic features in each text. Only normalized frequency (per 1,000 words) is used in the current paper. Unlike raw counts, normalized frequency accounts for
differences in text length and hence, insures accurate comparisons of the features’ counts across texts of various length (Biber, Conrad and Reppen, 1998). Normalizing can be accomplished by dividing the raw counts of the target feature by the length of the text and then multiply it by the norming number (1000 in this study) for which selection, e.g., 100 or 1000 is based on the “typical” length of the corpus texts (Biber, Conrad and Reppen, 1998, p. 264). The selection of linguistic features for the analysis of syntactic complexity takes into consideration the findings of differences between written and spoken registers (Biber and Gray, 2010; Biber, Gray and Poonpon, 2011) and the syntactic features associated with timed integrated argumentative essays in ESL writing (Biber, Gray and Staples, 20016; Taguchi, Crawford and Wetzel, 2013). For example, based on spoken vs. written variations, finite adverbials are more common in spoken whereas non- finite adverbials, nominalization, nouns and noun phrase modifiers including phrasal (e.g., adjectives, nouns, prepositional phrases) and clausal are typical of writing. In addition, features like That complement of verbs were found to be frequent in high rated integrated written essays. The targeted features are summarized by Table 3.1 with illustrative examples from the EPT corpus.

The five placement levels Pass and D for graduate students; Pass, C, and B for undergraduates are compared in terms of normalized rates of occurrence for the 12 targeted syntactic features. The comparisons among groups will be based on descriptive statistics of the mean and standard deviation. The use of inferential statistics namely MANOVA was not possible with the current data because 1) the majority of variables did not meet the normality assumption across all placement levels as was evident from histograms and Shapiro_Wilk test (p < .05) and 2) some of these variables have very low or almost no correlation ($r^2 = 0.00$) rendering them inappropriate to use in MANOVA. The results of quantitative analysis will
further be explicated through further analysis of how the targeted features are used across levels.

Concordance lines of AntConc- a freeware corpus analysis toolkit developed by Anthony, 2014- will be used to examine the general patterns of use for prominent linguistic features.

Table 4.1 Syntactic Features that are Included in the Analysis of Syntactic Complexity

<table>
<thead>
<tr>
<th>Grammatical features</th>
<th>Syntactic functions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dependent phrases + nominalization and nouns. (Note: Nouns are in bold)</td>
<td>1A. Adjectives (pre-noun modifiers)</td>
<td>1A. These small and light weight devices has some benefits (D Level)</td>
</tr>
<tr>
<td></td>
<td>1B. Prepositional phrases (of prepositional phrase) as post-noun modifier</td>
<td>1B. Using them in school classes to replace textbook and to increase the quality of education has some benefit and disadvantages that we are going to talk about (D Level)</td>
</tr>
<tr>
<td></td>
<td>1C. Nouns as nominal pre-modifiers</td>
<td>1C. For this reason, I think iPads in school classrooms will be helpful for their challenges. (C Level)</td>
</tr>
<tr>
<td></td>
<td>1D. Nominalization</td>
<td>1D. exposure of technology to students is definitely significant and beneficial. (PG)</td>
</tr>
<tr>
<td>2. Dependent Clauses (noun phrase constituents)</td>
<td>2A. Non-finite relative clauses (ed-clauses+ ing clauses+ to-clauses)</td>
<td>2A. Ing-noun modifier: They are so efficient and easy to carry that almost every student attending a college has one today. (PU)</td>
</tr>
<tr>
<td></td>
<td>2B. Relative clauses (that+wh)</td>
<td>2B. Ed-noun modifier: From the information provided in the two texts, I strongly believe that the advantages of using iPads in schools certainly outweigh the challenges (D Level)</td>
</tr>
<tr>
<td></td>
<td>2C. That complement clauses of nouns</td>
<td>2B. The experience that we have from our iPad is completely different from the usual way we were used to learn in school. (PG)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2C. Though the second text points out the fact that the iPad is expensive for students, I’m sure the fast-growing technology is going to produce very efficient and low cost tablets exclusively for students, in the days to come. (PG)</td>
</tr>
<tr>
<td>3. Dependent clauses (clause constituents)</td>
<td>3A. Finite Adverbial clauses</td>
<td>2A. If you do not have paper, you do not need to be worried about the final destination. (PG)</td>
</tr>
<tr>
<td></td>
<td>3B. That complement clause of verb</td>
<td>2D. I believe that the decision-making situation should always done by human (B)</td>
</tr>
<tr>
<td></td>
<td>3C. Non- Finite Adverbials</td>
<td>3D. To support my argument, I will use arguments given in the second article by Erica Loop, and some of my own (PG)</td>
</tr>
<tr>
<td></td>
<td>3D. Non- finite complement clauses of verbs</td>
<td>3E. I want to focus on the only one victory by a human player (B Level)</td>
</tr>
</tbody>
</table>
Results and Discussion

The results of descriptive statistics will be presented under two main sections: non clausal and clausal features. This organization meant to highlight group variations in the use of syntactic structures (clausal vs. non clausal) and the differences in the use of clauses with different syntactic roles (clauses constituents of a clause vs. clauses constituents of a noun phrase). Each section commences by comparing the distribution of the respective syntactic features within the placement levels of 1) graduates 2) undergraduate and finally 3) comparing graduates to undergraduate level students. To further understand how features of noun phrase complexity in particular contributed to writing quality across levels, and how argumentative genre promotes use of certain structural structures, examples from students’ writing will be inspected as well.

Non-Clausal Features

The distribution of nominal modifiers, nouns and nominalization across placement levels is presented in Table 4.2. To facilitate the visual representation of the distribution of the constituents of the noun phrase, Figure 4.1. illustrates the distribution of noun phrase complexity features across the five levels. At the graduate level, the descriptive statistics show that with the exception of nominalization and prepositional phrases where graduate Pass seems to slightly outperform their fellows in D level, negligible differences are noticed between the graduate students in using nominal modifiers specifically nouns, adjectives and nouns as prenominal modifiers.
Table 4.2 Mean Rates of Occurrence for Phrasal Complexity Features in EPT Writing (Normalized to 1,000 Words)

<table>
<thead>
<tr>
<th>Linguistic Features</th>
<th>B Mean (SD)</th>
<th>C Mean (SD)</th>
<th>D Mean (SD)</th>
<th>P (UG) Mean (SD)</th>
<th>P (G) Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectives</td>
<td>3.838 (2.166)</td>
<td>3.015 (1.412)</td>
<td>4.793 (2.061)</td>
<td>3.728 (1.707)</td>
<td>4.86 (1.775)</td>
</tr>
<tr>
<td>Pre-nouns</td>
<td>2.048 (1.332)</td>
<td>2.05 (1.197)</td>
<td>1.927 (1.098)</td>
<td>1.976 (1.156)</td>
<td>1.995 (1.175)</td>
</tr>
<tr>
<td>Of phrase</td>
<td>2.096 (1.246)</td>
<td>2.359 (1.318)</td>
<td>2.655 (1.217)</td>
<td>2.585 (1.385)</td>
<td>2.834 (1.063)</td>
</tr>
<tr>
<td>Normalization</td>
<td>3.283 (1.799)</td>
<td>3.593 (1.562)</td>
<td>4.096 (1.735)</td>
<td>3.937 (1.642)</td>
<td>4.498 (1.926)</td>
</tr>
</tbody>
</table>

Figure 4.1 Mean Rates of Occurrence (Normalized per 1,000 Words) of Phrasal Types Functioning as Constituents of NP across EPT Writing Levels

As for undergraduate levels, somewhat surprisingly, the mean frequency of attributive adjectives in level B is higher than that of their undergraduate peers in C Level (M = 3.01, SD = 1.4). In addition, both B level (M = 3.84, SD = 2.17) and the undergraduate Pass level (M = 3.73, SD = 1.7) exhibit similar frequency use of attributive adjectives.

Similar to graduate levels, a clearer pattern of distribution across placement levels of undergraduates is noticed in the use of prepositional phrases and nominalization with negligible
variation in the use of nouns. The undergraduates show more of a linear increase in the use prepositional phrases (see Figure 4.1) and nominalization as the level of writing advances. Less distinction among placement levels is, however, reflected in the use of nouns as premodifiers - a feature posited to be characteristic of high-level academic writing- and the use of nouns in general. In fact, contrary to expectations, the level B and undergraduates in C show as high means (M = 2.05, SD = 1.2) as Pass level (M = 2., SD = 1. 18) in the frequency use of nouns as a prenominal modifier. Similarly, little distinction is found among undergraduate placement levels in using nouns with low- level B is shown to have almost as high mean of nouns (M = 25.61, SD = 3.49) as their undergraduate peers in Pass Level.

When comparing non clausal complexity features in graduate vs. undergraduate levels, the descriptive statistics indicate that with the exception of nouns and nouns as a prenominal modifier, graduate student (Pass and D) used a higher frequency of phrasal modifiers of nouns (adjectives and prepositional phrases) and nominalization than their undergraduate fellows.

The frequent use of attributive adjectives and nominalization in higher level of writing particularly graduates calls for further inspection of how these features are used across high- and low-level writing. The analysis of concordance lines suggests differences not only in terms of frequency, but also structural and lexical complexity between graduate vs. undergraduate levels. As Example 1 illustrates, the use of attributive adjectives in the Pass level (particularly for graduate students) comes often in compound and coordinated adjectival structures “very efficient and low cost,” compared to simple single adjectival structures in Level B.

(1) I'm sure the fast growing technology is going to produce very efficient and low cost tablets exclusively for students, (Level PG)
(2) *For example, robot can helps people go into small hole, dangerous job that might kill a person. Robot can be the new way of security and safety.* (Level B).

In addition to structural simplicity, the examination of concordance lines suggests that the frequency use of attributive adjectives in B level might be influenced by either the prompt or the reading passage. The examples below indicate that students in B level tend to rely on repetition or borrowing to compensate for their lack of vocabulary knowledge. The examples show how the attributive adjective “unique” is used repeatedly across different texts in Level B.

(3) *First, human have many unique talent and ability.* (B Level)

(4) *The professor said that we all need to focus on which jobs are the best use of human's unique talents and abilities.* (B Level)

(5) *Rainwater said there are job that only can human do. Something like needs human's unique talents and abilities. In my personal opinion, there are some job that needs people's unique talents and abilities, but I think its much less than robots take over our job.* (B Level)

(6) *Humans can be focused fully on which jobs that are best use of humans' unique talents and abilities* (B Level)

The other phrasal feature that was found to increase along with the writing level is nominalization. The use of nominalized structures allows for more compression of information—and can also be used to organize the flow of ideas by making reference to a previous or a forthcoming argument. In example 7, the use of “this successful development and technological and urban progress” allows for presenting the same information in a compressed rather than elaborated and lengthy structures like “to develop abstracts created by human brains successfully into actions and to progress technology in urban areas, we need…”. Furthermore, the
nominalized word groups “development and progress” facilitate the transition to new ideas by functioning as what Francis (1994) refers to as “retrospective labels” i.e. nominal groups that refer to a previous discourse.

Similar function is noticed in in example 8 in which the noun phrase “management and planning” is used to encapsulate and summarize an earlier proposition. Example 9 illustrates how nouns can be used to introduce and evaluate propositions. The noun “possibility” is not only used to introduce an argument, a function that Francis (1994) referred to as “advance label”, but also, to question the validity of the claim that “robots will replace human…”.

(7) The abstracts created by human brains can be put in action by robots. This successful development and technological and urban progress needs both the human and the robot to working harmony (PG Level)

(8) Thus Robots can be utilized in our society with proper planning and assessment so that the whole balance and fulcrum of our society does not get effected. The authorities need to make sure that everyone gets to learn and work whoever gets replaced by a Robot.

Proper Management and planning can help us make the best use of technology without disturbing the life balance. (PG)

(9) Some experts are worried about the future for the possibility that robots will replace human in almost all the careers. However, others held a more optimistic view they regard this phenomenon as a good thing for human beings. (PG)

To sum up the discussion of non-clausal features, the overall findings of normalized frequency support previous research conclusions that the writing of high rated essays display more nominal complexity. Overall the frequency use of attributive adjectives, prepositional phrases and nominalization tend to increase along placement levels. The only exception is the
use of attributive adjectives in B level which surpasses their undergraduate peers in C. The investigation of concordance lines suggests tentatively that the frequency use of adjectives in lower levels might be inflated by the repetition and borrowing from the reading passage or the prompt.

**Clausal Features**

The finite and non-finite clausal structures that were examined can be used in two syntactic roles: clausal features that are constituents of the noun phrase and clausal features that are constituents of a clause. I will first discuss the results of clausal structures constituents of a noun phrase followed by those of a clause. With respect to clausal modifiers of nouns, the graduate students in Pass and D levels display as illustrated by Figure 4.2 similar frequency use of finite and non-finite relatives as well as *That* complement clause of nouns.

### Table 4.3 Mean Rates of Occurrence (Normalized per 1,000 Words) for Clausal Complexity Features in EPT Writing

<table>
<thead>
<tr>
<th>Linguistic Features</th>
<th>B Mean (SD)</th>
<th>C Mean (SD)</th>
<th>D Mean (SD)</th>
<th>P(UG) Mean (SD)</th>
<th>P(G) Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite Relatives</td>
<td>0.824 (.759)</td>
<td>0.78 (.766)</td>
<td>1.061 (.678)</td>
<td>0.938 (.682)</td>
<td>1.05 (.663)</td>
</tr>
<tr>
<td>Non-finite relatives</td>
<td>0.376 (.481)</td>
<td>0.367 (.459)</td>
<td>0.509 (.529)</td>
<td>0.335 (.438)</td>
<td>0.52 (.519)</td>
</tr>
<tr>
<td><em>That</em> Complement Clause of Noun</td>
<td>0.096 (.222)</td>
<td>.085 (.22)</td>
<td>.204 (.355)</td>
<td>.157 (.288)</td>
<td>.155 (.264)</td>
</tr>
<tr>
<td>That verb Complement</td>
<td>0.523 (.596)</td>
<td>0.672 (.641)</td>
<td>0.612 (.568)</td>
<td>0.637 (.633)</td>
<td>0.636 (.66)</td>
</tr>
<tr>
<td>Non-finite Verb complement</td>
<td>0.243 (.361)</td>
<td>0.272 (.397)</td>
<td>0.276 (.316)</td>
<td>0.434 (.511)</td>
<td>0.28 (.366)</td>
</tr>
<tr>
<td>Finite Adverbials</td>
<td>1.017 (.728)</td>
<td>0.862 (.792)</td>
<td>0.733 (.556)</td>
<td>0.819 (.698)</td>
<td>0.791 (.59)</td>
</tr>
<tr>
<td>Non-finite Adverbials</td>
<td>0.061 (.169)</td>
<td>0.038 (.147)</td>
<td>0.095 (.213)</td>
<td>0.077 (.175)</td>
<td>0.069 (.16)</td>
</tr>
</tbody>
</table>

More variations in the use of clausal modifiers of nouns are witnessed among placement levels of undergraduates. The frequency use of finite relatives was higher in Pass level (M = .94, SD = .62) compared to both C (M = .8, SD = .77) and B (M = .82, SD = .76) levels. Similar pattern of frequency is shown in the use of *That* complement of nouns as the Pass level (M =
1.57, SD = .29) outperformed the lower writing levels in C (M = .09, SD = .22) and B (M = .1, SD = .22). No noticeable distinctions, however, are revealed in the use of non-finite relatives.

When comparing the frequency of clausal modifiers of a noun phrase in graduate vs. undergraduate levels, post noun modifiers (finite relatives, noun that-complement clauses, and non-finite clauses after nouns) are shown to be of relatively higher frequency in graduate level compared to lower undergraduate levels namely in B and C.

![Figure 4.2 Mean Rates of Occurrence (Normalized per 1,000 Words) of Finite and Non-Finite Clausal Modifiers of Nouns](image)

The frequency use of post noun modifiers namely That complement of nouns in graduate levels reflects its function in establishing a stance in argumentative language. The structure is commonly used with nouns indicative of stance such as fact, claim, belief, possibility...etc. Example 10 illustrates how that clauses can be used to present a stance and a proposition at the same time. A major difference in using That complements in lower levels, is that the head nouns are of more personal connotations, e.g., opinion, side, and view. Also, unlike higher levels of writers where the subjects of the main clause are usually impersonal (Example 10), the subjects of the main clauses in lower levels of C and B are mainly personal pronouns as demonstrated by examples 11-12. The use of impersonal pronouns is common in academic writing because it allows the writer to distance himself from the truth value of a proposition and at the same time
place more focus on particular elements, e.g., “a well-known fact” (Biber et al. 1999). To the contrary, the frequency use of personal pronouns reflects more personal and conversational style as these features are more frequent in conversation than in academic writing (Biber et al., 1999).

(10) *It is a well known fact that robots can perform tasks with extreme accuracy in a repetitive manner* (PG)

(11) *It will be a tendency in the future life. Therefore, I hold my opinion that its weakness.* (Level C)

(12) *I stay in the side that some of the jobs can be replaced by the robots* (Level B)

As for clausal structures functioning as clause constituents i.e. finite and non-finite adverbial clauses, Figure 4.3, shows similar frequency distribution of both finite and non-finite adverbials in graduate level. As for undergraduates, B Level is shown to have the highest mean (M = 1.017, SD = 0.73) of finite adverbial followed by the undergraduates in C (M = 0.862, SD = 0.79) and Pass respectively (M = 0.819, SD = 0.698). Compared to undergraduates, the finite adverbials were least frequent in both graduate placement levels: Pass (M = 0.791, SD = 0.59) and D (M = 0.73, SD = 0.56). In contrast with finite adverbials, it is hard to detect a clear pattern of the frequency use of nonfinite adverbials as they are relatively rare in the corpus.

![Figure 4.3 Mean Rates of Occurrence (Normalized per 1,000 Words) of Finite and Non-Finite Adverbials](image_url)
The examination of finite adverbials’ frequency discloses, as shown in Table 4.4, that “because” (M = .241, SD = .39) and conditional “if” (M = .397, SD = .49) have the highest mean frequency in the whole corpus. The analysis of concordance lines of how finite adverbials are used in low and high levels suggests that the differences are not only in the frequency, but also in the accuracy and the functions served by subordinators. Writers in B level display apparent problems in their ability to construct sentences with subordinate clauses (examples 13-14). In addition, the inspection of concordance lines point to some variations in the use of true and untrue if conditionals. The use of true or in Biber’s et al. (1999) term “real” if conditionals implies that the condition may or may not be fulfilled, e.g., in a sentence like “if it rains, the temperature will drop”, it is not clear if the condition in the subordinate clause has been met or not. On the other hand, the use of untrue or “hypothetical “if “ indicates that the condition in the subordinate clause has not been met, e.g., “if I had known, I would have studies harder” (Biber’s et al., 1999, Note: examples are mine). The use of if conditionals in low level writing seems to rely more on true conditionals as illustrated by example 15. When untrue conditionals are used, they are, in many instances, associated with inaccuracies in tense form (see example 16). Higher levels, however, tend to show better mastery of using both true and hypothetical conditionals as illustrated by examples 17 and 18 respectively.

Table 4.4 Means of Normalized Frequency Counts of Finite Adverbials

<table>
<thead>
<tr>
<th></th>
<th>Because</th>
<th>Although</th>
<th>Though</th>
<th>Since</th>
<th>If</th>
<th>Unless</th>
<th>When</th>
<th>While</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>.241</td>
<td>.058</td>
<td>.039</td>
<td>.062</td>
<td>.397</td>
<td>.003</td>
<td>.043</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>(.39)</td>
<td>(.18)</td>
<td>(.15)</td>
<td>(.21)</td>
<td>(.49)</td>
<td>(.032)</td>
<td>(.14)</td>
<td>(.11)</td>
</tr>
</tbody>
</table>
Though they have the possibility to influence our life in a bad way, we can't stop the use of it just because of this. (B Level).

In conclusion, Although iPad and other new gadgets can be beneficial because using them tasks can be done easier (B Level).

For me, I think robots are useful if we use them in the right way. (B Level)

If students were not good at using iPad, they can teach each of them (B Level)

Take an example, if they get a mechanical problem, they are fixed by humans. (PG)

However, if the robots were to replace humans in repetitive tasks, that would require a large manufacturing, maintenance, and research and development industry which would also consist of humans. (PG)

Overall, the distribution of finite adverbials points to their decrease as students move into higher writing levels. Nevertheless, the use of finite adverbials particularly in low level writing indicates that despite the frequency use of these structures, learners exhibit problems in constructing structurally accurate structures. No clear pattern can, however, be detected in the use of non-finite adverbials due to their scarcity in the corpus.

Contrary to the frequency use of finite adverbials, the use of finite verb complement tends to be more frequent in higher level writing than B level. Although graduate level students displayed similar frequency use of both finite and non-finite verb complements, more differences appear between undergraduate levels. The low-level B shows a low frequency use of finite or that verb complement compared to other levels. The distribution of non-finite complement of verbs tends as visualized in Figure 4.4 to increase from low level B to Pass undergraduates. The only pattern that can be delineated when comparing the frequency use of finite or That verb complement and non-finite complements in graduate vs. undergraduate is that all levels tend to have higher frequency use of these structures compared to B level.
Similar to *that* noun complement clauses, *that-clauses* as verb complements can also be used to express one’s stance (example 19) or convey others’ position toward a particular proposition (example 20). It is hence not unexpected that they are commonly used in argumentative essays across levels, particularly high-level writers.

(19)  *I personally believe that benefits of use of iPads in classrooms, do outweigh their challenges.* (PU)

(20)  *Experts believe that if we are free and out of jobs which are being done by robots we are able to do our hobbies and do activities.* (C Level)

![Figure 4.4 Mean Rates of Occurrence (Normalized per 1,000 Words) of Finite and Non-Finite Verb Complements](image)

To sum up, the results of descriptive statistics pointed to some expected and other unexpected results. In line with the previous research findings regarding the linguistic characteristic of academic writing (Biber, 1985,1986, Biber and Gray,2010; Biber, Gray and Poonpon, 2011, Taguchi, Crawford and Wetzel, 2013; Crossley and McNamara, 2014; Lu, 2011, Kyle and Crossley, 2018), the higher placement levels, particularly graduate pass and D level,
included a more frequent use of both prenominal (i.e., adjectives) and post-nominal modifiers (prepositional phrases, nonfinite relatives, *that*-clauses as noun complements). Although the frequency use of attributive adjectives is shown to be more frequent in higher level writing particularly graduate levels, it appears to be a poor indicator of syntactic complexity in low level writing. The tendency for lower level writers to compensate for their low lexical repertoire through borrowing or repetition seems to mask differences between writers in adjacent low levels. The other unexpected result is that the differences in the use of nouns as prenominal modifiers were found to contradict Biber, Gray, and Poonpon’s (2011) hypothesis on writing development which postulates that these features are acquired at a later stage than attributive adjectives and hence are indicative of high-level writing. The use of pre-modifying nouns was not discriminating enough of writing levels. In fact, the low- Level B tends to use pre-modifying nouns as frequent as high levels.

In the same manner, the use of nouns found to be common across all levels with small distinctions among levels. In the author’s view, the lack of distinctions in using nouns in general and nouns as pre noun modifiers can be attributed to the task characteristics; this interpretation is supported by findings of Biber, Gray, and Staples (2016) that both noun and pre-noun modifiers are significant predictors of both mode and task type (integrated versus independent). Another potential reason is that noun+ noun sequence is more common in informational texts, e.g., newspaper and academic prose (Biber et al., 1999) than in argumentative essays.

**Conclusion**

This study examined the frequency use of linguistic features postulated to be characteristic of syntactic complexity in academic writing. The results generally support the validity argument for the EPT writing test as higher levels of writing display more syntactically complex language. The findings come in line with the EPT rubric that higher levels use more
complex sentences. This was reflected in the higher means of pre and post noun modifiers used by higher placement levels. In addition, the high frequency use of nominalization and *That* complement clauses of nouns conform with the rating rubric postulation that high-level writers especially graduates use a good summary of previous arguments.

The examination of learners’ writing, however, illuminates some lack of specification in the rating rubrics. Even though simple sentence structures can be reflected in various manners, no illustration of simple vs. complex structures are provided in the EPT rubrics. The current study supports the contention that noun phrase complexity is an indicator of development in academic writing ability as it allows for more specific yet compressed construction of meaning. Hence, complexity in the noun phrase should better be posited as an aspect of complex sentence structure in the EPT writing rubric. The nature of integrated tasks, however, seems to promote more use of nouns which may in turn contribute to the confusion regarding differences in features (nouns, pre-modifying nouns) considered to be major characteristics of academic writing.

Also, while the results of this study conform with the assumption of the EPT rubrics that the Low -level B relies mainly on common subordinate conjunctions and that high-level writers make an adequate use of both true and untrue conditionals, the findings show that the frequency of finite adverbials is not a sufficient criterion. The accuracy in using these structures is of equal if not more importance given their relative frequency across all levels.

**Implications and Future Research**

The major implication of investigating syntactic complexity in ESL writing is that examinations of features of syntactic complexity in L2 writing need to consider the L2 context, education level and the task characteristics. The range of language proficiency reflected in a
particular test or how homogenous or heterogeneous the groups are, affect the distribution of syntactic complexity features. In addition, the education level of students tends to influence their mastery of specific syntactic structures. Rating rubrics of writing need hence, more specific description or at least examples from actual learners’ writing that illustrate how complexity is displayed in various levels. This step is not only important for decreasing the variations among raters in terms of what constitute simple and complex structures but also for highlighting specific linguistic needs and, accordingly, inform the development of appropriate writing materials for learners in different placement levels. The pedagogical practices need also to highlight the discourse functions that certain linguistic structures mainly complex noun phrases, nominalization play in constructing syntactically complex, yet explicit and easy to follow arguments.

The study is not without limitations. One major limitation is that the analysis of syntactic complexity did not take the accuracy of using the syntactic structures into consideration. It also did not consider the interface between the linguistic features and the complexity of the lexis used on one hand and their semantic functions on the other hand. For example, the attributive adjectives can convey evaluative or mere descriptive meaning and they can also differ in their lexical complexity. Given their frequency across all levels, it would be insightful to know how semantic functions and lexis complexity vary with the writing ability levels. In addition, the conclusions regarding how finite adverbials particularly real and unreal conditionals were used in low- and high-level writing were based on general inspection of concordance lines, a more extensive and systematic analysis is needed in future research. Another area worth of research is how certain syntactic features particularly nominalization and subordinate clauses contribute to the organization of texts across levels particularly when they are used in sentence initials.
CHAPTER 5. DISCOURSE ANALYSIS OF TOPICAL DEVELOPMENT

A considerable amount of research has examined writing quality through various textual features including vocabulary complexity, length of texts, and syntactic complexity (Cumming et al., 2005, 2006; Knoch, Macqueen & O'Hagan, 2014). Fewer studies, however, have focused on topical development or organizational patterns and the specific syntactic forms that writers use to initiate and their sentences. Topical development is concerned with how writers shift from, repeat, or return to a previously discussed topic (Connor, 1987; Lautamatti, 1987). A topic is defined as “the main topic of the sentence”. Topics in English sentences are typically followed by a comment or “what is being said about the topic” (Connor, 1987, p. 683). One possible reason for the relative scarcity of research compared to syntax and lexical features is that such type of analysis requires a manual coding, which can be time consuming particularly with a large number of texts. In the present dissertation, topical development is designated as the third major component of linguistic complexity beside lexical and syntactic complexity. The inclusion of topical development as a major component of linguistic complexity construct reinforces the connection of the of three linguistic aspects i.e. lexical, syntax and discourse organization on one hand and their relatedness to the context of language use on the other hand.

The relatedness between organizational patterns and other aspects of linguistic complexity is reflected by the emphasis of recent theories of writing ability on textual aspects at the word, syntax and the overall textual organization (see chapter 2, section 2.1.1). For example, Bachman and Palmer (1996) endorses textual organization, grammar and vocabulary as indicators of “organizational competency” (see more discussion in Ch.2, section 2.1.1). McCarthy (1991) also stressed that the use of appropriate progression patterns is an indicator of the overall writing ability whereas the lack of thereof denotes limited grammatical resources:
“Low-level learners might be trapped in unnatural patterns [of topical development] owing to limited grammatical resources… but most advanced learners are likely to have a good feel for creating topic frameworks and orienting their audience” (McCarthy, 1991, p. 58).

Similar to lexical and syntactic features, the distribution of progression patterns and the linguistic forms that are utilized in connecting ideas found to be influenced by genre (narrative vs. argumentative) and by register (e.g., spoken and written). For example, some progression types were found to be more frequent in narrative and spoken discourse (Tod, 2016) whereas others were more prevalent in argumentative writing (Knoch, 2007; Schneider & Connor, 1990).

Acting as indictors of complex language use and of quality writing, topical development has been used in evaluating textual coherence. One method that has been used in investigating organizational patterns favored by proficient writers is topical structure analysis (TSA) (Lautamatti, 1987; Knoch, 2007; Schneider & Connor, 1990). TSA focuses on how the topical subject of a sentence is related to a previous topic or comment in a text and what structures are utilized to maintain the flow of ideas throughout the text.

Acknowledging the lack of research on organizational patterns in ESL writing in general and validation research of writing in particular and that linguistic complexity goes beyond individual lexical and grammatical features, the current study aims at investigating complexity of organizational patterns in EPT written responses. Complexity at the discourse level is defined in this paper in terms of the ability of constructing textual connection using appropriate progression patterns and linking devices that have been designated by previous research as indicators of writing development. With this purpose in mind, TSA will be utilized to investigate how topical development is reflected in different placement levels.
To identify progression patterns that found to be indicative of writing level, the first part of this study will focus on reviewing and evaluating the findings from TSA research. The second part will be dedicated to a discussion of three components that are important for analyzing topical development. Building on the conclusions and limitations of previous research using TSA and the preliminary analysis of EPT responses, the third section discusses the modified description of topical organizational patterns that is adopted in this study. The fourth section provides an overview of the methodology, followed by a discussion of the findings and limitations of the current research.

**Characteristics of Topical Structures in High Level Writing**

TSA has been used as a tool to analyze differences in topical development across different levels of writing. Some studies examined differences in organizational patterns between simplified and authentic texts (Lautamatti, 1987; Witt, 1983). Others were concerned with how writers of different proficiency levels (Knoch, 2007; Knoch, Macqueen & O’Hagan, 2014; Schneider & Connor, 1990) or how native versus nonnative writers of English vary in their topical development patterns (Simpson, 2000). Even though it highlights aspects beyond mere individual lexical and grammatical patterns, research on topical development is flawed by methodological limitations, including small sample sizes and a lack of clear guidelines for defining different organizational patterns, as will be fleshed out below.

One of the first studies that used TSA as an analysis tool for patterns of discourse development was that of Lautamatti (1987). Several language teachers and applied linguistics were asked to read an authentic text and to simplify it in a way that makes it more readable to foreign learners at the tertiary level. The simplified texts versus the authentic texts were examined using three types of topical progression: parallel, sequential, and extended progression. Parallel progression refers to the progression in which “topical subjects of successive sentences
have the same referents” (p. 95). Sharing the same referent in parallel progression is not limited to a single referent but can also refer to ideas developed throughout the entire text (Lautamatti, 1987). In sequential progression, “the rhematic part of one sentence gives rise to the topical subject in the following” (p. 95). Extended parallel is used when the same topical subjects are interrupted by at least one sentence interval. The analysis of simplified and authentic reading texts revealed that the former contains more parallel types, whereas the latter is distinguished by more frequent sequential progression. In her study, Lautamatti noticed that parallel, as well as sequential, progression can be related not only through direct relationship of mere repetition, but also through indirect relationships such as “contrast or conceptual inclusion” (p. 108). As for extended progression, it was relatively rare with only one instance in simplified texts. Using descriptive statistics, Lautamatti’s study provided a detailed analysis of both the patterns and the intricate nature of analyzing topical progression. However, the findings are based on a small sample of texts (one authentic text and four simplified ones) written by proficient native speakers. Supporting findings with quantitative data on typical trends (as well as inferential tests) was, hence, not possible.

Witt (1983) conducted a study in which 80 native speakers of English were asked to revise an authentic text (the same text used in Lautamatti’s study) for a college educated audience familiar with the topic at hand. The resulting texts were then rated by four raters of different backgrounds (two English teachers, one lawyer, and one CEO executive) using a four-point scale of comprehensibility (1 indicating the most difficult to understand to 4 representing the easiest to understand). The revised texts were then classified based on their scores into high-level scores (easily comprehensible texts) and low-level scores (difficult-to-comprehend texts). The T-test showed that the texts that presented comprehension difficulty used statistically more
sequential structures, whereas the easily comprehensible texts used statistically more parallel structures. The results are reminiscent to that of Lautamatti’s study in that the simplified texts in her study, reflected in the easy-to-comprehend texts in Witt’s study, utilized more parallel structures, whereas the authentic texts, reminiscent of the difficult-to-comprehend texts, incorporated more sequential structures. Witt’s study then gives support to Lautamatti’s findings that sequential analysis contributes to the complexity of language by providing more elaborations and that parallel structures reinforce the topic and, therefore, help facilitate cognitive processing.

These results suggest that the use of sequential and parallel progression creates a balance between reiteration and elaboration in discourse and, as a consequence, facilitates cognitive processing of information flow. The structures reported in Witt’s study reflect advanced levels of writing, as the participants were native speakers of the language. However, the division of high and low scores was based on the ease of comprehension, rather than the quality of the text; this means that, apart from complexity, both versions still represent high-quality writing. Another shortcoming of Witt’s study is that the raters were from different backgrounds and no inter-rater reliability was reported.

Shibayama (1987) compared the average of progression types in a widely used nursing text book. Five samples each with 10 paragraphs were randomly selected and analyzed. Sequential progression followed by parallel progression were found to be the most commonly used structures in textual connection.

In 1988, Connor and Schneider examined argumentative essays from TWE (Test of Written English) written by two groups of ESL examinees (from both high and low proficiency levels). Fifteen texts from each group were randomly selected and analyzed in terms of their topical progression from a pool of thousands of essays for TWE. A T-test was used to compare
the mean differences between the high and low groups’ writing in the number of T-units, and the proportion of parallel, sequential, and extended parallel progressions in each essay. The results showed that the high-level group used a statistically greater proportion of sequential progression than the low-level group. No statistical differences were found between the two groups in the proportion of other types of progression (extended parallel and parallel). Their findings come in partial support of that of Witt (1983) and Lautamatti (1987). One common finding among the three is that sequential progression is a characteristic of proficient writing. However, given that parallel progression was used by native speakers in Witt’s (1983) and Lautamatti’s (1987) studies, one would expect that the writing of more advanced levels would also show more parallel progression than the lower level writing, which was not the case in Connor and Schneider’s study. In the author’s point of view, one major reason for such a discrepancy is that none of the studies cited above account for the differences in realizing parallel structures, which subsequently blurs the picture of whether the proficient- and low-level writers use the same strategies for developing parallel topical structures.

In 1990, in a more of a confirmatory type of research, Schneider and Connor used TSA to compare an additional 15 texts from writers at an intermediate level of proficiency, with a total of 45 argumentative essays (15 advance, 15 intermediates, 15 low) written by ESL examinees in TWE analyzed. Three groups (high, intermediate, and low proficiency levels) were compared in terms of their frequency use of three topical progression (parallel, sequential, and extended parallel). ANCOVA with the T-unit length as a covariate, was used to run three univariate tests on the proportions of three types of progression. Similar to their previous findings, they found that high-level learners used a statistically more sequential progression. Contrary to the findings from their first study, the high-level learners also used a statistically lower proportion of parallel
progression instances compared to lower level groups. Schneider and Connor should be credited for using raters who were experienced in rating ESL writing and for providing detailed information about their coding procedures for the three progression types.

However, in both of their studies, 1988 and 1990, the authors adopted a broad definition of sequential progression and a narrow definition of parallel progression, both of which were different from Lautamatti’s original framework, which they claimed to follow. Parallel structures were restricted to exact repetition, pronoun reference, singular versus plural forms, affirmative versus negative forms, and synonyms. Sequential analysis on the other hand, involved any topic that is different from the previous topic. This means that cases in which parallelism of progression is achieved through other means, such as antonyms (e.g., advantages, disadvantages), derivatives (exam, examination) and extended reference -- referring to long stretches of discourse rather than one preceding sentence (e.g., this, that) -- were included under sequential progression. In fact, the researchers do admit that their coding guidelines contributed to the sizable increase of sequential progression counts in their findings.

In a more recent study, to examine potential differences in organizational patterns favored by English and Spanish proficient writers, Simpson (2002) analyzed 40 paragraphs from published journals in Spanish and English in terms of thematic progression and other physical features. Twenty paragraphs from English and Spanish academic journals in the Humanities were selected. Using Lautamatti’s (1987) TSA, the study found that English paragraphs used a higher percentage of parallel progression (17.7 versus 12.2) and more sequential development (16.8 versus 6.6). No differences were found in the use of extended progression. Simpson’s findings are important in that they shed light on organizational patterns preferred by proficient English writers. The study also shows that parallel progression is a preferred organizational
pattern in combination with sequential progression. This conclusion emphasizes the problem discussed earlier of the lack of distinction between parallel structures based on the linguistic devices used.

In an attempt to measure coherence in ESL students’ essays in the Diagnostic English Language Needs Assessment (DELNA), Knoch (2007) used TSA results to develop a TSA-based rating scale. She modified the typology suggested by Schneider and Connor (1991) to account for various cases encountered in DELNA. In Knock’s study, extended parallel progression was termed “extended progression” to account not only for topics interrupted by one or more sentences; but also, to cases in which the comment on a topic becomes the topic of another sentence after one or more sentence interval(s). The term “indirect sequential progression” was changed into “indirect progression” to account for cases in which the topic or a comment of a sentence becomes the topic of a successive sentence through indirect semantic relationships (e.g., part-whole or through inference). Knoch also introduced another two types of progression: coherence break and superstructure. A coherence break encompasses cases in which the topical progression is disrupted due to an error in the use of a pronoun or an erroneous use of a linking device. Superstructure refers to cases in which the progression is signaled by a linking device rather than the progression of the topical subject per se.

Using the DELNA rating scale, ESL texts were rated into five writing levels and then analyzed using TSA. It was found that direct sequential, indirect progression, and superstructure features are associated with higher levels of language proficiency, whereas coherence breaks, unrelated progression, and parallel progression are associated with lower writing levels. Extended progression was used equally by all levels. The findings of the TSA analysis were then used to develop a rating scale for measuring coherence, with the trends in TSA used as level
Descriptors. Scores based on the DELNA scale were compared to those based on the TSA rating scales. The results show that the TSA-based scale provides clearer discrimination than the DELNA scale, as raters “were able to discern more levels of ability among the candidates and ranked the candidates more similarly” (p. 112).

A drawback to Knoch’s study is that, similar to Schneider and Connor (1991), the indirect parallel and indirect sequential categories were collapsed into indirect progression. Even though this label intended to account for linguistic realization of topical subject aside from mere repetition, it fails to account for the difference between parallel and sequential structures; with the former, the topical subject is repeated in the subsequent sentence, while in the latter, the topic is derived from the previous comment. Such an analysis, as Schneider and Connor (1991) themselves admit, inflates the number of sequential of progression counts and further generates confusion between two different progression types (indirect parallel and sequential). Another issue in Knock’s study is the subjectivity of distinguishing direct sequential and indirect sequential types, a fault that was recognized by the researcher.

TSA has also been used in a slightly different way by Knoch, Macqueen, and O’Hagan (2014) as a part of a validation study with a large number of features. The authors collapsed five types of progression under “coherent discourse” (parallel, extended parallel, direct sequential, indirect sequential, superstructure) and “non-coherent discourse” (unrelated progression and coherence breaks). It was found that writers with higher levels of language proficiency used more coherent T-units. However, Knoch et al. indicated that this trend did not hold consistent in writers of higher proficiency levels. This, in my view, is mainly caused by the collapsing of various structures, a decision that obscured obtaining a clear and more informative picture on how learners develop the discourse topics.
Contrary to findings of discourse analysis of informative written language, analysis of spoken discourse and narrative texts revealed that the use of parallel structure is more frequent. In his analysis of classroom oral discourse and anecdotal text, Todd (2016) found that parallel progression is three times more frequent than sequential progression. He concluded that the frequent use of parallel structures might be due to the small number of concepts and hence is an indicative of high dependency on topic maintenance and a lack of concept development.

Overall, three main conclusions can be derived from the research on topical progression: 1) the discourse functions served by discourse organizational patterns influence their distribution across different contexts, 2) the ability of constructing well developed texts necessitates an appropriate lexical and syntactic knowledge, and 3) high rated essays display more frequency of sequential progression patterns.

The differences reported between prominent progression patterns in spoken vs. written forms suggests that the distribution of progression patterns reflects the functions that they serve in discourse. Sequential progression which serves the function of elaboration is more prominent in writing. Parallel which preserves and reinforces the topic focus is relatively frequent in narrative and spoken language. The use of both, however, complements the need for a balance between elaboration and reiteration in facilitating the cognitive processing of information in a text. The findings from TSA do not seem to suggest a particular pattern of distribution for extended parallel. This, however, does not negate its importance in bringing the focus back to the topic at hand when the flow of ideas has been interrupted by subsidiary or complementary details or ideas.

Another important conclusion from TSA research is that in their development of a topic, proficient writers, emphasize “the semantic meanings of sentences and their sequences” (Conner
& Farmer, 1990, pp. 127-128). In other words, the development of ideas in a text is not arbitrary. Instead, it moves in a manner that facilitates the processing of ideas. The ability of creating an appropriate connection of ideas highlights the intricate relation between vocabulary, syntax and organization in writing. For example, developing contrastive/comparative topical relations and the maintenance of a topic without excessive repetition requires an adequate knowledge of both vocabulary and syntactic structures.

The review of TSA research pointed also to some indicators of complex use of language at the discourse level. High-level writers tend to use more sequential progression. This comes in line with the Quirk et. al. (1985) contention that theme/rheme progression or sequential -using the terms of the current study- ensures the order of given/ new information in English discourse. Even though sequential progression is preferred, texts with overdependence on sequential progression or too much development beyond the main topic can cause difficulty maintaining the focus of the topic (Connor & Farmer, 1990) and consequently render the text to be incoherent (Todd, 2016). The distraction caused by overdependence on sequential progression was clear in Witt’s (1983) study, where the most difficult-to-comprehend texts portrayed more sequential progressions. Findings from proficient native speakers’ writing in Simpson (2002), Witt (1983), and Lautamatti (1987) and academic writing in text books (Shibayama, 1987) illustrate that proficient writers use parallel patterns in addition to sequential progressions.

In fact, what seems to be a discrepancy between the results of ESL and native speaker research with respect to using parallel lies mainly, I believe, in the failure of studies to capture the differences between ESL and native speakers’ use of parallel structures. In contrast to Lautamatti’s suggestions, the parallel conceptualization in Schneider and Connor (1991) and Knoch (2007) was restricted to that signified by mere repetition or synonymous relationships.
Considering the lack of transparency in defining different strategies of constructing parallel structures, the next section provides also more elaboration on the definition of parallel structures. The elaboration on progression types will, in addition, consider instances found in this corpus but were not clearly delineated in the research of topical developmental patterns.

**Modification and Examples of Topical Progression Types**

The discussion of the findings from the research on topical development and the analysis of corpus in the present study, raised the need for more elaboration on the definition of parallel and sequential progressions. A preliminary analysis of the current study’s corpus supports Lautamatti’s observations that the parallel structure can be reflected through contrast, part-whole relationships, or a reference to an extended part of the text rather than a single referent. This type of progression or the derived theme has also been recognized by several studies on discourse progression including Danes (1974) and Todd (2016). Bearing this in mind, the author of this study decided to use Lautamatti’s model with minor modifications based on the subsequent study of Knoch (2007) and the preliminary analysis of the corpus in hand.

To this end, a distinction between parallel using exact repetitions and indirect parallel using forms other than mere repetition is considered. The indirect parallel designation involves instances when repetition of the topic is realized through reiteration strategies other than exact repetition. Reiteration strategies allow writers to repeat themselves, keeping the reader on track, as well as allow for the introduction of new ideas (Hoey, 1991).

They (reiteration strategies) include using synonyms, contrasts, part-whole relationships, and extended reference (Halliday & Hasan, 1976, p. 52-53). Synonyms refer to words of similar meaning, whereas contrasts imply oppositeness and can be realized through the use of lexical contrast (e.g., merits and demerits) or conjunctions only (however, while...etc.). Part-whole relationships occur when one word is semantically part of another (e.g., the internet is part of
technology). Extended reference refers mainly to a word or expression with reference to an extended portion of a text (e.g., a sentence or even one or more paragraphs) (Halliday & Hasan, 1976; Lautamatti, 1987). Extended reference can be realized by personal *it*, deictic demonstratives and the use of what is called encapsulating nouns (Goutsos, 1997). Personal *it* is used to refer to a clause or a string of clauses rather than a single nominal instance (Halliday & Hasan, 1976). Deictic demonstratives (*this* and *that*) are used to refer to chunks of a preceding discourse block. “Encapsulating” or “shell nouns” (*e.g.*, *the issue, the action*) refer to abstract nominal phrases that refer to an extended proportion of a text anaphorically or cataphorically. These can also be realized through complex subject structures that combine pieces of information from previous discourse into one complex structure (Alonso, 2006).

The preliminary analysis of topical patterns also revealed cases that were not discussed by previous studies. Particularly, the analysis of progression types revealed instances in which a new topic is initiated with reference to the source reading. In other words, the reader is prepared for the next proposition through some topicalization expressions (e.g., *with regard to*). This type of structure is labeled as sequential in this paper, because one can argue that the relatedness of the topic presented is established through a linking device to the source text. Table 5.1 provides a list of the progression types examined, along with their definitions and illustrative examples of each progression pattern, mostly originating from the EPT sample corpus investigated. The progression patterns are presented with their abbreviated forms between brackets (Note: *<T>* marks the beginning of a T-unit; topical subjects of the respective progression types are bold and underlined, comments and the referents of topical subjects are underlined).
Table 5.1 *Types of Progression with Examples*

<table>
<thead>
<tr>
<th>Type of progression (short forms between brackets)</th>
<th>Definition and Sub-Types</th>
<th>Examples (topical subjects representing respective progression types are bold and underlined)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Parallel (P)</strong></td>
<td>The topical subject is repeated exactly across T-units. Specific types include:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Exact repetition: using the same word, pronoun, or expression.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;In my experience, I don’t have an iPad, &lt;T&gt;I had a notebook &lt;T&gt;and I learned a lot using it because I didn’t use the same operation system or programs of the other people. &lt;T&gt;so I had an advantage like professional, in my C.V. (B Level)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Pronominal reference: using a pronoun to directly refer to a word or phrase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;In my opinion, iPad represents one age &lt;T&gt;Actually, it has parts functions of phone and other parts functions of computer. (B)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Affirmative or negative forms (e.g., developers, no developers, everybody, nobody)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;There are apps for sharing files. &lt;T&gt; But there are no apps which provide a complete platform for reading (PG)</td>
<td></td>
</tr>
<tr>
<td><strong>2. Indirect parallel (IP)</strong></td>
<td>The topical subject of the sentences is repeated indirectly in the next successive T-unit. Specific types of indirect repetition include:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Direct or indirect inclusion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;Most people said, with the developing of the whole world, everything becomes fast &lt;T&gt;and life becomes easy. (Life is conceptually included in everything)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Synonyms, antonyms or contrast relationships</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;advantages include, easier and faster access to article and reopening materials, being much more interesting than printed books, providing with new communication methods and so many others. &lt;T&gt;disadvantages&lt;IP&gt; include being expensive to use (D Level)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Extended reference rather than a single word or a phrase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;In my opinion, the robots will be placed on almost workplaces &lt;T&gt; and this is not beneficial for people (B Level)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Qualifiers that specify a noun, so that it refers to a different referent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;It is crucial too that they are not used to fast track juvenile offenders towards custody &lt;T&gt;we need to reduce rather than increase the prison population. &lt;T&gt;Ensuring the accountability of those using the new powers to the communities they serve is vital. (Laura Edwards, &quot;Local authority&quot;. The Guardian, 24-11-2004, Alonso, 2006)</td>
<td></td>
</tr>
<tr>
<td><strong>3. Sequential (S)</strong></td>
<td>The comment of a previous sentence is further developed to become the topic or the theme of a successive sentence with or without a linking device. This development is established through semantically direct or indirect relation with a previous comment. This include:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Using derivatives (science, scientists)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The science of biotechnology has advanced tremendously in recent years. Biotechnology research Scientists are working on manipulating living organism for the benefit of humans. (My Example)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. A topic is a part of or an example of a preceding topic.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;T&gt;First, although iPad makes student feel comfortable to study, I will bring a lot of inconvenience to students. &lt;T&gt;For example, students can easily distract when studying with iPad. (PG)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2 (continued)

<table>
<thead>
<tr>
<th>Type of progression (short forms between brackets)</th>
<th>Definition and Sub-Types</th>
<th>Examples (topical subjects representing respective progression types are bold and underlined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c. Inferences, semantic sets, contrast/sequence with or without a linking device</td>
<td>&quot;These applications are available for free. &lt;T&gt; hence the student is no required to spend more than compared to searching and gathering data the normal way. (PG).&quot;</td>
<td></td>
</tr>
<tr>
<td>d. Relatedness to the source reading text.</td>
<td>&quot;Coming to the cost of iPads pointed by Erica, This cost would be equivalent to the printed materials which would be used later on (PG).&quot;</td>
<td></td>
</tr>
<tr>
<td>4. Extended Parallel (EP)</td>
<td>a. Repetition of the topical subject after more than one T-unit interval</td>
<td>&quot;Nevertheless, if students buy electronic books online, these books can be transferred to other students, new and clean. Another advantage that should be emphasized is its convenience and portable. Anywhere and anytime, students just need to take the iPad and can do anything (PG).&quot;</td>
</tr>
<tr>
<td>5. Coherence Break (CB)</td>
<td>a. The progression from one topical subject to the topical subject in the successive T-unit is obscured. Specific examples of obstruction include:</td>
<td></td>
</tr>
<tr>
<td>b. Absence of a topical subject</td>
<td>&quot;I say a robot, which is loading the truck with differed fruits. &lt;T&gt; In addition (Missing TS) is measuring which trick how much has been loaded and checking every trick whether good to transport or not (B Level).&quot;</td>
<td></td>
</tr>
<tr>
<td>c. Unclear relatedness between one topical subject and the previous topical subject or comment.</td>
<td>&quot;Some goals may be easy is a person can achieve them easily. &lt;T&gt; However, when a person cannot achieve his/her goals, they will be difficult. &lt;T&gt; If robots will do all our work, I will not achieve my goals (B Level).&quot;</td>
<td></td>
</tr>
<tr>
<td>d. Erroneous use of a pronoun referent.</td>
<td>&quot;Some robots can take over some simple... jobs so that make our life can be more efficiency &lt;T&gt; and it can do better than human labor (C Level).&quot;</td>
<td></td>
</tr>
<tr>
<td>e. Vague use of a pronoun referent.</td>
<td>&quot;physic cannot be done by any robot. &lt;T&gt; Thus, there are still many works, which cannot be replaced by robot. &lt;T&gt; On top of that it can support humans work. (D Level)&quot;</td>
<td></td>
</tr>
<tr>
<td>f. Detachment of a pronoun from its referents by one or more T-units</td>
<td>&quot;they also want to gain agreement or verification through collecting the number of &quot;like&quot;. &lt;T&gt; But it doesn't work, &lt;T&gt; they need to take action to get rid of their bad habits (B Level)&quot;</td>
<td></td>
</tr>
</tbody>
</table>

An important step before investigating topical progression is the identification of topical subjects. This requires an understanding of three basic notions: ISEs, topical subject, and mood.
subject. ISEs refer to “initially placed discourse material in sentences, whatever its form or type” (Lautamatti, 1987, p. 91). The topical subject is the linguistic element that represents the topic of a sentence. The mood subject is a grammatical subject that may or may not align with the topic of a sentence (Lautamatti, 1987). Investigating the relationship between the topical subject and the initial sentence elements is important for illuminating not only how to identify topical subjects, but also how the topical subjects are introduced to the reader. In other words, how the relationship between the initial sentence elements and the topical subject affect the overall information flow in discourse. Therefore, the forthcoming section will investigate topical structures i.e. the topical subject and ISEs. The discussion concludes with an operationalized definition of discourse level complexity in this paper.

**Topical Structures**

Each sentence in English has a topical subject or a linguistic expression that signifies what the sentence is about (Conner & Farmer, 1990; Lautamatti, 1987). The topical subject can, but not necessarily, coincide with the mood (grammatical) subject of a sentence. Various linguistic devices for framing the discourse topic are at writers’ disposal (Schiffrin, 1987). In fact, in her detailed analysis of simplified vs authentic texts, Lautamatti noticed that the topical subjects in authentic texts are usually preceded by initial sentence elements (ISEs) and or mood subjects. ISEs are operationalized in this paper as the initial adjuncts that precede the first mood subject in the T-unit, be it topical or non-topical.

The topical subject can be preceded by a single or multiple sentence-initial adjunct, as illustrated by the examples in Table 5.2. Examples 1-5 illustrate single ISEs i.e. one constituent or one type of syntactic structure, e.g., coordinate conjunction or prepositional phrase. Example 6 includes multiple ISEs i.e. more than one type of syntactic structures, e.g., coordinate
conjunction+ prepositional phrase. ISEs can be reflected in one word, phrase, or clause with various semantic functions. In Example 1 below, the topical subject in is preceded by a single adverbial word signaling stance. ISES can also be in the form of subordinate clause adverbials, as in Example 2, a non-finite supplemental clause, as in Example 4, or a prepositional phrase, as in Examples 3 and 5. Moving from given or previously mentioned discourse, initial clausal adverbials, or ISEs using the current study’s terminology, contribute to the informational flow of discourse (Biber, Johansson, Leech, Conrad & Finegan, 1999).

Table 5.2 Examples of Initial Sentence Elements (ISEs)

<table>
<thead>
<tr>
<th>ISE</th>
<th>Topical mood subject</th>
<th>Predicate of the topical mood subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Undoubtedly video games</td>
<td>created more problems than benefits.</td>
<td></td>
</tr>
<tr>
<td>2. When The knowledge of utilizing the iPad devices can be easily acquired, students should be well equipped with those devices.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. With the development of technology more and more advancing electronic products are created (D Level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. To read an article or an essay students need to staring at iPad's screen for a long time (B Level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. According to text 1 mentioned, some of the regular works such as pharmacists, prison guards can operate well with the robots (B Level )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. but by using an iPad they can download all there books in the iPad (B Level )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The examples above illustrate that writers utilize various forms of adjuncts with different functions to create intersentential connections. Adverbials or intersentential connecters, regardless of their forms, however, can be categorized based on their functions into three major types (Biber et.al, 1999): 1) circumstance adverbials, which relate a proposition to a time (e.g., in the near future), a place (e.g., in the classroom), or a manner, 2) stance adverbials expressing
feelings or attitudes (e.g., without question, obviously), and 3) linking adverbials (e.g., however, therefore).

In fact, the research suggests that the semantic functions served by various syntactic forms of initial sentence elements can be indicative of the mode (spoken vs written) and of writing proficiency as well. In their study of the frequency of rhetorical functions of specific syntactic forms used by Native vs. ESL researchers in oral presentation and articles, Rowley-Jolivet and Carter-Thomas (2005) found that prepositional phrases as adverbials of place in initial sentence position (e.g., in this picture) are more commonly used in oral presentations by native speakers. Schleppegrell (1996) compared the semantic functions of because in ESL vs. native speakers’ writing. Using because as a cohesive device that links given and new discourse was found to be among the dominant functions utilized in native speakers’ writing.

In addition to ISEs, the topical subject can be introduced by a mood subject. Mood subjects refer to the lexical elements that occupy the slot of the grammatical subject of a clause. The mood subject may or may not be the topical subject (Lautamatti, 1987). For example, in Figure 5.1, the mood subject in the first sentence “the study” is topical. In other words, it refers to the topic that is further explicated by the rest of the sentence. The non-topical mood subject, however, appears mainly in the form of a dummy mood subject: anticipatory it, as in the third sentence in Figure 5.1 and existential there “there are several reasons that this could happen”. In both examples, anticipatory it and existential there have no lexical meaning and merely function as the grammatical subject of the clause (Biber et.al, 1999). The non-topical mood subject can also be lexical. In Figure 5.1, the second sentence includes two mood subjects (one in the main clause, and one in the embedded clause). The topical subject i.e. “system” is introduced by the
mood subject “the study”. This example illustrates that the topical subject can be preceded by mood subject that can be either lexical or a referentially empty element.

*Figure 5.1 Types of a Mood Subject*

To sum up the discussion on topical development and topical structures, both topical progression patterns and ISEs provide insights on how proficient writers introduce their arguments to the reader. They also shed light on problems that result from constructing and introducing topics in discourse. Unfortunately, ESL research on writing and validation research has paid little attention to these aspects. The review of literature on topical progression highlighted several limitations from previous research regarding the misinterpretation of sequential progression and oversight of the differences between different realizations of parallel structures. Also, even though some studies examined the frequency of certain types of ISEs, e.g., linking devices in ESL in writers’ texts (e.g., Shaw, 2009), the use of *because* as a liking device (Schleppegrell, 1996), few studies have examined the different syntactic forms and the semantic meanings of ISEs that writers of different placement levels utilize in their writing. Such a line of research is needed to highlight the connection between the use of certain grammatical structures and their rhetorical functions in discourse.

**Purpose of the Study and Research Questions**

The review of literature on topical development patterns suggests a relatedness between the use of specific organization patterns and writing quality. However, and compared to lexical
and syntactic aspects, organizational patterns received less attention particularly in validation
research of writing. Moreover, despite the plethora research on linking devices in ESL writing,
there is less research on the different syntactic forms writers of different levels use in initiating
their sentences and how they contribute to discourse connection. With these limitations in mind,
the current study attempts to answer the following questions:

1. To what extent do ESL writers of different proficiency levels differ in their use of topical
   progression (parallel, indirect progression, sequential, extended progression, coherence
   break)?

2. To what extent do ESL writers of different proficiency levels differ in their use of initial
   sentence elements in terms of frequency, linguistic realizations, and semantic functions?

Methodology

Corpus

The corpus used for the analysis of topical development is the EPT corpus (ISU
EPT Corpus of Learner Writing, Release 2.0) which includes the written responses for read-
argue tasks (see chapter 3, section 3.2.1) for all test takers from January 2016 through January
2017 (N = 991). The use of read-argue tasks was to ensure compatibility with the task used to
investigate lexical and syntactic complexity. The file numbers of the whole corpus texts were
first arranged in an ascendant order in an Excel sheet, then 20 files were selected from each
placement level. The files were selected from either the top, mid or the last 20 file names in each
placement level. However, because of time constraints and the intensiveness of manual coding of
progression types and ISES, only 92 texts were coded. Even though the number of texts is not
equal across all levels, it is relatively balanced ranging from 17 to 20 texts in each placement
level (see Table 5.3). The sampled essays are from spring, summer and fall 2016. The topic of
the text prompt for these semesters focused on using a particular kind of technology, which
students provided their opinions on based on both the given texts and their personal experience.

A summary of the texts and the mean of T-units is provided in Table 5.3.

Table 5.3 *The Sample of EPT Writing Corpus used for Topical Analysis*

<table>
<thead>
<tr>
<th>Level</th>
<th>Graduate</th>
<th></th>
<th></th>
<th>Undergraduate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of</td>
<td>Mean</td>
<td>SD</td>
<td>Number of</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>Texts</td>
<td>number of</td>
<td></td>
<td>Texts</td>
<td>number of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T-units</td>
<td></td>
<td></td>
<td>T-units</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>20</td>
<td>14.50</td>
<td>3.76</td>
<td>17</td>
<td>17.65</td>
<td>4.21</td>
</tr>
<tr>
<td>Level C</td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td>18.63</td>
<td>5.40</td>
</tr>
<tr>
<td>Level D</td>
<td>18</td>
<td>16.06</td>
<td>4.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level B</td>
<td>18</td>
<td>15.06</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coding Data**

Texts were first divided into T-units to facilitate the identification of ISEs and the topical subject. Each T-unit was first coded manually for the type of progression and for whether the topical subject of the T-unit is preceded by ISES. Given the purpose of this study and the need to calculate the frequency of ISES, the coding of the ISEs was based on their occurrence at the beginning of each T-unit, regardless of whether the ISEs were realized by one or multiple ISEs that precede the first mood subject (see Appendix C for an example of a coded text). To account for the influence of the number of T-units on the number of the targeted features, the total number of each progression type and ISEs was divided by the number of T-units in each respective text. For example, the proportion of parallel structures is calculated by the total instances of Parallel progression in a text/ total number of T-units in that text. Similarly, the proportion of ISEs was calculated by the number of T-units initiated by an ISE in a text/ total number of T-units for that text.

The counts of T-units, number of different progression types, and ISEs in each text were obtained using a simple Python code. To identify the syntactic forms of ISEs, AntConc software
was used to identify T-units initiated by ISEs. The T-units initiated by ISES were then exported into Microsoft Excel sheets. Syntactic forms of different ISES were then identified and coded accordingly in the Excel sheet, e.g., if the ISE is in the form of a prepositional phrase or a single adverb. Syntactic forms were then analyzed qualitatively for their semantic meanings. For example, an ISE in the form of a prepositional phrase can have the meaning of circumstantial (e.g., in the factories) or function as a linking device (see section 5.6.3 for examples).

To ensure the reliability of coding the categories of topical progression, a subset of data (9 texts) were coded by another researcher who is also a Ph.D. student in applied linguistics and who has experience in discourse and grammar analysis. Before the independent coding of the 9 texts, 2 texts (not included in the analysis) were used for training purposes. The proportion of the exact agreement was 99% (100 agreement cases out of a total of 101 coded segments) for the T-unit and 98% (99 agreement cases out of 101 coded segments) for topical subjects. The inter-coder agreement of topical progression types reached 80% (73 cases out of 92 coded segments). To account for potential chance agreement between the two coders, Cohen Kappa, a measure of inter-rater agreement for categorical scale, was also calculated, and it was found to be of moderate strength .74.

**Statistical Analysis**

To answer the first question regarding the differences in the use of topical progression, descriptive statistics of the mean proportion are used to describe the distribution of the progression types in each writing level. To examine whether the differences in the distribution of topical progression are statistically significant, Multivariate test of MANOVA was conducted with the progression type (parallel, indirect parallel, extended progression and sequential) as the dependent variables and the placement level as the independent variable. One progression type (coherence break) was not included because it violates the assumptions of normality and the
homogeneity of variance.

As for the second question regarding the frequency use of ISEs, one-way ANOVA is used to examine if the differences in the mean proportion of ISEs among groups are statistically significant. To obtain a general idea on the most common syntactic realization of ISES in the corpus, the average proportions of each syntactic form in the corpus will be described. For example, the proportion of finite clauses out of all ISES in the corpus is calculate by dividing the number of finite adverbials used in initiating T-units to the total number of ISEs in the corpus. The placement groups will then be compared in term of the normalized frequency of each syntactic form of ISEs. The normalized frequency for the realization of the ISEs is calculated (number of instances of a feature / total number of occurrence instances of a feature in all corpora*100) for each group. Because the counts of ISEs are calculated from T-units initiated by ISEs, the normalization is based on the number of T-units not the number of words.

Results and Discussion

Topical Progression

1. To what extent do ESL writers of different proficiency levels differ in their use of topical progression (parallel, indirect progression, sequential, extended progression, coherence break)?

The descriptive statistics of the mean proportions of each type of progressions, as shown in Table 5.4 and Figure 5.2, revealed differences in topical developmental patterns among the five groups. As for graduate levels, the Pass seems to have higher means of parallel (M = .18, SD = .09) compared to their D level collogues (M = .17, SD = .17). Undergraduate levels also display differences in their mean use of Parallel structure as the mean use tend to decrease from the B to the Pass level.
Table 5.4 Means of Topic Progression Types

<table>
<thead>
<tr>
<th>Status</th>
<th>Level</th>
<th>n</th>
<th>Parallel Mean (SD)</th>
<th>Indirect Parallel Mean (SD)</th>
<th>Sequential Mean (SD)</th>
<th>Extended Progression Mean (SD)</th>
<th>Coherence Break Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate</td>
<td>Pass</td>
<td>20</td>
<td>.18 (.09)</td>
<td>.15 (.16)</td>
<td>.44 (.15)</td>
<td>.14 (.09)</td>
<td>.01 (.02)</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>18</td>
<td>.17 (.15)</td>
<td>.11 (.07)</td>
<td>.46 (.16)</td>
<td>.13 (.12)</td>
<td>.05 (.07)</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>Pass</td>
<td>17</td>
<td>.19 (.17)</td>
<td>.14 (.11)</td>
<td>.48 (.15)</td>
<td>.08 (.06)</td>
<td>.04 (.06)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>19</td>
<td>.21 (.13)</td>
<td>.11 (.08)</td>
<td>.36 (.11)</td>
<td>.14 (10)</td>
<td>.09 (.09)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>18</td>
<td>.24 (.13)</td>
<td>.09 (.06)</td>
<td>.35 (.13)</td>
<td>.10 (.05)</td>
<td>.13 (.08)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>92</td>
<td>.20 (.14)</td>
<td>.16 (.16)</td>
<td>.41 (.15)</td>
<td>.12 (.09)</td>
<td>.06 (.08)</td>
</tr>
</tbody>
</table>

Figure 5.2 Mean Proportion of Parallel Progression in EPT Writing

To the contrary of parallel, the mean use of indirect parallel, as illustrated by Figure 5.3, tends to increase with the placement levels. The graduate Pass (M = .15, SD = .16) are shown to outperform their peers in D level (M = .11, SD = .07). Similar pattern is also noticed among the placement levels of undergraduates as the mean use increases along with the placement levels.
Figure 5.3 Mean Proportion of Indirect Parallel Progression in EPT Writing.

The use of sequential patterns is also shown to increase along with placement level. However, the increase tends to be linear across undergraduate but not graduate levels. Graduate students in D level, as demonstrated by Figure 5.4, show slightly higher means of sequential progression than their colleagues in Pass. The use of sequential progression, however, is shown to increase along with the placement levels of undergraduate with the least mean in B level (M = .35, SD = .13) and the highest in Pass level (M = .44, SD = .15).

Figure 5.4 Mean Proportion of Sequential Progression in EPT Writing

With respect to coherence break, the mean occurrence of topic digression is shown to decrease as the level of writing increases in both undergraduates and graduates with the graduate Pass showing the least mean of coherence break (M = .01, SD = .02). When compared to other
progression types, no clear patterns can be discerned from the distribution of extended progression.

**Figure 5.5 Mean Proportion of Coherence Break in EPT Writing**

To examine whether the differences in means among groups are of statistical difference, MANOVA test was conducted. Before conducting MANOVA, the assumptions of normality, multivariate-normality, homogeneity of variance, multicollinearity and linearity were tested. For the progression type of parallel, Shapiro-Wilk test found to be statistically insignificant across all five levels. As for indirect parallel, the normality test of Shapiro-Wilk found to be violated at two levels Pass graduates and Pass undergraduates test (p = .02 < .05). In extended progression, Shapiro-Wilk test of normality was statistically significant in level D only (p = .002 < .05).

Nevertheless, the examination of histograms and Q-Q plots for these levels show no major violations of normality. The normality test for the progression type ‘coherence break’ was found to be statistically significant in graduate Pass and D (p < .001), undergraduate pass (p = .001 < .05) and C level (p = .013 < .05). Therefore, it was decided to exclude it from further analysis. Mahalanobis test was used to examine multivariate-normality. Only one case found to be statistically higher than the critical value of Mahalanobis distance. Nevertheless, the analysis with and without this case shows no difference. Accordingly, it was decided to include this text in the analysis. The correlation among the dependent variables ranges
from .2 to .5, so multicollinearity was not an issue. Before running MANOVA and as a precaution for the minor violation of normality, the non-parametric Kruskal-Wallis test was run. Since no differences in the results were found between MANOVA and the non-parametric test, it was decided to move on with using MANOVA test.

The results of MANOVA revealed a statistically significant main effect for progression types on placement levels, $F(16, 348) = 2.836, p < .001$; Pillai's Trace = 0.466, partial $\eta^2 = .12$, which is a medium size effect (Gray & Kinnear, 2012, p. 328). Univariate tests as shown in table 5.5 showed that there were significant differences across the placement levels in the distribution of sequential progression $F(4, 87) = 3.42, p = .012 < .05$, partial $\eta^2 = .14$, which is a medium size effect (Gray & Kinnear, 2012, p. 328). No statistical differences were found among the groups in the distributions of parallel, indirect parallel and extended progression. Post hoc tests (LSD) -as shown by table 5.6- revealed that the graduates in Pass and Level D outperformed Level B students in the use of sequential progression and the difference was statistically significant: ($p = .046 < .05$) and ($p = .02 < .05$) respectively. In addition, Level D displayed statistically higher means for sequential progression than Level C ($p = .03 < .05$). Similarly, undergraduate Pass level used statistically more sequential progression than the lower levels of C ($p = .008 < .05$) and B ($p = .006 < .05$).
Table 5.5 Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>Parallel</td>
<td>.065^a</td>
<td>4</td>
<td>.016</td>
<td>.890</td>
<td>.473</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>.052^b</td>
<td>4</td>
<td>.013</td>
<td>1.243</td>
<td>.299</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>.278^c</td>
<td>4</td>
<td>.069</td>
<td>3.420</td>
<td>.012</td>
<td>.136</td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>.056^d</td>
<td>4</td>
<td>.014</td>
<td>1.728</td>
<td>.151</td>
<td>.074</td>
</tr>
<tr>
<td>Intercept</td>
<td>Parallel</td>
<td>3.571</td>
<td>1</td>
<td>3.571</td>
<td>195.311</td>
<td>.000</td>
<td>.692</td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>1.298</td>
<td>1</td>
<td>1.298</td>
<td>124.178</td>
<td>.000</td>
<td>.588</td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>15.980</td>
<td>1</td>
<td>15.980</td>
<td>786.316</td>
<td>.000</td>
<td>.900</td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>1.270</td>
<td>1</td>
<td>1.270</td>
<td>155.602</td>
<td>.000</td>
<td>.641</td>
</tr>
<tr>
<td>Level</td>
<td>Parallel</td>
<td>.065</td>
<td>4</td>
<td>.016</td>
<td>.890</td>
<td>.473</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>.052</td>
<td>4</td>
<td>.013</td>
<td>1.243</td>
<td>.299</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>.278</td>
<td>4</td>
<td>.069</td>
<td>3.420</td>
<td>.012</td>
<td>.136</td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>.056</td>
<td>4</td>
<td>.014</td>
<td>1.728</td>
<td>.151</td>
<td>.074</td>
</tr>
<tr>
<td>Error</td>
<td>Parallel</td>
<td>1.591</td>
<td>87</td>
<td>.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>.909</td>
<td>87</td>
<td>.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>1.768</td>
<td>87</td>
<td>.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>.710</td>
<td>87</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Parallel</td>
<td>5.234</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>2.273</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>18.008</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>2.066</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>Parallel</td>
<td>1.656</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indirect Parallel</td>
<td>.961</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequential</td>
<td>2.046</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extended Progression</td>
<td>.767</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .039 (Adjusted R Squared = -.005)
b. R Squared = .054 (Adjusted R Squared = .011)
c. R Squared = .136 (Adjusted R Squared = .096)
d. R Squared = .074 (Adjusted R Squared = .031)
Table 5.6 *Multiple Comparisons of LSD Posthoc Test*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(I) Level</th>
<th>(J) Level</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequential</td>
<td>Pass</td>
<td>Pass (UG)</td>
<td>-.042</td>
<td>.047</td>
<td>.374</td>
<td>-.136</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pass (UG)</td>
<td>-.018</td>
<td>.046</td>
<td>.704</td>
<td>-.110</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pass (UG)</td>
<td>.086</td>
<td>.046</td>
<td>.062</td>
<td>-.004</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pass (UG)</td>
<td>.094*</td>
<td>.046</td>
<td>.046</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>Pass (UG)</td>
<td>Pass(G)</td>
<td>.042</td>
<td>.047</td>
<td>.374</td>
<td>-.051</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pass(G)</td>
<td>.024</td>
<td>.048</td>
<td>.615</td>
<td>-.071</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pass(G)</td>
<td>.128*</td>
<td>.048</td>
<td>.008</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pass(G)</td>
<td>.136*</td>
<td>.048</td>
<td>.006</td>
<td>.040</td>
</tr>
<tr>
<td>D Level</td>
<td>Pass(G)</td>
<td>.018</td>
<td>.046</td>
<td>.704</td>
<td>-.074</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td>Pass(UG)</td>
<td>-.024</td>
<td>.048</td>
<td>.615</td>
<td>-.120</td>
<td>.071</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.104*</td>
<td>.047</td>
<td>.029</td>
<td>.011</td>
<td>.197</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.111*</td>
<td>.048</td>
<td>.021</td>
<td>.017</td>
<td>.206</td>
</tr>
<tr>
<td>C Level</td>
<td>Pass(G)</td>
<td>-.086</td>
<td>.046</td>
<td>.062</td>
<td>-.177</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>Pass(UG)</td>
<td>-.128*</td>
<td>.048</td>
<td>.008</td>
<td>-.223</td>
<td>-.034</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>-.104*</td>
<td>.047</td>
<td>.029</td>
<td>-.197</td>
<td>-.011</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.007</td>
<td>.047</td>
<td>.877</td>
<td>-.086</td>
<td>.100</td>
</tr>
<tr>
<td>B Level</td>
<td>Pass(G)</td>
<td>-.094*</td>
<td>.046</td>
<td>.046</td>
<td>-.186</td>
<td>-.002</td>
</tr>
<tr>
<td></td>
<td>Pass (UG)</td>
<td>-.136*</td>
<td>.048</td>
<td>.006</td>
<td>-.231</td>
<td>-.040</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>-.111*</td>
<td>.048</td>
<td>.021</td>
<td>-.206</td>
<td>-.017</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>-.007</td>
<td>.047</td>
<td>.877</td>
<td>-.100</td>
<td>.086</td>
</tr>
</tbody>
</table>

Note: * The mean difference is significant at the .05 level.

The current study’s findings on topical development are consistent with previous research findings that sequential progression is characteristic of advanced-level writing (Knoch, 2007; Lautamatti’s 1987; Schneider & Connor, 1991). The use of sequential progression is shown to be associated with placement levels. The writers in higher placement levels of Pass and D levels used a greater proportion of sequential progression than both the C and low-level B. The difference between Pass (graduates and undergraduates) and D level vs. the lower level B reached statistical significance at .05 level. Even though the distribution of topical structures
mainly sequential is associated with higher level writing, its sole use does not guarantee quality writing. Realizing the discourse functions of progression structures in developing argument and the appropriateness of grammar and lexical choices impact the quality of writing as well. This proposition can be best illustrated through examples of how progression types are used across placement levels.

The relative low means of sequential progression in B and C levels compared to higher level writers in Pass and D indicates a lack of ability in providing further support for a given argument. For example, instead of providing further elaboration on how the use of robots would help “people concentrate on their work”, the student in example 7c use “it” -an erroneous grammatical reference to the plural noun robots- to present redundant information “help increase ability of people in their workplace”. In some cases, instead of using sequential progression to present more supportive details, they are used to jump to new information whose semantic connection with the previous argument is hard to establish. For example, the comment- initiated topic “a lot of people” in example 7d is used to jump to a new point “stressed out by their work” rather than elaborating on how robots increase “ability of people in their work place”. Another common aspect of low-level writing is the use of comment-based themes that are indirectly or partially related to the previous comment. For example, the comment -initiated theme” a lot of people” is indirectly related to the comment of the previous sentence “ability of people in their work”.(Note: Examples 7-8 each represents a series of T-units produced by an undergraduate level B student. Each T-unit is reproduced in order, with the topical subjects underlined).
b. For example, if **robots** could work chores instead of people, people would concentrate on their own works.

c. In addition, **it** help increase ability of people in their workplace.

d. **A lot of people** are usually stressed out their work loads.

In addition to the relative low use of sequential progression, the use of sequential progression, in many instances, deviates from its main discourse function of providing further details to presenting redundant information on the comment-based topical subject. In example 8, the topical subject “robot” is used to repeat the same information presented in the previous T-unit.

(8) Level B Progression type

a. **The skills a dairy-men** has cannot be replaced by a robot no matter how advanced the robot is the knows the art

b. **Robot** can never do no matter how advanced the robot is. (C Level)

Compared to lower level writers in B and C, higher level writers in Pass and D levels show better understanding of the discourse function of sequential progression, i.e. elaborating on a previous argument. The main difference between Pass and D levels is not in their understanding of the discoursal function of the sequential progression but in the syntactic and lexical appropriateness of their propositions. Compared to Pass, the use of sequential progression in D level is in many instances inundated with errors in grammatical structures and/or ISEs that mask the connection between related topical subjects. In Example 9, the student is trying to
elaborate on benefits of iPads by enumerating some examples: “replacing text books, accessibility to a lot of data and also the entertainments options encourage students to utilize this device as their first choice”. The connection between the list of noun phrases, however, is masked as shown by the excerpt in 9b by the predicate “helps environment”, which disrupts the structural parallelism of the T-unit.

A close examination of the semantic relationship between adjacent sequential structures suggests some over-dependence on linking devices, which weakened rather than strengthened the flow of ideas. For example, the indirect causal relationship between Examples 10a-10c are masked by the inappropriate use of ISEs: “After… but later… in the end”. The statement in Example 10c is intended to introduce an outcome of the act “trying new applications”. However, the use of the linking device “in the end” turns the argument into a more of a narrative chain of actions.

(9) Level D Progression type

a. *These small and lightweight devices has some benefits.*

b. *Replacing textbooks helps environment, accessibility to a lot of data* S

c. *and also the entertainments options encourage students to utilize this device as their first choice.* S

(10) Level D

a. *After I bought, the first few months I did download a lot of papers and English material,*

b. *but later I just couldn't help trying new applications like games, videos etc.*

c. *In the end the iPad just became a device for me to surf the internet.* S
d. *I rarely used it to study, although I tried to use it for study.*

e. *To be honest, iPad is indeed a good and easy device for reading and study.*

By contrast, sequential patterns are used by Pass Level writers to solidify a previous argument through reiteration or provision of further examples. This is usually initiated by expressions that remind readers of previous discourse before introducing a further argument. The first sequential pattern in Example 11b elaborates on the discussion of using iPad in classrooms through a contrastive argument of “cost and time invested” versus benefits. In example 11c, the comment-based theme “the ipad” follows a prepositional phrase “for its popularity” that acts as a paraphrase to the previous comment. In Example 12c, the use of complex structures, rather than single key words or expressions, facilitates the connection of ideas. The writer uses a complex non-finite subject “*Keeping a student interested in studying*” to cement and summarize the previous comment and to prepare the reader for the important new information “*has been an issue in numerous communities.*” (Note: Topical subjects are in bold)

(11) Graduate Pass Level Progression type

a. *A hot topic among educators for the last few years has been on how should new technology be used in the classrooms,*

b. *while some argue that costs and time invested dont counterweight the benefits, the use of technology has definitely aided in all levels of education, from kindergarten to doctorates*  

S

c. *For its popularity, the iPad has been one of the most discussed pieces of technology.*  

S
(12) Graduate Pass Level

a. *the iPad has been one of the most discussed pieces of technology*

b. *and in many cases, it has proven to have a positive effect on children that support their studies with tablets like the iPad.*

b. *Keeping a student interested in studying has been an issue in numerous communities S*

The results of other types of progression namely parallel, indirect parallel and coherence break show no statistically significant impact on placement levels. Nevertheless, the distribution patterns of these progression types are informative and worth considering from a discourse research perspective. The frequency use of topical progressions is in line with the findings of previous research on advance students’ writing (Schneider & Connor, 1990, Knock, 2007) and it also aligns with the features of proficient expository writing (Shibayama, 1987; Simpson, 2002). The use of indirect parallel progression has shown to increase along with the placement level. Surpassing other levels in the mean proportion of indirect parallel progressions, the Pass demonstrated an ability of utilizing semantic relationships of extended reference, contrasts, synonyms, and antonyms, among other strategies, in solidifying arguments or as a summary strategy. Examples 13 and 15 illustrate the use of referents that span across long stretches of the text such as demonstratives “this” in 13c, demonstratives followed by a nominalized word as in 14b “this expenditure” or shell nouns that summarize paragraph long argument as in 15b “a gradual implementation of ipads”.

(13) Graduate Pass Level

<table>
<thead>
<tr>
<th>Progression type</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. <em>as it became obvious that students were not spending more time reading books than before</em></td>
</tr>
</tbody>
</table>
b. but also their non-academic reading was also seen to rise.

c. This made the coursework interesting.)

(14) Graduate Pass Level

a. Second, as pointed out by Erica Loop that iPads add a huge cost to the education fees.

b. This expenditure, I believe, is unnecessary and can be spent elsewhere for a better purpose.

(15) Graduate Pass Level

a. Such a way of learning may not therefore being appropriate for students at young age because at that stage a human being has a very high feeling of his senses…

b. I believe that a gradual implementation of iPads into schools can be beneficial.

In contrast, lower level writers in B and C levels demonstrated less competence in exploiting strategies of reiteration. The overdependence on parallel progressions in low-level writing, compared to other levels, reflects a lack of diverse vocabulary and lack of competence in reinforcing topic continuity without undue repetition. In fact, a close examination of low-level writing particularly level B, as shown in Example 16-17, reveals that parallel structures are realized mainly by pronouns. In many cases, low-level writing of B uses pronouns to refer to themselves not to the topic at hand. In other words, the use of pronouns switches the focus from the topic at hand to personal narratives.

(16) Level B Progression type

a. People tend to eat a lot on their free time
b. most of people will spend their time eating if they don't have job.

Thus,...

P

c. they are threat to gain a lot of weight which might cause them to die.

P

(17) Level B

a. When we use robots they can do better thanks in some area which we are good at

P

b. Thus, we have enough time to concentrate on creative thinking and problem-solving.

P

c. Human could do advanced, which robots cannot do,

P

d. we could solve those questions which we are puzzled for centuries

P

Unlike other types of progression, the distribution of EP shows no clear pattern. However, the examination of how this structure is used demonstrate differences among the groups in understanding its discourse function. The use of extended parallel serves the function of reminding the reader of the main topic and hence its usually used in conclusion paragraphs. When comparing how this structure is utilized by Pass and D compared to low-level writers, it can be noticed that the discourse functions of EP is adequately utilized in advanced writing.

Example 18 below is a conclusion paragraph. The topicalizing of Kindle in line 18c comes to conclude a previous comparison between iPads and Amazon Kindle devices:

(18) Undergraduate Pass Level Progression type

a. In conclusion, iPad can harm education more than the potential help

P

b. It has an expensive price just because of the capabilities that impede an efficient education instead of any help

P
c. A simpler electronic gadget such as Kindle can provide the students the main usefulness of iPad, without having any unpleasant aspect

The use of EP in levels B and C does not, in many cases, serve its discourse function as a summarizing and a topic reminding strategy. To the contrary, it is used to present unexpectedly redundant information. In the example below, the personal pronoun I in line 19a is topicalized when the writer talks about his personal interest in technology. The focus then, shifts to how technology reflects human greatness in line 19c. Instead of presenting further details on how technology reflects “how great a man is”, the student in 19d “I played robots” shifts back to the previously presented personal narrative (i.e. the writer’s personal interest in technology).

(19) Level C Progression type
a. For my own, I'm interest in latest technologies including AI and Robots
b. It's made by human with our intelligent brains.

(19) Level C Progression type
c. It shows how great human is.

(19) Level C Progression type
d. I played robots when I was a small kid. (C Level) EP

In addition to redundancy, the flow of ideas in in lower levels namely in C and B levels is usually abruptly coherence breaks triggered mainly by the absence of topical subjects (example 20-21), a grammatically inadequate use of subordinating structures (example 22b) and inadequate lexicons (example 22d). Another feature that contributes to coherence break in low level writing is the use of vague pronouns which, in the author’s view, stems mainly form a lack of awareness of the impact a distance between the referee and its pronoun referent has on textual coherence. In example 23c, it is hard for the reader to determine whether the “it” pronoun...
is refereeing to “development of artificial intelligence” or is a grammatically inaccurate referent to robots. The distribution pattern of coherence breaks in this study support Knoch’s (2007) findings that coherence breaks are more frequent in lower proficiency levels and are rare in advanced levels.

(20) Level C
Progression type

a. *The future* is very unpredictable.
b. *Humanity* takes sudden turns or leaps forward that,
c. *at first glance,*[] seems to be beneficial for us

d. *but[*] turn at to be our own grave

(21) Level C

a. *In the most jobs* [] is required the creativity of the human to do all better.

b. *[] Is very important in every work the decision that need to make for have a success company and this just can do humans.*

(22) Level B

a. *Secondly the widely using of automatical machines and robots* will gives us more opportunities to concentrate on creative thinking, innovation, imagining and issue solution thinking

b. *For we[*] could be unlocked by automation from repetitive tasks and aroles of production claim*

c. *we could solve those questions which we are puzzled for centuries.*
d. *Innovates today’s artisans, are developing new ideas for our society.*

(23) Level B

a. *I think that development of artificial intelligence* could improve human’s life more convenient than it was.
b. *For example, if robots* could work chores instead of people, people would concentrate on their own works

c. *In addition, it help increase ability of people in their workplace.*
To sum up the discussion of this section, the analysis of topical progression in read-argue EPT essays is in line with the EPT rubrics, as higher-level writers displayed better ability in developing the discourse topic through examples and reiteration. Specifically, the limited use of sequential progression in level B compared to other levels support the EPT rubric’s description of this level’s writing as lacking “development of ideas” (see Appendix A). Similarly, the infrequent use of sequential progression in level C corroborates with the rubric regarding the functional aspect of topical development that this level “argue with unclear examples/ or lack of examples” (see Appendix A). It also reinforces the rubrics that this level display “moderate repetition of the same expressions or sentence structures”. The results of topical analysis do not, however, support incorporating the same rubrics for both C and D levels. For example, the frequent use of cohesive devices in D level, the understanding of the discourse functions of sequential progression and the lack of repetition compared to C level do not support some of the descriptors of this level. In particular, the findings do not corroborate the description of this level that points to using limited number of cohesive devices, lack of examples and the “moderate repetition” of structures or expressions.

The findings for parallel and indirect parallel progressions corroborate previous research findings that even though repetition is used by both skilled and less skilled writers, it is redundantly used by low-level writers (Stotsky, 1986; Witt & Faigely, 1981). Overall, it can be concluded that the group differences in using parallel, direct parallel, and sequential structures sheds light on the role of reiteration as “the single most important form of cohesive tie” (Hoey, 1991, p. 9). The differences revealed by the topical analysis with respect to different placement levels aligns with Knoch (2007) conclusion that topical analysis can provide better discrimination of levels and hence, can be used to develop rating scales. The findings also
highlight the importance of introducing topics in facilitating connections of ideas. The next section discusses ISEs, important devices for introducing and connecting discourse.

**ISEs Frequency and Linguistic Realizations**

The results of the analysis of ISEs in this section discusses in two main parts the answer to the second question in this study regarding ISEs frequency and linguistic realizations. The first part reports the results of the descriptive statistics and ANOVA on the distribution of T-units initiated by ISEs across placement levels. The second section focuses on how placement levels differ in the distribution of different syntactic forms. To address this question, the proportion of each type of syntactic forms to the total syntactic forms will be first discussed. Accordingly, the placement levels will be compared in terms of the normalized frequency of the most prominent syntactic forms. The semantic functions of these syntactic forms will also be discussed along with illustrative examples.

With respect to differences in the mean use of ISEs, the descriptive statistics as shown in Figure 5.5 shows a relative increase of sentences initiated by ISEs as the placement level increases in both graduate and undergraduate levels. To examine whether the differences are of statistical difference, one-way ANOVA test was then conducted with the proportion of ISES as a dependent variable and the placement level as the independent variable. The results as shown in table 5.8 indicate that there is a statistically significant difference in the distribution of ISEs: $F(4, 87) = 3.747, p = .007 < .05$. A Tukey HSD post hoc test revealed that the graduate Pass ($p = .045 < .05$) and the D level ($p = .03 < .05$) use of ISEs was statistically significantly higher than their pass undergraduates’ peers.
Figure 5.6 The Mean Proportion of ISEs across EPT Placement Levels

Table 5.7 ANOVA Test for the Mean Proportion of ISEs across EPT Placement Levels

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>.548</td>
<td>4</td>
<td>.137</td>
<td>3.747</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3.182</td>
<td>87</td>
<td>.037</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.730</td>
<td>91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8 Tukey HSD Post Hoc Test for the Mean Proportion of ISEs

<table>
<thead>
<tr>
<th>(I) Level</th>
<th>(J) Level</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass (G)</td>
<td>Pass (UG)</td>
<td>.178*</td>
<td>.063</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pass (G)</td>
<td>- .015</td>
<td>.062</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pass (G)</td>
<td>.123</td>
<td>.061</td>
<td>.268</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pass (G)</td>
<td>.137</td>
<td>.062</td>
<td>.186</td>
</tr>
<tr>
<td></td>
<td>Pass (UG)</td>
<td>Pass (G)</td>
<td>-.178*</td>
<td>.063</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pass (UG)</td>
<td>-.193*</td>
<td>.065</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pass (UG)</td>
<td>-.055</td>
<td>.064</td>
<td>.911</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pass (UG)</td>
<td>-.041</td>
<td>.065</td>
<td>.969</td>
</tr>
</tbody>
</table>

*The mean difference is significant at the 0.05 level.
Having higher frequency of T-units initiated with ISEs in high level writing may not on the surface support some previous research findings that low proficiency writers overuse linking adverbials compared to proficient writers (Shaw, 2009). The explanation for this difference is that in the current study all syntactic forms (linking adverbials and other syntactic forms) that are only sentence initials are considered. Also, unlike most of the previous research in which the normalized frequency is based on the number of words in texts, the current findings are based on the number of the T-units. The findings that proficient writers tend to use more linking devices in creating textual cohesion, however, align with studies in which frequency was adjusted based on the number of sentences (e.g., Anderson, 2014). Overall, the findings of the current research suggest that not only the placement level but also the education level (graduate vs. undergraduate) impact the frequency use of ISEs. Having more exposure to academic writing, graduate level students have probably more awareness of the role ISEs play in developing textual cohesion. It should, however, be stressed that if there is any conclusion to be drawn from research on frequency use of cohesive devices is that the mere use of cohesive devices, as will be shown later, does not guarantee textual coherence and, hence writing quality. The frequency of ISEs is not a sufficient indicator of writing quality (Carrell, 1982). The range of syntactic forms, the appropriateness and the semantic meanings served by the ISES play even more important role in judging the writing quality.

The analysis of syntactic realization of the ISEs as shown by Figure 5.6 revealed that out of the total syntactic forms of ISEs, the most commonly used syntactic forms across levels are linking adverbials (43%), coordinated conjunctions (16%) prepositional phrases (19%) and finite clauses (11%). Single adverbs constitute about 8% and they are usually accompanied by other types of ISEs such as linking adverbials and finite clauses. Non- finite clauses are relatively rare
in all levels with only 3% of the T-units initiated by non-finite clauses. To examine how placement levels differ in the use of various syntactic forms, normalized frequency of the syntactic realization is calculated. Because they are usually used as inter-sentence connectors (connecting two main clauses within sentences) and they are usually accompanied by other syntactic forms such as linking adverbials, the frequency of coordinated conjunctions will be excluded from the discussion of normalized frequency of ISES’ syntactic forms. Also, and due to their rare occurrence as ISEs in the current corpus (only 3%), non-finite clauses are also excluded from further discussion.

The normalized frequency for the various syntactic forms of ISEs as illustrated by table 5.9 suggests similar distribution of linking adverbials, prepositional phrases and adverbs across levels. Interestingly, the use of finite clauses as ISEs is relatively frequent in all levels except in low level B.
Table 5.9 *The Normalized frequency of Syntactic Realizations of ISES per 100 T-unit*

<table>
<thead>
<tr>
<th>Status</th>
<th>Level</th>
<th>Linking adverbials</th>
<th>Finite clauses</th>
<th>Prepositional phrases</th>
<th>Adverbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate</td>
<td>P</td>
<td>24.5</td>
<td>24.5</td>
<td>19.9</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>21.0</td>
<td>16.0</td>
<td>20.5</td>
<td>22.4</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>P</td>
<td>16.2</td>
<td>19.1</td>
<td>20.5</td>
<td>22.4</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>18.6</td>
<td>25.5</td>
<td>19.9</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>19.7</td>
<td>14.9</td>
<td>19.3</td>
<td>16.4</td>
</tr>
</tbody>
</table>

Considering the assumption that low level writing would reflect more aspects of spoken language whereas high level writing would display more resemblance to academic writing, the findings with regard to the frequency of finite and PP adverbials appear to contradict, on the surface, Biber, Gray, and Poonpon’s (2011) and Biber and Gray’s (2010) contention regarding the distribution of these features. Being more prevalent in spoken language, finite clauses are expected to be of higher frequency in low compared to high level writing. To the contrary, being a prevalent feature of academic writing, PP adverbials are expected to be of a lower frequency in low level writing.

One reason for such a difference in the findings, in the author’s view, is that the results of Biber and his colleagues were based on the frequency of prepositional adverbials and finite clauses regardless of their position in the sentence. However, in this study, only PP adverbials and finite clauses in initial sentence positions are considered. Another potential reason might be related to the relatively lower range of writing proficiency levels represented in the EPT written responses compared to the corpus of professional writing used in Biber, Gray, and Poonpon’s (2011) and Biber and Gray’s (2010). It is possible that the writing level of the EPT students has not reached the point where the use of these features begins to decline. The relative frequency of
finite clauses in the Pass Level and the frequency of PP adverbials in low levels prompts further investigation of their functions in an initial sentence position. The next section will focus on the discourse functions of linking adverbials, PP, finite clauses and adverbs across levels.

**Discourse Functions of ISES**

Linking adverbials constituted the largest syntactic forms used across levels in initiating T-units. The main discourse function of linking adverbials is to connect ideas rather than adding more information (Biber, Conrad & Leech, 1991). Linking adverbials in the current corpus are usually accompanied by other syntactic forms including prepositional phrases, finite clauses and adverbs. Each of these syntactic forms can be used to express different meanings. Prepositional phrases can be used as circumstantial adverbials (e.g., *at this time*), stance (i.e. *in my view*) or as a linking device that connects given with new information (e.g., *with all these unanswered questions in mind, we need to reshape our thinking of how this issue should be approached*).

When used in initial position, adverbial clauses which include finite clauses, as Biber et al. (1991) noticed, carry two main discourse functions: 1) maintaining English information structures either by containing given information or by setting the scene for the next stage of discussion, or 2) framing subsequent events, a function common in fiction discourse. When considering how initial adverbials in general are used in low level compared to high level writing, it is noticed that the low -level writing display problems in understanding the discoursal functions of initial adverbials in academic writing, not to mention the difficulties in constructing grammatically accurate complex structures with dependent and independent clauses.

Examination of how linking adverbials (LAs), prepositional phrase (PPs) and finite clauses (FC) are used in B and C compared to D and Pass levels reflects differences in learners’ ability of appropriate use of ISEs in maintaining textual coherence (Note ISEs are in bold). Low
level writers particularly in level B rely excessively on linking adverbials. Initiating short sentences, LAs are inadequately used to jump from one idea to another and to add more information rather than connecting two units of discourse (see examples 24b, 24c, 25c and 25e). The use of linking adverbials reflects often a lack of awareness of the role of LAs in establishing a connection between a previous and a forthcoming discourse unit. Example 25c exemplifies such lack of connection as the use of “in another words” defies readers’ expectations of presenting a rewording of a previous discourse. When finite clauses are used, they usually introduce new events - a style more reminiscent of spoken modes of communication - through the use of mainly hypothetical conditions, as illustrated by Examples 25f, 26c. This level exhibits also difficulty in constructing subordinate structures (see example 27b) particularly when using because, though and although:

(24) Level B
a. *Robots are created to do many jobs to make life easier and comfortable.*
b. *However, having robots into humans’s fields will impact our world in a bad way.*
c. *In sum, there are disadvantages to the world with robots having jobs.*

(25) Level B
a. *There is no doubt that robots could save time if they occupy our jobs.*
b. *Virtually we only need to do is "innovation"*
c. *In another word all the humanity will have the same job*
d. *In this circumstance, the finance will crap people just thinking all the useless things*
e. *Finally, the society will face destroy*
f. *What is more if the robot have some error and make the tragedy. Who is the sponsor will be a cracial issue.*
(26) Level B

a. *In addition,* it help increase ability of people i their workplace. A lot of people are usually stressed out their work loads.

b. *Therefore,* improvement of Ai technologies will take advantages of human life

c. *Moreover,* if one family have a patient, they really need to help from others.

(27) Level B

a. *Secondly* the widely using of automatical machines and robots will gives us more opportunities to concentrate on creative thinking, innovation, imagining and issue solution thinking

b. *For we could be unlocked by automation from repetitive tasks and aroles of production claim*

c. *So* we can be allow to do superior work with the help of robots.

C level shows less dependence on LAs and more attempts of using various syntactic forms mainly finite clauses. The use of different subordinators in level C particularly *though, because, as* and *when* reflects awareness of their discourse function in initial sentence position. The use of LAs as well as finite clauses is, however, mostly inappropriate semantically and grammatically. In line 28f, the attempt to link previous argument with the new proposition failed due to the use of fragment sentences (dependent clauses without independent clauses). Similar semantic disconnection between two consequent ideas is noticed in the use of “however” in example 28h.

(28) Level C

a. *Some people state that robots are not good for humans's development in the future years, because they take place the humans's jobs and lives.*

b. *However,* others say human are still the supervisor of the world.
c.  From my perspective...
d.  First, because people have different unique talents, there are many people in the world.
e.  People create the robots, we can control them.
f.  Even though they can do some jobs which human can't do.
g.  However, when they complete the jobs, they still need to human's conduct.
h.  However, everyone in the world has the different talents and thoughts.

Lower levels of C and B show also similar frequency use of prepositional phrases as higher levels. Nevertheless, the examination of the semantic functions of PP adverbials revealed that the circumstantial (see example 29-30) and stance (example 31) meaning are more dominant in low level writing. The use, in many cases, is inappropriate. The use of prepositional phrases particularly those of manner contribute to the disrupt rather than the cohesion of the information flow, as exemplified by "in many ways" in example 29. Prepositional phrases of stance are mainly epistemic adverbials reflecting either view point (see example 31) or certainty and doubt as with "for some reason" in example 29.

(29)  Humans will become less active and unproductive. In many ways, but for some reason they will help too. (Level B)

(30)  In the manufacture, robots could do their job in 24-hour and 7 days per week, with low fee because they just need electricity (Level B).

(31)  As we know, robots take more and more jobs in our daily lives, in my opinion, however, it is not beneficial to humanity for two reasons. (Level B).

Unlike lower levels, Pass and D level students show more proficiency in alternating between different syntactic forms to introduce their arguments. Generally, the use of linking adverbials reflect appropriate knowledge of their cohesive functions. Finite clauses are used to
link given and new information as shown in example 32-34, despite some lexical and grammatical errors in Pass undergraduate and D level writing. Similar function is served also by the initial prepositional phrases as shown in examples 35-37 and with the phrase “with the development of technology” in example 38. Prepositional phrases in initial sentence position reflect also other meanings that function as preparatory phrases to a new discourse unit. This include stance as in “from my point” in example 38 and circumstance “(example 39). Prepositional phrases reflecting stance are mainly epistemic (expressions of the extent of truthfulness of the forwarded information) with two major meanings: personal view and source of knowledge. For example, in excerpt 40, the forwarded proposition is initiated by a reference to the reading passage.

(32) **Though** the second text points out the fact that the iPad is expensive for students, I'm sure the fast growing technology is going to produce very efficient and low cost tablets exclusively for students, in the days to come (PG)

(33) I personally believe that benefits of use of iPads in classrooms, do outweigh their challenges. **Though** there are a few negative effects of iPad use on students, I believe use of iPad is one of the best way for students to keep up with the growing technology (PG)

(34) **Above all, although the robot** take a lot of convenients for people such as the cleaner and sweeper. The disadvantages are more than the advantages of the robots. (PU)
(35) **these digital machines are expensive.** With rising education costs, it makes no … (PG)

(36) **We have so much into these things that we even started compromising with our social life as well.** Apart from this, Robots made us and will make us more in future so free that we even forgot the meaning of life. (PU)

(37) **From all the above reasons,** I strongly support and recommend the use of iPads in schools. (D Level)

(38) **With the development of technology,** more and more smart devices are invented for convenience. Ipad is one of such inventions, which leads to an argument whether the school should offer ipad study. From my point, I think ipad study has more advantages than disadvantages. **To begin with,** ipad is more convenient for students to use than textbooks. **Firstly,** it makes the backpacks much lighter. (PG)

(39) **In the new era of technology,** a lot of devices are built to improve the quality of life and solve or erase the problems. **Specifically with advance in computer technology,** a lot of progress gain using new technology. (D level).

(40) **Can Robots write a song? In another word, according to Reading text 1, robots can do a lot of jobs better than human it is because computer programs are good at doing the same thing.** (PU).
Unlike finite clauses and prepositional phrases, the use of adverbs is relatively less frequent, and the meanings conveyed are similar across levels. The two major meanings postulated by the adverbs in the current corpus are time (e.g., now, nowadays, recently, today) and epistemic stance that expresses perspective (e.g., personally) or doubt and certainty (e.g., perhaps, absolutely, actually).

Overall, the distribution of the various syntactic forms of ISEs and their discourse functions across placement levels implicate that more proficient writers portray more awareness of the role ISEs play in structuring their arguments and also more competence in utilizing different syntactic realizations of ISES in connecting discourse. The findings with regard to the semantic functions employed across levels is in line with the research findings on differences between spoken and written language as well as research findings on aspects of proficient writers. The use of place adverbials in the form of prepositional phrases is common in low level writing. This suggests that the writing at this level is influenced by spoken language given that place adverbials are found to be more common in oral presentation (Rowley- Jolivet & Carter-Thomas, 2005). Also, the predominance of if conditional structures in low level writing of B in particular reflects resemblance to spoken language as if clauses of condition are found to be of high frequency in conversation (Biber et al.,1999). To the contrary, the use of finite clause in high- level writers of Pass and D aligns with the findings of Schleppegrell (1996) that non- ESL writers tend to use because to cohesively connect previous and subsequent discourse.

Conclusion

The above discussion examined how ESL learners of various placement levels maintain topical development in written discourse. The analysis of the progression types revealed that both the writing of Pass and D Levels portrayed more elaboration of the topic in hand and less redundancy. This was mainly reflected in the use of more sequential progressions, indirect
parallel and less instances of coherence breaks than their lower level peers in C and B levels. The analysis of ISEs’ frequency and their linguistic realizations highlighted differences in the frequency, linguistic realizations, and the discourse functions of ISEs between higher and low placement levels. High level writers in Pass and D levels tend to utilize various forms to introduce and link their propositions. Lower level writers mainly in level B displayed not only over dependence on linking adverbials but also a lack of knowledge of their discoursal functions. The discourse analysis of the topical progression, the frequency and discourse functions of ISEs in the EPT writing section provides support for the validity argument for the EPT. The results back the assumption that ESL students’ performance on the EPT test reflects differences in their linguistic complexity. The findings align with the level descriptors of Pass and B level. In line with the EPT rubrics, the pass level found to use a variety of cohesive devices and to display appropriate organization with little digression from the main topic. Level B, on the other hand, shows limited ability of developing an argument through examples and details, more instances of coherence breaks and a limited use of cohesive devices compared to other levels. Also, and in line with EPT level descriptors, B level portrayed more use of If conditional in initial sentence position compared to other levels.

While the discourse analysis of topical development and ISEs tends to align with the level descriptors of Pass and B levels, they do not support some of the rubrics of C/D levels. Even though the EPT rubrics do not distinguish between C and D levels, the discourse analysis of topical development and the distribution of ISES revealed substantial differences between the two. While the EPT level descriptors that point to a lack of examples and a redundant use of expressions found to match discourse features of C level writing, they are in conflict with the discourse analysis findings of D level. Contrary to what is stated in the rubrics, the D level shows
less redundancy and more frequent use of various linking devices and more awareness of their meanings and discourse functions than its C counterpart.

**Implications and Limitations**

Investigating organizational patterns of discourse for ESL students of various placement levels sheds light on the intricate relationships among lexis and grammar in discourse development. Reinforcing and elaborating on specific topics requires knowledge not only of the general essay structure, but also an awareness of the functions of lexical strategies and of grammatical devices in facilitating topical progression, in general, and in projecting opposing views in particular.

This conclusion is supported by the analysis of how writers at different levels utilize the discourse functions of ISEs in structuring their essays. Syntactic accuracy of using certain structures (e.g., finite clauses and prepositional phrases) does not equate semantic appropriateness. Accordingly, the study of grammar in writing classes for lower level students should not only focus on the accuracy of using individual aspects, such as tenses, transition words, and subordinate structures, but also emphasize how syntactic structures can be adapted to enhance the rhetorical persuasiveness in argumentative writing. At higher level writing where learners show more awareness of the discourse function of grammatical structures namely adverbials, the accuracy of constructing complex structures with dependent and independent clauses should be emphasized.

The results of this study are not without limitations. The number of texts representing each placement level is relatively small. Also, since LSD post hoc test does not account for Type I error or family-wise error rate (rejecting the null hypothesis when it is actually true), the results of sequential should be perceived as informative but not conclusive. Also, the results regarding cohesive devices are based on syntactic forms used in initial sentence positions. It would be more
insightful to compare how certain structures such as finite clauses, propositional adverbials are used in initial as well as median and final sentence positions. Moreover, the definition of ISES was restricted to elements preceding the first grammatical subject be it topical or non-topical, it would be far more informative to further examine how non-topical subjects (e.g., *There is no doubt that the new plan would impact the way they work*) that usually precedes topical subjects contribute to the cohesiveness and the management of information structure in a text.
CHAPTER 6. THE IMPACT OF EPT USE ON ESL TEACHERS AND STUDENTS

Investigating the Consequences of EPT Placement Decisions (Ramification Inference)

With respect to placement tests such as the writing section of the EPT, “an ideal course” for the courses that students are placed into is expected to focus on the type of competencies, skills and abilities that students have problems with or are not mastered sufficiently (Kane, 2006). The claim underlying the ramification inference, then, is that the consequences of using the EPT writing assessment and of the decisions that are made are beneficial to stakeholders namely learners and teachers. (Bachman & Palmer, 2010). Two warrants underlie this claim with three types of backing that utilize qualitative and/or quantitative methods. Relevant to learners, the first warrant is that test takers will be able to recognize their strengths and weaknesses in writing and work accordingly. This assumption can be backed by a survey study of students’ self-assessment of their abilities before and after ESL writing courses. Pertinent to teachers, the second warrant is that the EPT writing assessment helps promote good instructional practices in ESL writing classes. This warrant can be backed by a mixed method research (questionnaires, interviews) with ESL teachers to evaluate teachers’ satisfaction with the curriculum objectives and materials, the compatibility between teachers’ opinions about characteristics of academic writing, common problems in their ESL students’ writing and the respective instructional practices in ESL writing classes.

Investigation of the ramification inference in terms of the impact on teachers and learners requires an overview of the related literature. Therefore, the first section of this paper discusses research findings on the relationship between instructors’ beliefs and instructional practices. The second section focuses on self-assessment as a validation tool in language assessment. This will be followed by the questions addressed in this study, the methodology and findings. The paper
concludes with a discussion of implications for ESL writing particularly in the EPT context. Limitations and avenues for future research will also be highlighted.

**Teachers’ Perceptions and Instructional Practices**

Investigating the alignment between teachers’ beliefs and teaching practices in light of research findings is of a pedagogical importance particularly for in-house tests where the major goal is to insure learners’ readiness for university level writing. Emphasizing the role of teachers’ perceptions in any curriculum innovation, Shi and Cumming (1995) stressed that “personal qualities of teachers’ conceptualizations of their work are particularly vital in considering initiatives for curriculum change in the domain of second language writing” (p. 104).

It is expected that teachers’ beliefs influence their focus of teaching and their perceptions of error fatality (Richards & Lockhart, 1994, Farrell & Lim, 2005). This, however, is not always true, as research has revealed that what teachers believe to be important may not always be reflected in their classrooms’ practices (Farrell & Bennis, 2013, Phipps & Borg, 2009). In many cases, teachers may not be aware of their theoretical orientations let alone potential discrepancies in their teaching focus (Farrell, 2008). Aside from the syllabi assigned by institutions, various factors contribute to shaping teachers’ thoughts and knowledge about ideal practices including “their training, their teaching experience, or … their own experience as language learners” (Richards and Lockhart, 1994, p. 34). This highlights the importance of raising teacher’s awareness of research-based characteristics of academic writing and insuring a convergence between such beliefs and actual classroom practices. It is generally accepted that improvement in schooling, including ESL writing development, depends to a large extent on what Dobson and Dobson (1983, p. 21) called “beliefs-practice congruency”. By necessity such alignment should be shaped by research findings of characteristic of academic writing rather than intuitive beliefs about language. The need for surveying ESL teachers’ perspectives and classroom practices is
then justified by the importance of alignment between research findings with respect to features of academic writing and ESL classrooms practices in general and the congruency between teachers’ beliefs and their writing instruction in particular.

There is a general agreement among teachers and researchers that there is a positive correlation between advanced vocabulary use, syntactic complexity and appropriate development of topic and the judgement of writing proficiency (e.g., Gue, Crossley, & McNamara, 2013; Grant & Ginther, 2000). Despite the general agreement among researchers and teachers on the significance of these three aspects, research pointed to an imbalance in covering these areas in ESL classrooms. For example, even though lexical complexity is perceived to be important in academic writing, Folse’s observations of 50-hour ESL classes (2004, p. 9) revealed that “vocabulary is not systematically covered in most curricula” and that daily writing activities do not contribute to the expansion of vocabulary knowledge and were only introduced when needed. In fact, the lack of focus on vocabulary may affect how students of different backgrounds benefit from ESL courses. Derived largely from Graeco-Latin origin (more than 80%), Academic Word List or AWL which constitutes more than 10% of vocabulary in any academic text put speakers of Romance languages at an advantage over other L1 learners (Coxhead & Byrd, 2007).

In their attempt to measure the extent with which courses of academic writing have achieved their goals of preparing ESL students for university level content courses, Leki and Carson (1994) surveyed international students who had or were at the time of the study taking ESL writing courses in two US universities. It was found that the need for expansion vocabulary in particular and for speed processing of grammatical structures were the primary needs that ESL courses fall short from achieving. Learners voiced frustration over the time it takes them to access the proper lexical and grammatical forms in their academic written assignments. Accurate
grammatical use, however, was voiced to be of less important in their content courses. Similar findings were reported by Silva (1992) study in which difficulties of accurate and precise use of lexical expressions and grammar were the main concerns for ESL learners.

As for syntactic complexity, there is a scant research on how indicators of syntactic complexity are perceived and addressed in actual ESL classrooms compared to the vast majority of research on the frequency of use of such features in ESL writing across different proficiency levels. The observation of ESL classrooms has pointed to an emphasis on accurate use of grammatical structures (Folse, 2004) rather than what and how grammatical features contribute to the complexity level of a text. In a survey of 60 ESL teachers with a mean experience of 14 years in the US and Puerto Rico, Eisenstein-Ebsworth and Schweers (1997) found that teachers’ definition of grammar is restricted to syntactic elements at the sentence level with no reference to discourse level. Understanding of how a cluster of grammatical features are rhetorically used in academic texts is necessary, yet, it is hampered by the general focus either by teachers or textbooks or even both on single features mainly verbs and types of tenses (Bunting & Byrd, 2008).

Corpus-based studies, however, have concluded that academic writing is characterized by dense noun use and long complex noun phrases, along with other features such as the use of adverbials for organizational functions (e.g., in the next section) and non-finite clauses (Biber, 1988; Biber et al., 1999). The prevalence of long complex noun phrases in academic texts may explain ESL learners struggle with processing grammatical structures in university level content courses, as was reported by Leki and Carson (1994). Accordingly, ESL writing courses are expected to promote a comprehensive approach to grammar that considers how clusters of
dominant grammatical features in academic discourse are used and utilized rhetorically by proficient writers.

The capacity of developing well organized ideas, unlike lexical and grammatical features, have has been given priority in judging both ESL and native speakers writing, particularly from the perspectives of university professors in content area (Leki and Carson, 1994; Song & Caurso, 1998). The emphasis on discourse organization and content has also been reflected in retrospective comments made by experienced ESL writing teachers. In their analysis of weekly interviews with experienced ESL writing teachers in a Canadian university over their conceptualizations of their pedagogical practices, Shi and Cumming (1995) found that the theme of content and organization dominated 47% of their reflections.

To sum up, the discussion above indicates that there could be a disconnection between what teachers believe to be important in writing development and their actual instructional practices. Also, the brief literature review on instructional practices of ESL teachers suggests a potential discrepancy between the focus of writing courses and learners’ voiced difficulties. There seem to be a more focus on accuracy and organization in ESL classes at the expense of the expansion of vocabulary and grammatical knowledge relevant to academic writing.

In fact, despite the semi consensus among researchers on the importance of vocabulary, syntax and discourse level features, the findings from research on ESL classroom practices suggest that the three aspects are not equally addressed in classrooms of writing. Teachers tend to place more emphasis on the general structure of the text (e.g., supporting thesis statement with examples, introduction, body and conclusion) rather than the rhetorical functions of grammatical structures.
The reciprocal influence of beliefs about quality writing and instructional practices highlights the importance of exploring teachers’ beliefs and their relatedness to teachers’ pedagogical focus in ESL writing classes. In addition to exploring the alignment between ESL teachers’ perceptions and instructional practices, investigating ESL learners’ perceived improvement and difficulties in various aspects of academic writing is needed for better understanding of the ramification inference for the use of the EPT writing section on both teachers and learners. The next section, hence, discusses self-assessment as a tool for investigating ESL learners’ perceived development in writing.

**Self-Assessment of Writing Abilities**

Self-perceptions of strengths and weaknesses have been widely used in assessment research. Several terms are used in the literature to refer to the same concept, including self-assessment, self-evaluation, self-concept, self-rating or self-perceived knowledge. Two major types of self-assessment can be identified depending on their purposes: performance-oriented and development-oriented self-assessment (Bachman, 2000). In the first type, learners’ performance is assessed once and it is typically used for diagnostic purposes, placement decisions, or in correlational studies to provide evidence of a concurrent validity of a criterion under question (Enright, Bridgeman, Eignor, Lee, & Powers, 2008; Johns, 2006; LeBlanc & Painchaud, 1985; Li, 2015; Ross, 1998). Unlike performance-oriented self-assessments which are administered once, development-oriented self-assessments are used to detect patterns of improvement in learning over an extended period of time (Dornyei, 2001).

The increasing use of self-assessment by educators and researchers has been accompanied by further research on the reliability and validity of peoples’ self-evaluation of their abilities. Reliability refers to the consistency of scores obtained from self-assessment tools (Ross, 2006). Reliability of an assessment tool can be estimated through comparing results
across tasks and times. In a meta-analysis of research using self-assessment in classroom settings, Ross (2006) concluded that self-assessment measures are reliable and consistent across items, tasks and short time periods. The review shows that the results from self-assessments tend to be less consistent across long time periods, particularly with children.

The validity of scores is verified by examining the correspondences between self-assigned scores and other objective measures, e.g., standardized tests or teachers’ evaluations (Ross, 2006). Examining the validity of self-evaluation scores, Cassidy (2007) compared the self-assigned grades of 160 students in a faculty of health science with the scores assigned by their tutors. A positive but relatively low correlation of .25 was found between the two types of evaluation. The study also found that 56% of students underestimated their ability whereas 40% overestimated their performance. Overestimation was most likely to occur when the self-assessed scores contribute to students’ overall scores in a course (Boud & Falchikov, 1989). Agreement with teachers’ rating is reported to be high when learners are familiar with the content or subject area that they assess (Ross, 1998).

In a meta-synthesis of 22 meta-analytical studies of the relation between self-assessment and objective performance measures, Zell and Krizan (2014) found a moderate mean correlation ($M = .29, \text{SD} = .11$) between self-assigned scores and other objective measures (e.g., standard tests, teachers’ assigned grades). Zell and Krizan’s study also concluded that the strength of correlation between self-evaluation and performance measures is influenced by the scope and the type of the domain, and the simplicity and the familiarity of the tasks. The correlation between self-evaluation and performance criteria is stronger with more specific domains, as well as with less complex and more familiar tasks. The accuracy of self-assessment was also found to vary across domains. Compared to self-assessments of vocational skills and sports ability, research on
self-assessments of language learning was found to have the highest reported effect size of \( r = .63 \), Zell & Krizan, 2014). The researchers attributed the difference in the accuracy of self-assessment across domains to the nature of language competence domain, where unlike other fields, learners receive ubiquitous feedback on their performance in different language skills.

Despite some downsides of using self-assessment, e.g., overestimation or underestimation, the discussion of research findings of self-assessment as a measure of performance suggests the utility of this tool in language learning settings, particularly with adults who are familiar with the domain being assessed. The research also suggest that the results of self-assessment can be more reliable when used in a non-threatening environment, i.e., when the self-assessments are related to scores not contributing to students’ grades in a course.

Considering the context (language learning and adult students) and the purpose of the current study, self-assessment is considered a viable cost-efficient tool - in terms of time and resources-to evaluate ESL learners’ perceived development in writing. Triangulating learners’ perspectives of their strengths and weaknesses with the perspectives of ESL teachers in the same educational context strengthens the validity of findings regarding the impact of the EPT placement decisions on both learners and teachers. In addition, an examination of the extent of congruence between ESL teachers’ beliefs about quality writing and their instructional practices contributes to our understanding of factors that may impede or enhance students’ writing development. Bearing this in mind, ESL students’ self-assessment of writing ability and ESL teachers’ perspectives of quality writing will both be surveyed to investigate the ramification inference for the EPT writing section.

**Purpose of the Study and Research Questions**

Research findings of aspects of academic writing are important for shaping our understanding of what should be included and emphasized in ESL writing courses. Nevertheless,
it is not expected that all ESL teachers have the same background knowledge, nor that they have the time to keep updated with the findings of research. There is, then, a need to survey teachers’ views of what researchers have found to be signals of proficient academic writing and to further examine the relation between teachers’ beliefs of proficient academic writing, the common problems in their ESL students’ writing, and the focus of their pedagogical practices. Triangulating ESL teachers’ perceptions of quality writing and of strengths and weaknesses in their students’ writing with ESL learners perceived improvement in writing is important to understand potential factors affecting improvement in ESL courses. In light of these concerns, the following questions are investigated:

1. To what extent do ESL teachers’ instructional practices align with their beliefs about good academic writing?
2. To what extent do ESL learners demonstrate self-assessed improvement in linguistic ability after ESL writing courses?

**Methodology**

To answer the questions of this study, surveys and interviews with ESL teachers of writing and ESL students taking ESL writing courses of 101B, 101C and 101D were utilized. Further details on participants, instruments and data analysis are presented below.

**Participants**

**ESL teachers**

ESL teachers who were teaching or have taught ESL writing courses at ISU (specifically, English 101B, 101C, and 101D) were invited to participate in filling out a survey on teachers’ perceptions of important language skills in writing. A total of 9 teachers completed the survey (pseudonyms are used). Four teachers (3 females and 1 male) agreed to participate in semi-structured interviews. All the teachers had at least three years of experience teaching writing.
Table 6.1 *Demographic Information for ESL Teacher Participants*

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sue</td>
<td>Female</td>
<td>Master degree</td>
</tr>
<tr>
<td>Diana</td>
<td>Female</td>
<td>Master degree</td>
</tr>
<tr>
<td>Nancy</td>
<td>Female</td>
<td>Master degree</td>
</tr>
<tr>
<td>Dan</td>
<td>Male</td>
<td>PhD. degree</td>
</tr>
</tbody>
</table>

**ESL students**

Students who were enrolled in ESL writing courses (English 101B, 101C, and 101D) during Fall 2018 were invited via email to participate in the survey, administered through Qualtrics. Sixty students completed the first survey and only forty-seven completed the post survey. The forty-seven participants were 24 males and 23 females from different L1s: Chinese (35%) followed by Malay (30%), Hindi (19%) and other L1s (30%).

**Instruments and Procedures**

Two surveys and two semi-structured interviews were used to answer the above stated questions. The first survey was distributed to ESL teachers (see Appendix D) and was followed by a semi-structured interview (see Appendix E). The second survey (see Appendix F) was administered to ESL learners to examine their self-assessment of their writing ability before and after taking English 101B, 101C, and 101D at ISU. The survey of ESL learner was also followed by a semi-structured interview (see Appendix G) with ESL students. Before conducting the research, the approval of Office of Responsible Research (IRB) at ISU was obtained for ESL teachers’ (see Appendix J) and ESL students’ (see Appendix K) surveys and interviews. The present study is labelled under the exempt form of research because, based on IRB rules, it meets the federal requirements of exemption which include using surveys and interviews in which percipients’ identities are kept confidential (pseudonyms are used).
The ESL teachers’ questionnaire (see Appendix D), which was built and distributed through Qualtrics online survey platform, has 36 Likert scale items and two open-ended questions. The scale items address teachers’ perceptions of what constitutes critical skills for academic writing and should accordingly be reflected in ESL writing instruction practices. The survey has three major sections: 1) linguistic skills that are emphasized in teaching practices; 2) teachers’ perceptions of indicators of proficient academic writing; and 3) aspects of writing that ESL students are struggling with. The first section has 15 items answered on a Likert scale of six ordinal responses coded from 1 (never) to 6 (always). The six level responses aimed at establishing the frequency with which writing activities targeting vocabulary knowledge, syntactic complexity aspects and discourse development are addressed in ESL writing classrooms. The second and the third sections use six agreement levels ranging from strongly disagree coded as 1 to strongly agree coded as 6. The second part has 11 items addressing teachers’ perceptions of the importance of certain aspects related to vocabulary, syntactic and discourse complexity in academic writing. The third section explores teachers’ views of problems related to these three areas in students’ writing. The questionnaire has also 2 open-ended questions, one of which asks teachers about activities that they believe enhance students writing. The second question solicits any additional comments. The questionnaire was distributed to ESL teachers via emails at the end of 2017 Spring semester. Teachers were explicitly instructed to fill in the questionnaire based on the current or the most recent ESL writing course that they have taught. The questionnaire also included demographic information about gender, number of courses taught, current course being taught and years of experience in teaching ESL writing. To increase the rate of participation, teachers were not required to write their names. ESL teachers were also invited to participate in semi structured interviews in order
to explore issues mentioned in the survey more fully. The interview questions (see Appendix E) were designed to elicit information about the teachers’ satisfaction with the EPT writing placement decisions, their impressions on students’ improvement and the appropriateness of course materials, their beliefs about aspects pertinent to academic writing and the problems commonly noticed in their learners’ writing and how they are addressed.

The second survey targeted ESL students’ self-assessment of their writing ability at the beginning and toward the end of ESL writing courses. The pre- and post-surveys (see Appendix F) have 19 Likert scale items ranging from strongly disagree coded as 1 to strongly agree coded as 6. The survey addressed three major components of writing ability: vocabulary knowledge (items 1-8, 16), syntactic knowledge (items 9-14) and discourse level knowledge (15, 17, 18, 19). Four items (1, 4, 7, 9, 19) were reverse coded items. The post survey has, in addition to the previous items, another section that explores participants’ perceptions of the frequency of instructional practices addressing vocabulary, syntactic complexity and discourse development. This section has 11 Likert scale items (items: 20-31) ranging from never to always. The survey has also one optional open-ended item in which participants are asked about class activities they think would help them in improving their writing skill. To ensure the clarity of the survey statements, three PhD students and two professors in Applied linguistics were invited to comment on the appropriateness of the items for ESL students. The survey was also piloted to ESL students in the last two weeks of the Spring 2017. The piloting revealed no problematic issues with the items or questions.

After obtaining the IRB approval and the permission from the English department, the ESL students’ survey was distributed on the second week of Fall 2018 semester and the link remained active for a week. Students enrolled in 101B, 101C and 101D were invited to
participate via Qualtrics email. In order to increase the rate of participants, ESL teachers were also invited by the researcher to post the survey on Moodle. Students were also offered $5 cash as incentives. The internal consistency estimates of the survey calculated using Cronbach Alpha found to be acceptable at $\alpha = .82$. The internal consistency estimates of lexical ($\alpha = 0.68$), syntactic ($\alpha = 0.68$) and discourse ($\alpha = 0.69$) sub-components were also found to be acceptable.

The post-self-assessment survey was then distributed on the fourteenth week of the same semester and the link remained active for a week as well. To encourage participation, each participant was offered $15$ cash. To triangulate the results from the survey, another email invitation for a semi-structured interview was sent to those who participated in the survey. Four participants (all female) volunteered to participate. The interviews were scheduled to be about 10 minutes. Before the interviews, the researcher walked each participant through the terms of consent form (see Appendix H) which emphasizes that their identities will remain anonymous. The interview questions (see Appendix G) were set to address 1) students’ overall satisfaction with the placement decisions of the EPT writing section, 2) their perceptions of their overall writing ability after ESL writing courses and the specific areas in which they perceive development i.e. lexical, syntax and discourse aspects and finally 3) the problems they perceive to be common in their writing.

**Data Analysis**

To answer the first question regarding the convergence between ESL teachers’ beliefs of quality writing and their instructional practices, descriptive statistics will be used to analyze the survey of teachers’ perceived importance of certain writing skills and to describe the frequencies of class activities targeting these skills.

To answer the second question regarding the ESL learners’ self-assessments of their writing development, descriptive statistics and one-way within-subjects repeated ANOVA were
used to compare students’ self-assessed writing ability before and after their ESL writing course. The semi-structured interviews with both ESL teachers and students were transcribed and analyzed for common themes and patterns. In other words, themes addressing vocabulary, syntax and discourse level aspects as well as the overall satisfaction with the EPT placement decisions and ESL courses were identified in each transcript.

Results and Discussion

ESL Teachers

1. To what extent do ESL teachers’ instructional practices align with their beliefs about good academic writing?

The ESL teachers’ questionnaire examined the frequency of instructional practices at three levels of linguistic knowledge: vocabulary, phrasal and clausal grammatical complexity, and discourse level using a Liker scale with six levels indicating frequency ranging from never coded as 1 to always coded as 6. The mean and standard deviation of questionnaire results were calculated to reveal participants’ general attitudes to each linguistic aspect. Table 6.2 illustrates instructional practices arranged from the most to the least frequent (for space consideration, statements are shortened, full statements are provided in appendix D). At the level of vocabulary, the vocabulary instruction is addressed in ESL writing classrooms, though not on frequent or systematic manner. The means reflecting the frequency of vocabulary-related instruction range from \( M = 2.89 \) to \( M = 4.22 \) on a six Likert scale. As shown in Table 6.2, encouraging students to avoid repetition seems to be a highly frequent practice with a mean of 4.8 and with the majority of responses ranging between very frequent (44%) to always (22%). Similarly, encouraging students to use newly learned words in their writing seems to be a common practice \( M = 4.33 \) with 44% of teachers emphasizing that it is an “always” practice. Despite the emphasis on avoiding repetition and the use of new words, the frequency of instructional
practices that targets vocabulary expansion such as discussion of synonyms and antonyms seem not to be a common practice in ESL writing classes. The discussion of antonyms is the least common practice with a mean of 2.89 and 89% of responses indicating the frequency instruction on these topics as rarely or sometimes. Synonyms, words with similar meaning and different formality status and family members of words are also not frequently introduced (M = 3.44 to 3.67) respectively.

As for frequency of class activities related to syntactic complexity, particularly complexity of noun phrases, the survey showed that the teachers do direct learners’ attention to the complexity of noun phrases. The frequency of syntactic complexity related activities is, however, relatively modest ranging from (M = 3.22 to M = 3.44) on a 6-point Likert scale. With regard to discourse targeted practices, ESL teachers seem to devote much time to activities related to the development of ideas through examples and details (M = 4.00, SD = 1.32) with less focus on how writers begin their sentences (M = 3.56, SD = 1.67). Discussion of how grammatical structures can be used to maintain connectivity and how can they be utilized to reflect attitudes are, however, relatively infrequent practices ( with M = 3.1 and M = 3.00 respectively.
Table 6.2 *Frequency of Different Writing Related Practices in ESL Classrooms based on ESL Teachers’ Survey (1 = Never, Always = 6)*

<table>
<thead>
<tr>
<th>Linguistic Component</th>
<th>Class Practices</th>
<th>Item Mean</th>
<th>Item SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>Avoid repetition of the same words.</td>
<td>4.89</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>Encourage the use new words in their writing.</td>
<td>4.33</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>Attention is drawn to different forms of a word.</td>
<td>4.22</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Students discuss family members of words</td>
<td>3.67</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>Students discuss or provide words that are similar in meaning but have different formality degree</td>
<td>3.44</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>Students are informed of possible synonyms.</td>
<td>3.44</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>In reading activities that precede writing, the class discusses a range of synonyms, antonyms, collocations.</td>
<td>3.33</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>In my writing classes, students discuss antonyms (words of opposite meaning) of words.</td>
<td>2.89</td>
<td>0.93</td>
</tr>
<tr>
<td>Syntax</td>
<td>There are activities to explain how to expand a simple noun phrase</td>
<td>3.44</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>Students are shown how noun phrases are constructed in written academic language.</td>
<td>3.33</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>Students' attention is drawn to the number and types of modifiers used with nouns in written academic English.</td>
<td>3.22</td>
<td>1.56</td>
</tr>
<tr>
<td>Discourse</td>
<td>Students are shown how writers develop topics by adding more details to the idea in question.</td>
<td>4.00</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>My writing class activities involve discussion of how good writers begin their sentences.</td>
<td>3.56</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>Written texts are analyzed to show how writers alternate between grammatical structures to maintain the connection of ideas in a text.</td>
<td>3.11</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td>Learners' attention is drawn to how grammatical options can be used to emphasize certain elements in the argument.</td>
<td>3.00</td>
<td>1.32</td>
</tr>
</tbody>
</table>
The second part of the questionnaire (see Table 6.3) addressed teachers’ perception of what constitutes important aspects of academic writing. The aim of this section is to examine the extent to which teachers’ perceptions and instructional practices are congruent. The statements of this section are on a six-point Likert scale agreement ranging from strongly disagree coded as 1 to strongly agree coded as 6. The items reflect three aspects of linguistic knowledge: complexity of vocabulary, syntactic complexity and complexity at the discourse level. The results show a unanimous high agreement on the importance of all three aspects in academic writing. The mean agreement on the statements addressing the importance of vocabulary knowledge was relatively high ranging from (M = 4.56, SD = .50) to (M = 5.56, SD = 50).

Table 6.3 ESL Teachers’ Mean Agreement on Statements regarding Characteristics of Good Academic Writing

<table>
<thead>
<tr>
<th>Linguistic Aspects</th>
<th>Statements</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexical</strong></td>
<td>Good writers should avoid frequent repetition of the same words in their writing.</td>
<td>5.56</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>The accurate use of less frequent academic words is an indicator of good vocabulary knowledge.</td>
<td>4.56</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>To be a good writer, one should have an adequate knowledge of synonyms or words of similar meaning</td>
<td>5.11</td>
<td>.57</td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
<td>Good writers use effective balance of complex and simple grammatical structures.</td>
<td>5.56</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>The use of long complex noun phrases is a common feature of academic writing.</td>
<td>5.44</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>The use of multiple nouns is common in academic writing.</td>
<td>5.33</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Prepositional phrases are commonly used to structure long noun phrases in academic writing.</td>
<td>5.00</td>
<td>.67</td>
</tr>
<tr>
<td><strong>Discourse</strong></td>
<td>Good writers realize that beginning a sentence with a reference to something that has been mentioned before creates</td>
<td>5.33</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Avoiding the repetition of the same theme or the first structural element in a sentence is an indicator of good writing ability.</td>
<td>3.78</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>In academic writing, It is more appropriate to follow the typical information structure in English.</td>
<td>4.89</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>Abrupt progression of ideas affects the overall writing quality.</td>
<td>4.67</td>
<td>1.15</td>
</tr>
</tbody>
</table>
At the syntactic level, 76% of teachers voiced strong agreement that using balanced combination of simple and complex structures is a characteristic of professional writers. Statements regarding prevalence of complex noun structures, using multiple nouns and prepositional phrases to structure noun phrases in academic writing received high level of agreement ranging from (5.00 to 5.6). At the discourse levels, beginning statements with reference to previous discourse and the appropriate structuring of discourse from known to unknown elements received unanimous agreement as important aspects of good academic writing with means ranging from M = 4.67 to M = 5.3.

As for the common linguistic problems in ESL learners’ writing, Table 6.4 shows that teachers seem to agree that learners have problems at all the three levels: lexical, syntactic and discourse complexity. At the lexical level, the use of limited range of academic vocabulary and the inappropriate use of words of similar meaning are common from teachers’ perspective (M = 5.22, SD = .83). At the sentence level, the frequency use of clausal modifiers of nouns and the use of short noun phrases are perceived to be features of ESL writing with mean agreement ranging from (M = 4.67, to M = 5.22).

At the discoursal level, the teachers also agree that redundant topical structures and abrupt progression are major problems in ESL writing. The mean agreement on items featuring discourse level aspects of good academic writing ranges from (M = 4.33 to M = 4.89). The findings of the survey with regard to what teachers perceive to be important aspects of academic writing and what is voiced to be problematic in ESL writing was supported by the results of the open-ended question. In their answers to the open-ended question regarding the activities that may help enhance lexical and grammatical knowledge of ESL learners, six out of nine teachers mentioned that guided analysis of model texts and/or comparative analysis of both authentic
texts and learners’ texts in terms of certain structures, lexical usage, synonyms antonyms all contribute to attracting attention to how language is used in an academic manner. Some teachers (four teachers) emphasized sentence level activities focusing on vocabulary and grammar particularly for Level B students. One teacher argued that “according to [her] teaching experience, some 101B students after advancing to 101C still make some mistakes such as run-on sentences, fragment sentences; therefore, for intermediate students, it is a necessity to just focus on activities on sentence-level errors. When they advance to 101C (advanced-low), paragraph-level activities should be more focused for those students”.

Table 6.4 Common linguistic Problems of ESL Student Writing from the Perspectives of ESL Teachers

<table>
<thead>
<tr>
<th>Features</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary-Related Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>limited range of vocabulary</td>
<td>5.22</td>
<td>.83</td>
</tr>
<tr>
<td>Similar words but different usage</td>
<td>5.22</td>
<td>.83</td>
</tr>
<tr>
<td>Over use of common words</td>
<td>4.89</td>
<td>.93</td>
</tr>
<tr>
<td>Inaccurate use of word forms</td>
<td>4.22</td>
<td>1.72</td>
</tr>
<tr>
<td>Syntactically-Related Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>more clausal than phrasal noun modifications</td>
<td>5.22</td>
<td>.83</td>
</tr>
<tr>
<td>Mixed patterns of formality and informality</td>
<td>5.22</td>
<td>.83</td>
</tr>
<tr>
<td>Short NP</td>
<td>4.67</td>
<td>1.00</td>
</tr>
<tr>
<td>Discourse-Related Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redundant topical structures</td>
<td>4.89</td>
<td>.93</td>
</tr>
<tr>
<td>Problems with appropriate shift of topics</td>
<td>4.78</td>
<td>.83</td>
</tr>
<tr>
<td>Abrupt progression of topics</td>
<td>4.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The results from the survey of ESL teachers indicate a compatibility between teacher’s perceptions of the importance of textual organization and the focus on this aspect in ESL classrooms. There is, however, a lack of alignment between the beliefs about lexical and syntactic features and the extent of emphasis they receive in ESL classrooms. Even though lexical and syntactic aspects are signaled as important, they tend to be less frequent in
classrooms’ activities compared to discourse level instruction. Perceived to be an important aspect of academic writing, textual organization tends to be prioritized in ESL writing courses. When asked about the main characteristics of good academic writing, all teachers pointed to organization as the most important feature. All the teachers, also, emphasized that good written essays should have a clear thesis statement. Sue, a teacher for both B and C levels, commenting on good academic essays indicated that: “if a student can convey his own ideas and has a good structure, I think it is OK”.

Contrary to text organization, there is less computability between ESL teachers’ beliefs about the importance of lexical and syntactic related features and the focus of writing instruction. Aside from encouraging learners to develop their vocabulary repertoire and the correction of vocabulary errors, class instruction tends to allocate little time to vocabulary expansion activities. In the interviews, teachers attributed the lack of focus on vocabulary to different factors: time limitations, insufficient exercises in the assigned text books, the lack of motivation on the part of students and the lack of knowledge on how to connect vocabulary and grammar related activities to writing. Diana, a teacher of level C students, indicated that the text book includes sections for vocabulary exercises, “But [they] do not spend time on it because [they] do not have enough time”. Nancy, a teacher of B level, however, pointed to the insufficiency of vocabulary related exercises covered by the text book stressing that “it is not enough at all”. The lack of focus on vocabulary expansion activities was also attributed to the lack of motivation on the part of learners. Diana, indicated that vocabulary development is important, yet she believes that “vocabulary development is quite possible with motivated students, the majority of our students are not motivated. They are taking a required course because they have to, it was not their lucky day when they took the EPT”.
In addition to time and motivation factors, two teachers voiced difficulty in how to address lexical and grammatical issues or how this knowledge contribute to writing development. Nancy indicated that some of her students asked for help in improving their vocabulary and grammar, “but [she didn’t] have a clear answer for them”. Similarly, Diana stressed that she realized the importance of lexical and syntactic knowledge in writing and she includes it in her syllabus, however, she was not sure how- beyond leading to fluent production and accurate use, this knowledge impacts the written product: “it is really hard to tell how much it helps except that motivated students tend to have more accurate forms, more fluent English production”.

As for grammar, the interviews support the findings of the survey that syntactic complexity related activities are not common in ESL writing courses. The interview with teachers indicated that grammar-related activities in B and C levels focus on error correction, particularly for run-on sentences and incomplete sentences and conjunctions. Additionally, one teacher emphasized that accuracy of tense use is emphasized in the syllabus of 101C because, in her view, it affects the intelligibility of texts: “we look at the type of errors that affect comprehensibility and the intelligibility and we start with verb tense and word form that affects sentence level issue”.

When asked about the common problems noticed in ESL writing, the interviews also align with the survey results that ESL students exhibit problems across vocabulary, grammar and discourse levels. At the word level, repetition of vocabulary, translations from an L1, the use of contextually inappropriate words, and the use of informal language were, in teachers’ perceptions, common problems in level B and C. As for grammar, grammar accuracy and short sentences were voiced to be problematic in both C and B writing. Nancy, a teacher of level B,
indicated that short sentences are common in this level. Problems in expanding sentences which affect both meaning and cohesion was also voiced to be common in C level writing: “it is both the mechanics of stretching out their English and lack of cohesion” (Diana said). Another problem that seems to be more pertinent to level C is the mismatch between the form used and the semantic functions. Diana provided an example saying that “they use on the other hand not for comparison but because they are taught to use it”.

As for problems at the discourse level, the findings of the interview support that of the survey regarding the commonality of illogical sequence of ideas particularly in low level writing. Dan, a teacher for level B, indicated that his students need to rewrite sentences “so they can logically connect”. Diana attributed the lack of coherence in level C writing to the inappropriate use of grammar: “they think they finished their thoughts; they move to the next sentence and end up with syntactically broken sentences”.

Interviews with ESL teachers indicated an overall satisfaction with the placement decisions of the EPT despite some cases of misplacement. One teacher indicated that participating in the rating of the EPT, she realized the difficulty of differentiating between C and B levels. Sue, a teacher for level C, pointed to potential harsh rating at least, as “a record” number of students were moved to 150 out of one of her recent sections based on an in-class diagnostic. All teachers agreed that the diagnostic test at the beginning of the semester helps overcome potential misplacement decisions. One teacher emphasized the need for allowing students more time to take the diagnostic test stressing that being administered the first day, the test is usually missed by many students. Another short coming of diagnostic test based on the same teacher’s view is that regardless of the number that the teacher believes to be misplaced only 3-4 papers will be considered by the course coordinator.
The teachers also expressed, despite some complaints from students and some persistent problems in students’ writing, an overall agreement that the course materials and activities helped a lot of students to improve their writing. When asked about whether they believe that the syllabus and activities of ESL classrooms met the objectives of the course and the expectations of their students, ESL teachers indicated that they witnessed improvement in their students’ writing, despite some persistent problems in vocabulary and grammar. Nancy, a teacher of level B, indicated that “[she] saw no thesis statement in the first session, but now [she] can see that 99[of students] have thesis and divided paragraphs although many of them have long sentences in one paragraph. But the problem is vocabulary”. Teachers also pointed to some factors they believe affect their students’ satisfaction of the course. Sue, another teacher of level B, pointed that she believes that the course materials do meet students’ expectations “most of the time”. However, she thinks that some students were not satisfied with the lack of coverage of grammatical topics: “For 101B we have a grammar book but we did not have time to use it, so some students complain we did not use the book why did we buy it”. Diana pointed that motivation seems to affect students’ perceptions about the course: “The motivated students are incredibly very happy … I would say half of our students feel that they got something from 101C The others feel they did not for different reasons”.

To sum up the findings from ESL teachers’ survey and interviews, both the survey and the interview analysis pointed to a congruency between teachers’ beliefs about aspects of academic writing and the focus of instruction in some but not all areas. Structural organization which teachers perceive to be highly important ends being the focus of ESL writing courses. However, even though the use of complex vocabulary and structures are considered critical skills and that they were, from teachers’ perspectives, problematic in students’ writing, they tend to
receive less attention compared to discourse organization skills. The results overall tend to come in support with previous research findings (False, 2004) and Shi and Cumming (1995) regarding the prominence of organization and the lack of vocabulary and syntactic complexity relevant activities. Even though teachers voiced limited and inappropriate use of vocabulary and complex structures as common problems in ESL writing, the instructional focus is geared toward text organization in different genres. The interview findings suggest that the lack of focus on vocabulary expansion and syntactic awareness is linked to various factors including, from teachers’ perspectives, time, self-motivation and the insufficient knowledge on teachers’ part of how this knowledge contributes to the quality of writing.

**ESL Students’ Self-Assessed Writing Ability**

2. To what extent do ESL learners demonstrate self-assessed improvement in linguistic ability after ESL writing courses?

To answer the question of whether the students perceive improvement in their writing ability after taking ESL writing courses, the results from both the pre- and post- self-assessment surveys were analyzed. Interviews with students with regard to the EPT writing test, course evaluations, and the overall self-assessment of writing ability after ESL courses were also conducted. I will first present and discuss the results from the analysis of pre- and post-surveys using descriptive statistics and One-way Repeated Measures ANOVA. The results from the survey will then be further explored using the findings from the semi-structured interviews with ESL students.

Before analyzing the survey items, negatively worded items were reversed coded. The sum of scores of the pre and post surveys for each case were then obtained. The descriptive statistics as shown in Table 6.5 indicates an overall improvement in the ESL learners’ perceived writing ability from \((M = 71.1, SD = 10.95)\) to \((M = 76.98, SD = 10.44)\) toward the end of ESL
writing courses. Students in each of the three levels i.e. B, C and D (see Table 6.6) perceived an increase in their writing ability.

Table 6.5 *The Means of Pre and Post Self-Assessment Surveys*

<table>
<thead>
<tr>
<th>Self-Assessment</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-survey</td>
<td>71.11</td>
<td>10.953</td>
<td>47</td>
</tr>
<tr>
<td>Post-survey</td>
<td>76.98</td>
<td>10.435</td>
<td>47</td>
</tr>
</tbody>
</table>

Table 6.6 *Means of Lexical, Syntactic and Discourse level Skills in the Pre and Post Self-Assessed Surveys*

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Pre-Survey Mean (SD)</th>
<th>Pre-Survey CI (Lower-Upper)</th>
<th>Post-Survey Mean (SD)</th>
<th>Post-Survey CI (Lower-Upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Level</td>
<td>8</td>
<td>73.00 (13.87)</td>
<td>61.41-77.13</td>
<td>78.88(8.11)</td>
<td>72.09-85.66</td>
</tr>
<tr>
<td>C Level</td>
<td>20</td>
<td>73.05 (8.72)</td>
<td>68.97-77.13</td>
<td>76.35(8.57)</td>
<td>72.34-80.36</td>
</tr>
<tr>
<td>D Level</td>
<td>19</td>
<td>68.26 (11.72)</td>
<td>62.62-73.91</td>
<td>76.84(13.15)</td>
<td>70.50-83.18</td>
</tr>
</tbody>
</table>

To examine whether the self-perceived improvement in writing is statistically significant, repeated within subjects One-Way ANOVA was conducted (see Table 6.7). The assumptions of normality were first examined. The normality was checked using Shapiro-Wilk test and box plots. Assumption of normality was met for all independent variables except in Level C, where Shapiro Wilk test was statistically significant (p =.015 <.05). When the one case outlier was removed from level C, normality found to be met (p = .75 <.05). Nevertheless, and because no difference was found in the results after the removal of the one outlier, it was decided to proceed
with the analysis using all cases. The One-Way Within-Subjects ANOVA revealed that the mean scores for self-evaluation of writing ability has a statistically significant increase at .05 level: F (1, 46) = 13.40, partial eta squared = .23, which is a large effect (Gray & Kinnear, 2012, p. 328).

Table 6.7 Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power*</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>Sphericity Assumed</td>
<td>810.383</td>
<td>1</td>
<td>810.383</td>
<td>13.401</td>
<td>.001</td>
<td>.226</td>
<td>13.401</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>810.383</td>
<td>1.000</td>
<td>810.383</td>
<td>13.401</td>
<td>.001</td>
<td>.226</td>
<td>13.401</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>810.383</td>
<td>1.000</td>
<td>810.383</td>
<td>13.401</td>
<td>.001</td>
<td>.226</td>
<td>13.401</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>810.383</td>
<td>1.000</td>
<td>810.383</td>
<td>13.401</td>
<td>.001</td>
<td>.226</td>
<td>13.401</td>
</tr>
<tr>
<td>Error(time)</td>
<td>Sphericity Assumed</td>
<td>2781.617</td>
<td>46</td>
<td>60.470</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>2781.617</td>
<td>46.000</td>
<td>60.470</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>2781.617</td>
<td>46.000</td>
<td>60.470</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>2781.617</td>
<td>46.000</td>
<td>60.470</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To further examine the areas of writing in which students perceived improvement in their writing ability, the categories representing vocabulary, syntax and discourse ability were investigated separately. The descriptive statistics of within-subjects as shown below in Table 6.8 shows an improvement in the three components of the writing ability from the pre- to the post-self-assessment survey. The self-evaluation of the ability of using academic vocabulary has increased from (M = 33.64, SD = 5.99) to (M = 35.83, SD = 4.80) at the end of ESL courses. Similar increase is shown in the perceived ability of using syntactic structures as the mean increases from (M = 22.02, SD = 4.41) to (M = 24.51, SD = 4.36). The ability of discoursal organization moved from (M = 15.45, SD = 3.22) to (M = 16.64, SD = 2.96) toward the end of ESL courses. Three separate One-Way within subjects (or repeated measures) ANOVA were
conducted to compare the effect of ESL writing courses on ESL learners’ self-evaluation of using vocabulary, syntactic structures and discoursal knowledge in writing. The results revealed that the increase in the self-assessed writing ability from the pre to the post survey is statistically significant in vocabulary $F(1,46) = 8.88, p = .005, \eta^2 = .16$; in syntax $F(1, 46) = 12.43, p = .001, \eta^2 = .21$ as well as discourse $F(1,46) = 5.957, p = 0.019, \eta^2 = .115$. The size effects of the improvement ranges from medium in discourse to large in vocabulary and syntax.

Table 6.8 Descriptive Statistics of the Within-Subjects self-Assessment before and after ESL Writing Courses

<table>
<thead>
<tr>
<th>Self-Assessment</th>
<th>N</th>
<th>Vocabulary Mean (SD)</th>
<th>Syntax Mean (SD)</th>
<th>Discourse Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Survey</td>
<td>47</td>
<td>33.64 (5.99)</td>
<td>22.02(4.41)</td>
<td>15.45 (3.22)</td>
</tr>
<tr>
<td>Postsurvey</td>
<td>47</td>
<td>35.83(4.80)</td>
<td>24.51(4.36)</td>
<td>16.64 (2.96)</td>
</tr>
</tbody>
</table>

The results regarding learners’ perceptions of the frequency of class activities (Table 6.9) revealed that discourse relevant activities are more frequently addressed in ESL writing classes with a frequency mean of 4.06 to 4.45 on a six-point Likert scale. Compared to discourse-level activities, the activities targeting vocabulary expansion (such as discussion of synonyms, antonyms and word family members; items 20-22) were assessed as relatively less frequent, with a mean range of 3.34 to 3.85. Nevertheless, the accuracy of using words and lexical expressions receives more attention in ESL writing classrooms with a frequency mean of 4.43, from students’ perspectives.

As for the optional question regarding preferred activities for improving writing ability, only twenty-seven participants responded to that question. Most of the suggested activities (see
Appendix I) focused on analyzing academic or model texts and having more vocabulary focused activities. Specifically, almost all students from levels B and C voiced their preferences for more vocabulary-related activities whereas the majority of students in level D emphasized the importance of analyzing academic texts.

Table 6.9 Descriptive Statistics of the Frequency of Class Activities from ESL Students’ Perspectives

<table>
<thead>
<tr>
<th>Class activity</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q.28 In my writing class, texts are analyzed to see how writers support their ideas with details and examples.</td>
<td>4.45</td>
<td>1.16</td>
</tr>
<tr>
<td>Q.23 In my writing class, teachers give feedback on accurate use of words and lexical expressions</td>
<td>4.43</td>
<td>1.49</td>
</tr>
<tr>
<td>Q.29. In my writing class, there is a discussion of how good writers create a logical sequence of ideas in writing.</td>
<td>4.43</td>
<td>1.19</td>
</tr>
<tr>
<td>Q.30 In my writing class, there is a discussion of how professional writers express their agreement or attitudes towards certain arguments.</td>
<td>4.13</td>
<td>1.31</td>
</tr>
<tr>
<td>Q.31 In my writing class, teachers draw students' attention to expressions that professional writers use to engage their readers and make their argument more convincing.</td>
<td>4.13</td>
<td>1.24</td>
</tr>
<tr>
<td>Q.27 In my writing class, texts are analyzed to see how writers shift and return to the original topic in hand.</td>
<td>4.06</td>
<td>1.05</td>
</tr>
<tr>
<td>Q.20 In my writing class, teachers inform students of possible synonyms (words of same meaning) of vocabulary items in a text.</td>
<td>3.85</td>
<td>1.18</td>
</tr>
<tr>
<td>Q.24. In my writing classes, students analyze and compare complex nouns in academic texts. For example, the noun in sentence A is more complex than that in B: A. The newly developed technical improvements in the educational system attracted the audience attention. (complex noun). B. The improvements attracted the audience attention. (simple noun)</td>
<td>3.83</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Table 6.9 (continued)

<table>
<thead>
<tr>
<th>Class activity</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. 26 In my writing class, teachers draw students’ attention to the number and types of adjectives used with nouns in written academic English.</td>
<td>3.83</td>
<td>1.07</td>
</tr>
<tr>
<td>Q. 21 In my writing class, teachers discuss the antonyms (words of opposite meaning) of words.</td>
<td>3.53</td>
<td>1.16</td>
</tr>
<tr>
<td>Q. 22 In my writing class, teachers explain family members of words that are encountered in reading or class activities, e.g., policy, politics, political, politicians.</td>
<td>3.34</td>
<td>1.32</td>
</tr>
<tr>
<td>Q. 25 In my writing class, there are activities to explain how to expand a simple noun phrase into a complex long noun phrase in writing. For example, the simple noun &quot;system&quot; in sentence A has changed into a complex noun in B: A. The system needs to be improved. (simple noun) B. The current educational system of the newly admitted undergraduate students needs to be improved. (complex noun)</td>
<td>3.30</td>
<td>1.10</td>
</tr>
</tbody>
</table>

The results from the ESL survey were further supported by the findings of the semi-structured interviews with four ESL students (two from level D, one from C, and one from B). The students expressed general satisfaction with the placement decisions and the development they witnessed in their writing after taking ESL courses. Even though the students from level D felt initially that they did not need this course, they emphasized that it helped them improve their academic writing. Specifically, Level D students indicated that the course helped them with the organization, formulaic expressions and the tenses used in research papers. Students in Level B and C pointed that they experienced improvement in textual organization, the logical connection of ideas as well as vocabulary and grammar. Voicing her satisfaction with her writing after the ESL writing course, the C level student described the change in her writing saying: “I did not
have flow throughout my essay, it did not have much sense but now when I read my essays, I feel really good, it is a very huge improvement…my grades were in the 70s now my grades have gone to 90s”.

When asked about areas of language they have problems with, the students in both C and B indicated that using simple words or contextually inadequate vocabulary is still problematic: “in my writing, I use simple words, so I think I need to practice vocabulary, I think in 101B, we need to focus more, it is not a big part both in the book and the class”. Students in C and B also acknowledge having “some simple grammatical mistakes” in their writing.

To sum up the discussion of ESL students’ perceptions of their strengths and weaknesses, the results of the ESL learners’ responses suggest that ESL students perceive improvement based on their self-evaluations of their overall writing ability. Students have shown higher self-evaluation of their ability in terms of vocabulary, syntax and discoursal organization. The results from the questions regarding frequency of class activities from ESL students’ perspectives corroborated with the perceptions voiced by ESL teachers. The results of both surveys showed more focus on discourse level organization and less emphasis on vocabulary and syntactic complexity-related activities. The findings from the open-ended question suggest that students at higher levels of writing, particularly graduate students in level D, appreciate having more discourse-level activities through reading and analysis of academic writing. Low level writers in B and C levels, however, tend to prefer more vocabulary-focused activities. Voicing more need for vocabulary related activities support ESL teachers’ comments that vocabulary and grammatical structures are among the common problems confronted in ESL level writing.

**Conclusion**

The current study was conducted to investigate the assumptions underlying the ramification inference of the EPT writing. Score interpretations and placement decisions of the
EPT writing section are expected to be beneficial to both teachers and learners. As for teachers, the EPT scores or placement decisions are expected to enhance instructional practices, instructional materials to meet learners’ linguistic needs. Learning objectives and curriculum which are supposedly driven from the writing assessment use, should not be in conflict with ESL teachers’ beliefs of what constitutes writing quality (Bachman and Palmer, 2010). A congruence is then expected between teachers’ beliefs of quality writing and common problems in their students’ writing and the emphasis of their instructional practices. The results from the survey and interviews with teachers provide partial support for the ramification inference. While teachers showed an overall satisfaction with the placement decisions of the EPT and the course materials and objectives, the focus of instructional practices reflects less congruence with what teacher perceives to be important aspects of academic writing on one hand and the common problems in ESL essays on the other hand.

ESL teachers expressed unanimous agreement on the importance of vocabulary, syntax and organizational knowledge in professional academic writing. The emphasis of appropriate discoursal development and the logical connection of ideas seems, however, to relatively supersede aspects of vocabulary and syntactic knowledge when it comes to how teachers allocate classroom time and instruction. The use of inappropriate words and to a less extent inaccurate grammatical structures, simple sentences are voiced to be common problems in low level writing from the perspectives of ESL teachers’ and students. Accordingly, these areas need, from the authors’ perspective, more attention.

With respect to learners, the score interpretations and placement decisions are meant to place students in courses that address their linguistic needs. Both the survey and interviews with
learners revealed that learners do perceive higher self-assessment of their overall writing ability after ESL writing courses.

**Pedagogical Implications and Limitations of the Study**

The results of the current study highlight the importance of three major interrelated components of writing ability: vocabulary, syntax and discourse. Despite the established importance of these aspects in academic writing form both research on writing across different levels and from teachers’ perceptions, ESL writing courses tend to approach them in an unbalanced manner. The focus on discourse organization in terms of the overall structure of the text overrides attention to the contribution of vocabulary and grammar to discourse connectivity and the complexity of the language used. ESL courses should address these issues in a more systematic manner through specifying explicit goals for these three important aspects of academic writing especially in low level writing. Reading activities in writing courses should be geared toward achieving this goal. Learners attention should be drawn to how proficient writers utilize their vocabulary knowledge and grammatical devices to deliver more effective message.

While the findings of this study are pedagogically insightful, they are not without limitations. The major limitation is that the findings with regard to self-assessment of writing are based on relatively a small number of students who may not be representative enough for all EPT placement levels particularly for Level B learners. Also, the improvement of the students’ writing ability was only measured by their own self-evaluation, future studies are encouraged to examine the match between learners’ self-assessment of their writing ability and their actual written production before and after ESL courses. As for the findings regarding the congruency between teachers’ beliefs and instructional practices, the current study examined this relation from teachers’ own perspectives, observation of ESL writing practices, the type and the focus of teachers’ feedback across levels would provide a more informative picture about the
interrelatedness between teachers’ belief about language and their actual instructional practices on one hand and how this match is influenced by moderating factors such as learning objectives and learners’ linguistic needs.
CHAPTER 7. SUMMARY AND CONCLUSIONS OF FINDINGS

This chapter provides a summary of the questions addressed in the current study, the main findings, and the major theoretical and empirical implications. Adopting the framework of an interpretation and use argument, the current dissertation utilized a mixed method approach to investigate two main inferences underlying the interpretation and use of the EPT writing scores: the explanation and ramification inferences. The explanation inference addresses the extent that ESL students’ performance in the EPT writing reflects differences in their linguistic complexity, which is a major subconstruct of writing ability. Linguistic complexity has been investigated in the current study through three major dimensions: lexical, syntactic and discourse. Higher level writers are expected to display more complex use of vocabulary, syntactic structures and topical development. The ramification inference emphasizes the impact that the placement decisions of the EPT writing section has on both learners and ESL teachers. Placement decisions are assumed to 1) help learners address their linguistic problems and accordingly, improve their writing ability, and 2) help teachers undertake instructional practices that address students’ linguistic problems and hence, contribute to their students’ writing development. The next paragraphs present a summary of the findings pertinent to the explanation and ramification inferences. Additionally, a table summary of the EPT validity argument and an evaluation of the evidence obtained for the explanation and ramification inferences will also be presented.

**Lexical complexity, Diversity and Sophistication across Placement Levels**

The first question in the current study addressed the first assumption of the explanation inference in the EPT writing interpretation /use argument. The first assumption states that ESL students’ performance on the EPT writing is expected to reflect differences in students’ linguistic complexity. Specifically, the first research question focused on one aspect of linguistic
complexity, namely lexical complexity. Two main free software were used to examine five indices of lexical complexity pertinent to lexical diversity, sophistication and cohesion. A subsample corpus of 554 texts was used to analyze aspects of lexical complexity in five placement levels. The results of descriptive statistics pointed to differences between the placement levels and also to an influence of the education level (graduate vs. undergraduate) on the complexity of lexical production. The multivariate analysis (MANOVA) revealed that the Pass graduate levels displayed statistically higher use of complex vocabulary in terms of diversity, sophistication and cohesion than the low-level B. The differences in the use of diverse and sophisticated vocabulary were not consistent among placement levels with undergraduate students. In terms of diversity, the pass level used statistically more diverse vocabulary than the low-level B. Other pair-wise comparisons in diversity fall short from statistical significance.

Lexical sophistication was investigated through the use of AWL words, range of academic words and range of bigrams. The findings suggest that graduate vs. undergraduate student level was related to learners’ use of academic words. The graduate students in both Pass and D Levels used statistically more AWL words than their undergraduate peers. The use of contextually restricted words and more common bigrams revealed more differences on sophisticated use of vocabulary among the EPT placement levels. The Pass level use of low-range words was statistically higher than lower undergraduate levels of C and B. In addition, students in all levels displayed higher use of contextually diverse bigrams than level B. Similar to that of bigrams, the analysis of cohesion index i.e. hypernymy of nouns show that all levels used statistically higher hypernyms than level B. The overall results provide partial evidence for the explanation inference of the EPT validity argument. Even though the Pass level showed better knowledge of using complex vocabulary than other levels, the differences among
placement levels of undergraduates, namely in the use of AWL words and diverse vocabulary, do not reflect variations in lexical ability. Specifically, level B students were found to outperform their colleagues in level C in the use of AWL words. In addition, no statistical differences were found between graduate students in D and Pass in terms of lexical complexity. The inferential statistics suggests partial support for the expectations of variations in lexical complexity across levels. This conclusion should, however, be considered with caution for several reasons. First, with respect to the results of AWL, the close inspection of few samples from B and C levels’ writing suggests that the increase in using AWL in B level is inflated by the excessive repetition of academic words. Another important factor that should be taken into consideration is the relatively small range of proficiencies reflected in the EPT writing section which in turn, affects the magnitude of variations among groups particularly graduate level students.

**Syntactic Complexity across Placement Levels**

The second question addressed syntactic complexity in ESL learners’ writing, which is the second dimension of linguistic complexity. A subsample of the EPT texts (554) that were tagged using Biber tagger was used to investigate the distribution of 12 linguistic features representing clausal and non-clausal features across the placement levels. Descriptive statistics of the normalized frequency revealed differences in syntactic complexity across levels and also variations in syntactic complexity between graduates vs. undergraduates. In line with previous research findings, the graduate level students in both Pass and D levels used higher frequency of nominalization and nominal modifiers, specifically attributive adjectives and post nominal modifiers (prepositional phrases, non-finite relatives, that clause complements of nouns) than all other undergraduate levels. Nouns as nominal pre-modifiers did not vary by level, however. As for features constituents of a clause, the findings revealed that the use of finite adverbials
decreases in higher placement levels. No clear pattern was however, detected from the frequency of use of non-finite adverbials, which were rare in the corpus.

The frequency use of nominal modifiers and the decrease of finite adverbials in higher level writing support corpus research conclusions that noun phrase complexity is a prominent characteristic of academic writing and that clausal subordination is more prevalent in spoken as opposed to academic written texts. The lack of variations in the use nouns as prenominal modifiers might be affected by the genre of writing. In other words, this feature is not common in argumentative writing compared to informational texts such as newspapers and text books.

Furthermore, the qualitative analysis of complex syntactic structures sheds light on the intricate relationship between syntax and lexicons. Even when using the same syntactic structures, high and low-level writers exhibited differences in their lexical choices, which may in turn impact the judgement of complexity. For example, the use of *that* complement of nouns in high level writing has shown to be associated with less personal nouns than that in lower level writing. This indicates that the complexity in terms of both frequency and the constructed meaning changes as learners advance in writing.

Overall, the results of syntactic complexity provided support for the explanation inference of the EPT validity argument. Specifically, the frequency of use for phrasal and clausal complex structures shift in expected ways (based on patterns observed in the target domain of academic writing), thus providing evidence for the assumption that the performance on the EPT writing reflects differences in one aspect of linguistic complexity, i.e., syntactic complexity.

**Topical Development Patterns across Placement Levels**

The third aspect of linguistic complexity analyzed in order to investigate the assumption related to the explanation inference is topical development. High level writers are expected to
provide evidence of their ability to maintain and elaborate on the topic at hand. Research on topical development in expository writing revealed a more frequent use of sequential and indirect parallel topical progressions in both native English writing and in high-scoring ESL essays. Sequential progression is used to further develop the topic through details and examples, while indirect parallel progression helps maintain the focus on the topic without undue redundancy. A MANOVA test was used with the progression types (parallel, indirect parallel, extended parallel and sequential) as dependent variables and the placement levels as independent variables. The results pointed to statistically significant differences in the use of sequential progression. The pair-wise comparisons using LSD post hoc test showed that sequential progression was statistically higher in graduate Pass than the B Level. In addition, undergraduate Pass and D levels’ use of sequential progression was significantly higher than those in levels B and C.

Even though the differences in the frequency of use of other progression types were found to be statistically insignificant, the patterns illustrated by descriptive statistics are informative from a discourse analysis perspective. The descriptive statistics revealed an increase in the use of indirect parallel structures and a decrease in parallel and coherence break progressions as the placement level increases.

The mixed method approach adopted in the analysis of ESL learners’ writing has illuminated the importance of triangulating quantitative analysis of topical progression with a qualitative inspection of how different progression types contribute to textual coherence across levels. The qualitative analysis of progression types revealed that Pass (both graduates and undergraduates) and D levels have more awareness of the discourse functions served by certain topical progression types, despite problems in constructing syntactically complex sentences in D level in particular. While sequential progression is shown to be used by Pass and D levels to
provide further details on comment-initiated themes, it is largely used by lower levels (C and B) to introduce either redundant, irrelevant or weakly relevant information to a previously forwarded argument. Also, higher level writers in Pass and D levels have been shown to utilize different strategies to maintain the topic without unnecessary repetition. Unlike low-level writers in C and B, in which the topical subjects are commonly realized by pronouns or repeated lexical phrases, higher level writers in Pass and D avoided repetition through the use of synonyms and nominalization as topical subjects, among other strategies.

The results of topical analysis provide support for the assumption underlying the explanation inference for the EPT validity argument. Specifically, these results back the assumption that the performance of ESL students on the EPT writing section reflects differences in their ability to construct complex language at the discourse level. Higher placement levels in Pass and D show better ability to develop more coherent texts than both C and B levels. The topical analysis, however, highlighted some weaknesses in the rubrics for the EPT writing section. The substantial differences between C and D levels do not support the unification of rubrics for these two levels. While the descriptors which point to the lack of using examples, cohesive devices and the use of redundant language align with the characteristics of level C students, they are not supported by the discourse analysis of D level writing. This conclusion is further solidified by the differences in the use of initial sentence elements, which is summarized in the next section.

**The Use of Initial Sentence Elements (ISEs) across Placement Levels**

The fourth question complements the third question in providing evidence or lack of evidence for the explanation inference in the EPT validity argument. The question investigated another aspect relevant to topical complexity, namely the frequency, linguistic realizations and the discourse functions of initial sentence elements. It is expected that higher placement levels
would portray better use of ISEs in constructing textual coherence. The findings from the analysis of ISEs pointed to differences between high and low placement levels. As for the frequency of ISES, the descriptive statistics showed that higher levels (Pass and D) used ISES more frequently than both C and B levels. However, the pair-wise comparisons using Tukey HSD showed that the Pass (graduates) and D level use of ISES was statistically significantly different than their undergraduate Pass peers. The normalized frequencies for the linguistic realizations of ISEs revealed similar distributions of the following syntactic forms across levels: prepositional phrases, linking adverbials, and adverbs. In contrast, finite clauses as ISEs were found to be less frequent in level B compared to other levels. Assuming that low level writers in B show more resemblance to spoken language, the relative high frequency of prepositional phrases (PP) and the low frequency of finite clauses seem to misalign with previous research that found that PPs are more prevalent in writing whereas finite clauses are more frequent in spoken language.

The discourse functions served by these ISEs provided better explanation for their frequency across placement levels. Despite their relative frequency in both high and low levels, linking adverbials and prepositional phrases served different discourse functions. Prepositional phrases were used by higher level writers in Pass and D to mainly connect discourse and to a lesser extent to present stance. However, PPs were mainly used by lower level writers as circumstance adverbials or to express a personal stance. The use of finite relative clauses to connect given and new discourse in higher placement levels explains their higher relative frequency in higher levels compared to the lower B level. Even though linking adverbials were frequently used across levels, the qualitative analysis revealed that they are used differently.
Unlike higher level writers in Pass and D, lower level writers (particularly in level B) used linking adverbials to add rather than to link ideas in discourse.

Overall, the findings for the frequency and discourse functions of ISEs come in support for the assumptions underlying the explanation inference in the EPT validity argument. The analysis of ISES revealed variation in ESL students’ ability to construct coherent texts. Both Pass and D levels portrayed a more frequent use of ISES than lower levels (C and B) and also more appropriate use of their discoursal functions

**The Alignment between ESL Teachers’ Belief about Language and their Instructional Practices**

The fifth question sought to provide backing (or lack thereof) for the assumptions of ramification inference in the validity argument for the EPT writing section. Particularly, it assessed the impact that the EPT test use/interpretations has on both teachers and students. The use of EPT writing placement decisions is expected to direct focused attention on language features important for academic writing, and which are subsequently included in instructional activities in courses that address learners’ linguistic needs. Learners are also expected to benefit from their assigned ESL writing courses and improve their writing accordingly. Therefore, the investigation of the ramification inference focuses on the impact of the EPT writing on both teachers and ESL students. The fifth question focused on teachers’ perceptions of good academic writing and the alignment between their beliefs and their actual instructional practices. To answer this question, a survey and interviews with ESL teachers teaching English 101B, C and D were conducted. Both the survey and the interviews assessed teachers’ perceptions of features expected to be reflective of good academic writing, common problems in ESL writing, and the frequency of instructional practices addressing important aspects of academic writing. The survey addressed three main aspects of linguistic complexity: lexical, syntactic and discourse.
The results of the survey showed that teachers perceive the three aspects to be of relatively equal importance for quality academic writing. It also revealed that teachers believe that ESL students have difficulties in all three levels of linguistic complexity. Although teachers believe that students have difficulties at the three level of linguistic complexity, the inspection of frequency of instructional practices pointed to more emphasis on discourse development over aspects of lexical and syntactic complexity. Activities targeting syntactic complexity tend to focus on accuracy of using grammatical structures rather than development of complex language. The lack of vocabulary focused activities was attributed to various factors including time, insufficient coverage of this aspect in textbooks and the lack of motivation on learners’ part. The findings from ESL teachers’ survey and interviews concerning the lack of activities addressing expansion of vocabulary and syntactic complexity were also corroborated by the results from ESL students’ survey and interviews.

The results from the study of ESL teachers’ perceptions and instructional practices provides partial support for the ramification inference of the EPT validity argument. Even though lexical and syntactic complexity features are posited to be important for academic writing and despite being flagged as problematic areas from both teachers’ and ESL students’ perspectives, they do not seem to be sufficiently addressed in ESL classrooms.

**ESL Students’ Self-Rated Development in Writing after Taking ESL Courses**

To answer the question of whether ESL students perceive improvement in their writing ability after ESL courses, interviews and a pre- and post- survey of learners’ self-perceived writing ability were conducted. The survey addressed three main aspects of linguistic complexity in academic writing: lexical, syntactic and topical. One way within subject repeated ANOVA revealed statistically significant perceived improvement from ESL courses. The separate
univariate ANOVA tests disclosed that the perceived development in lexical, syntactic and discourse level features were statistically significant at .05 level. The quantitative results were further triangulated by qualitative analysis of the semi-structured interviews with ESL students. The results revealed an overall satisfaction with both the placement decisions and the ESL courses.

**A Summary of the Assumptions, Warrants and Backing for the EPT Validity Argument**

Table 7.1 presents a summary of the assumptions, warrants and the extent of support the findings from this study have provided for the explanation and ramification inferences in the EPT validity argument.

**Table 7.1 A summary Table of the Validity Argument of the EPT Writing**

<table>
<thead>
<tr>
<th>Inference</th>
<th>Warrant licensing the inference</th>
<th>Assumptions underlying warrants</th>
<th>Evidence</th>
<th>Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The attributes of performance in EPT writing test are related as expected theoretically to the construct of linguistic complexity.</td>
<td>Performance on the EPT writing test reflects complexity in test-takers' linguistic ability.</td>
<td>a. Lexical analysis: MANOVA test revealed statistically significant differences between high level writers in Pass and B level. Differences between other levels namely undergraduates were not consistent in terms of AWL use and range of AWL.</td>
<td>Partial Support</td>
<td></td>
</tr>
<tr>
<td>b. Syntactic analysis: the results of normalized frequency showed that graduate Pass used more complex structures (namely nominalization and prepositional phrases) than graduate D level. Also, a linear increase in the use of syntactically complex structures along with the increase in proficiency level is witnessed among undergraduate levels.</td>
<td></td>
<td></td>
<td>Full Support</td>
<td></td>
</tr>
</tbody>
</table>
Table 7.2 (continued)

<table>
<thead>
<tr>
<th>Inference</th>
<th>Warrant licensing the inference</th>
<th>Assumptions underlying warrants</th>
<th>Evidence</th>
<th>Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Topical Analysis: Descriptive statistics indicated that placement levels in Pass and D levels displayed better textual coherence. They used more sequential, indirect parallel and less coherence break than other levels. MANOVA post hoc tests showed that Pass (undergraduates) used statistically more sequential patterns than C and B levels.</td>
<td>Full Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The consequences of using EPT writing assessment and of the decisions that are made are beneficial to stakeholders</td>
<td>Full Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. ESL students will be able to recognize their strengths and weaknesses and will organize their study accordingly</td>
<td>Repeated within subjects One-way ANOVA revealed statistically significant improvement in ESL student’s self-assessed writing ability. Three separate one-way within subjects ANOVA pointed to statistically significant improvement in the self-rated ability at lexical, syntactic and discourse levels.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The descriptive statistics of teachers’ perceptions of good academic writing, common problems in ESL writing and the frequency of instructional practices targeting these features revealed a partial congruence between what teachers perceive to be important and the respective instructional practices. Discourse level activities tend to override other perceived important aspects of lexical and syntactic complexity</td>
<td>Partial support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The instructional practices of ESL writing courses will improve, and instructors will use placement decisions to plan and organize their instructional activities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**General Theoretical and Empirical Implications**

The methods of analysis and the results of the current dissertation study are insightful to research on the assessment of writing in general and to placement tests in particular. The study has shed light on the complexity not only of defining the writing construct but also its subcomponents or subconstructs. Linguistic complexity is a major subconstruct of writing ability that has attracted a considerable amount of research. It is also commonly designated as
a feature of advanced level writing in rating rubrics, yet there is no consensus on the definition of this construct. In fact, the difficulty of definition extends to its subcomponent i.e. lexical, syntactic and discourse levels aspects which have shown to be interrelated. For example, the analysis of syntactic complexity highlighted the importance of considering the interrelatedness between syntax and vocabulary. The same syntactic structures can be assigned a lower or a higher degree of complexity depending on the complexity of vocabulary used. Similarly, the analysis of topical progression patterns and ISEs indicated connection between topical development, syntactic structures and vocabulary in developing coherent texts. Higher level writers maintained their topics through a variety of reiteration strategies including synonyms, nominalization and encapsulating nouns. They also used specific structures such as prepositional phrases and finite adverbials to maintain the connection of given and new information. Therefore, a comprehensive practical definition of linguistic complexity should consider both contextual factors and the interrelatedness of the various components of complexity.

Another major theoretical implication of the current work is the need for connecting form to meaning and discourse functions, particularly when discussing and justifying the distribution of particular linguistic forms and structures. This is an indicator of the difficulty inherent in generalizing quantitative findings without an in-depth consideration of the connection between the predominance of certain features and their discourse functions.

The current dissertation work also has some practical implications that can be insightful for future research on writing assessment and placement tests. First, the emphasis of contextual factors and the interrelatedness of meaning and discourse functions highlight the need for replication studies in the field of assessment. Replication studies would contribute
meaningfully to the enhancement of theory and assessment practices and can largely work on bridging the gap between theory and practice in the field of language assessment, as well as in ESL classroom practices. Second, the results of the current study point to the importance of discourse analysis in refining and enhancing the rating rubrics of a given assessment tools. Rating rubrics are usually developed based on some standardized and commonly-used framework. Examination of the extent of alignment between actual students’ writing and rating rubrics provides backing or lack of backing to the adopted level descriptors.

Third, the study also highlights the need of probing both teachers’ and students’ perceptions of what constitute to be problematic areas in students’ writing, particularly for in house tests. Placement tests are used to help students identify their problems and improve their writing accordingly. Surveys and interviews with both teachers and learners can be time efficient methods of surveying learners’ linguistic needs. These can further be consolidated by discourse analysis of students’ writing at different times during their ESL courses.

**Practical Implications for the EPT Writing Section**

The results of linguistic complexity analysis of the EPT writing across placement levels have several practical implications for the EPT writing section and ESL writing courses. The lexical, syntactic and discourse level analysis revealed a need for refining level descriptors particularly for level C/D students. The differences that were found between graduate and undergraduate level students in lexical, syntactic and topical complexity suggest consideration of education level (graduates vs. undergraduates) in the description of these features. Also, the results of syntactic complexity suggest that certain linguistic features mainly normalization and postnominal modifiers can be used to provide examples of complex use of syntax across levels. Both the analysis of syntactic structures and the initial sentence elements suggest that
the discourse functions of finite adverbials can be used to discriminate higher vs. lower level writing.

The findings from the discourse analysis of ESL writing along with the surveys and interviews with ESL teachers and students suggest that constructing statistically complex structures is an area of struggle even for D/C levels. In addition to problems in developing syntactically complex structures, undergraduate level students have shown to have inadequate repertoire of vocabulary. ESL courses are encouraged to develop a balanced curriculum that interconnects lexical, syntactic and organizational level skills.

**Concluding Remarks**

This study was conducted to examine the validity of the score use and interpretations of the EPT writing test. Providing comprehensive and practical guide for evaluation research, the interpretation /use argument framework was used to guide the inquiries carried out in this dissertation. The investigation focused on two main inferences: explanation and ramifications. The analysis of syntax and topical patterns provided a full support for the EPT validity argument whereas the results of lexical analysis revealed partial support. The inspection of the ramification of EPT writing showed mixed results. While ESL students perceived significant development in their writing ability after ESL courses, the focus of instructional practices as reported by ESL instructors seems to fall short from addressing lexical and syntactic level problems of ESL learners. The researcher hopes that the results of this work will contribute to the improvement of both the EPT rating rubrics, curriculum planning and instructional practices in ESL classrooms. It is also hoped that the limitations and avenues of future research that were highlighted in previous chapters would encourage researchers to utilize the argument-based approach for more in-depth inspection of under-researched areas in writing assessment in general and English placement tests of writing in particular.
REFERENCES


ISU EPT Corpus of Learner Writing (Release 2.0). (2017). Corpus compiled by the Applied Linguistics and Technology program and Bethany Gray at Iowa State University.


## APPENDIX A. RUBRICS OF THE EPT WRITING

<table>
<thead>
<tr>
<th>Levels</th>
<th>General Descriptions</th>
<th>Organization</th>
<th>Grammar &amp; Vocab</th>
<th>Functional</th>
<th>Mechanics</th>
<th>Comprehensibility</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pass</strong></td>
<td>No need for ESL instruction. A few grammatical, mechanical and stylistic errors not hindering comprehension(ENGL150/250) and compete adequately with native speakers. Graduate have adequate writing skill for graduate work in their field. Roughly corresponding to Advanced Mid or higher in ACTFL proficiency guidelines 2012-writing.</td>
<td>• Utilize a variety of cohesive devices Effectively organize the flow of thoughts in several paragraphs Stay on topic with little digression</td>
<td>• Adequate use of time frames with moderate control of aspect(e.g., perfect, progressive, habitual, momentary, etc.). Adequate control of grammar(with some minor grammatical errors/mistakes observed) A variety of complex syntactic structures. Adequate vocabulary use.</td>
<td>• Narrate and describe with details Argue with specific evidence and or examples. Make effective comparisons and contrast. Adequate use of true and untrue conditionals, although few errors are still found. Good use of formal language; with little evidence of spoken expression(UG only).</td>
<td>• A few mistakes in punctuation and spelling</td>
<td>• Readily understood by native speakers not accustomed to non-native speakers’ writing. Little efforts to make to understand Ideas fully developed with conspicuous clarity.(albeit a few grammatical, mechanical and stylistic mistakes)</td>
</tr>
<tr>
<td><strong>101C(U) 101D(G) Students can benefit from one semester of ESL instruction. Communication on formal topics is somewhat effective but often involves redundancy or repetition corresponding to Advanced Low in ACTFL proficiency Guidelines 2012-writing.</strong></td>
<td>• Use a limited number of cohesive devices Somewhat awkward and incoherent thoughts of thoughts. Somewhat fail to stay on topic, making some digression. Little use of thesis statement and/or topic sentence.</td>
<td>• Moderate use of time frame with fair control of aspect Fair use of complex syntactic structures Moderate repetitions of the same expressions or sentence structures. Fair use of vocabulary sometimes making circumlocution.</td>
<td>• Narrate with less sophisticated descriptions Argue with unclear examples or lack of examples/evidence. Make comparison and contrast to some degrees—often with inconsistencies in details and subtopics. Adequate use of true conditionals, mostly incorrect use of untrue conditional Adequate use of formal language; often/sometimes relying on spoken expression</td>
<td>• Moderate degree of mistakes in spelling and/or punctuation.</td>
<td>• Can be understood by native speakers not accustomed to non-native speakers’ writing with some efforts and patience. Underdeveloped ideas.</td>
<td></td>
</tr>
<tr>
<td>Course</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101B</td>
<td>Students need more than one semester of ESL instruction. They will proceed to 101C/D after completing 101B. Communication on formal topics is limited corresponding to intermediate High in ACTFL Proficiency Guidelines 2012_Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Very Limited use of cohesive devices
- Lack of thesis statement and/or topic sentences
- Incoherent flow of thoughts
- Unclear formation of paragraphs
- Lack of development of ideas.

- Inadequate/inconsistent use of time frames with lack of control of aspect
- Mostly rely on simple sentence structures
- Connect sentences largely relying on coordinate conjunctions (e.g., and, but, or) and common subordinate conjunctions (e.g., because, if, when)
- Many grammatical mistakes hindering readers' comprehension
- Limited use of vocabulary (repetition of the same expressions or redundancy)
- Incomplete sentences.

- Simple communication
- Narrate, describe, and/or argue with lack of details
- Ineffective comparison and contrast
- Moderate use of conditionals, often making mistakes with verb forms
- Inappropriate use of formal/informal expressions.

- Frequent systematic errors or random mistakes in punctuation and/or spelling.

- Requires a lot of efforts to understand Lack of clarity of ideas.
APPENDIX B. ACTFL PROFICIENCY GUIDELINES 2012-WRITING

ADVANCED MID (CORRESPONDING TO PASS IN EPT)
Writers at the Advanced Mid sublevel are able to meet a range of work and/or academic writing needs. They demonstrate the ability to narrate and describe with detail in all major time frames with good control of aspect. They are able to write straightforward summaries on topics of general interest. Their writing exhibits a variety of cohesive devices in texts up to several paragraphs in length. There is good control of the most frequently used target-language syntactic structures and a range of general vocabulary. Most often, thoughts are expressed clearly and supported by some elaboration. This writing incorporates organizational features both of the target language and the writer’s first language and may at times resemble oral discourse. Writing at the Advanced Mid sublevel is understood readily by natives not used to the writing of non-natives. When called on to perform functions or to treat issues at the Superior level, Advanced-Mid writers will manifest a decline in the quality and/or quantity of their writing.

ADVANCED LOW (CORRESPONDING TO 101C/D IN EPT)
Writers at the Advanced Low sublevel are able to meet basic work and/or academic writing needs. They demonstrate the ability to narrate and describe in major time frames with some control of aspect. They are able to compose simple summaries on familiar topics. Advanced Low writers are able to combine and link sentences into texts of paragraph length and structure. Their writing, while adequate to satisfy the criteria of the Advanced level, may not be substantive. Writers at the Advanced Low sublevel demonstrate the ability to incorporate a limited number of cohesive devices and may resort to some redundancy and awkward repetition. They rely on patterns of oral discourse and the writing style of their first language. These writers demonstrate minimal control of common structures and vocabulary associated with the Advanced level. Their writing is understood by natives not accustomed to the writing of non-natives, although some additional effort may be required in the reading of the text. When attempting to perform functions at the Superior level, their writing will deteriorate significantly.

INTERMEDIATE HIGH (CORRESPONDING TO 101B IN EPT)
Writers at the Intermediate High sublevel are able to meet all practical writing needs of the Intermediate level. Additionally, they can write compositions and simple summaries related to work and/or school experiences. They can narrate and describe in different time frames when writing about everyday events and situations. These narrations and descriptions are often, but not always, of paragraph length, and they typically contain some evidence of breakdown in one or more features of the Advanced level. For example, these writers may be inconsistent in the use of appropriate major time markers, resulting in a loss of clarity. The vocabulary, grammar and style of Intermediate High writers essentially correspond to those of the spoken language. Intermediate High writing, even with numerous and perhaps significant errors, is generally comprehensible to natives not used to the writing of non-natives, but there are likely to be gaps in comprehension.
APPENDIX C. CODING OF TOPICAL DEVELOPMENT AND ISES

There's no doubt that with development of technology, robots almost are everywhere. But personally, the result of that robots replace people to do plenty of job isn't good and beneficial as we imagine.

First of all, a large amount of people will lose their job if robots take over their jobs. Some company prefer hiring robots instead of human to work to increase the workforce. That will cause terrible results, for instants, there will be more and more people choosing in legal ways to make living because they are out of work and they have to make money.

Secondly, robots may bring danger to human when they are working. To be honest, robots are mashamis, which haven't experiment, creativity and judgement. They can't handle some accident as soon as possible when it happen. That would hurt us and make people losed money.

To sum up, I don't think robots can make our life better if they take over some jobs. (Level C_Fa16.txt)

There's no doubt that with development of technology, robots almost are everywhere. But personally, the result of that robots replace people to do plenty of job isn't
APPENDIX D. SURVEY OF ESL TEACHERS

**Instructions:** Please indicate how often do you emphasize the points below in your writing classes. **PLEASE** respond based on the last writing course you taught.

<table>
<thead>
<tr>
<th>First Section: Instructional Practices</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>often</th>
<th>very frequently</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>In my writing class, students are informed of possible synonyms (words of same meaning) of vocabulary items.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing classes, students discuss antonyms (words of opposite meaning) of words.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing class activities, students’ attention is drawn to different forms of a word. For example (assume, assumption).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing classes, students discuss or provide words that are similar in meaning but have different formality degree e.g., conduct research is more formal than do research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing class activities, students are encouraged to use new words in their writing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing class activities, students are often encouraged to avoid repetition of the same words.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In reading activities that precedes writing, the class discusses a range of synonyms (same words) and antonyms (opposite words), collocations of new words.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing classes, students discuss family members of words that are encountered in reading or class activities e.g., policy, politics, political, politicians.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing classes, students’ attention is drawn to the number and types of modifiers used with nouns in written academic English.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing classes, students are shown how noun phrases are constructed in written academic language.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my writing class, there are activities to explain how to expand a simple noun phrase into a complex long noun phrase in writing. For example:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. The issues were discussed thoroughly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. A serious of a complex thorny issue that are of core interests to all parties were discussed thoroughly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My writing class activities involve discussion of how good writers begin their sentences (what grammatical structures they use to begin their sentences).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In my writing activities, learners’ attention is drawn to how grammatical options can be used to emphasize certain elements in the argument. For example, sentence A places more emphasis on attitude:

A. *It is unacceptable and disappointing to see that students gained less than 40% of the available marks.*

In writing classes, students are shown how writers develop topics by adding more details to the idea in question.

In my writing classes, written texts are analyzed to show how writers alternate between grammatical structures to maintain the connection of ideas in a text. For example, subordination can be used to introduce new subtopics:

*While it is known that some self-help books are best-sellers, their overall impact on mental health literacy is unknown.*

<table>
<thead>
<tr>
<th>Second Section: Aspects of Linguistic knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good writers should avoid frequent repetition of the same words in their writing.</td>
</tr>
<tr>
<td>The accurate use of less frequent academic words is an indicator of good vocabulary knowledge</td>
</tr>
<tr>
<td>To be a good writer, one should have an adequate knowledge of synonyms or words of similar meaning.</td>
</tr>
<tr>
<td>Good writers use effective balance of complex and simple grammatical structures.</td>
</tr>
<tr>
<td>The use of long complex noun phrases is a common feature of academic writing.</td>
</tr>
<tr>
<td>The use of multiple nouns (e.g., <em>earthquake damage</em> INSTEAD OF <em>the earthquake that caused damage</em>) is common in academic writing.</td>
</tr>
<tr>
<td>Prepositional phrases are commonly used to structure long noun phrases in academic writing. For example: <em>A comprehensive analysis of the features of management of long-term budget deficits</em> was conducted.</td>
</tr>
<tr>
<td>Good writers realize that beginning a sentence with a reference to something that has been mentioned before creates a more cohesive text or discourse continuity (i.e. a link between what has already been said and what is about to be said).</td>
</tr>
<tr>
<td>Avoiding the repetition of the same theme or the first structural element in a sentence is an indicator of good writing ability.</td>
</tr>
<tr>
<td>In academic writing, it is more appropriate to follow the typical information structure in English (i.e. begin sentences with information familiar to your readers and end sentences with information relatively new/unexpected to readers).</td>
</tr>
</tbody>
</table>
Abrupt progression of ideas affects the overall writing quality.

Third Section: To what extent do you agree that the following represent problems in ESL writing?

ESL Learners often misuse words that have similar meaning but different usage

28. ESL Learners have little problems in producing accurate word forms, e.g., historical. Historic.

29. ESL Learners tend to overuse commonly frequent words in their writing.

30. ESL Learners use limited range of academic words in their writing.

31. ESL learners often tend to use more clausal type of modification. For example, the first sentence tends to be more common in learners’ writing than the second one:

The statement that was recently released shows.

The newly released statement shows...

ESL Learners tend to use short rather than long complex noun phrases.

Writing of ESL learners display mixed patterns of formality and informality.

ESL learners tend to use redundant themes (i.e. start their sentences with the same linguistic elements).

35. ESL learners often show problems in making an appropriate shift of topics in writing.

36. ESL writing display abrupt progression of the topic in hand.

Section Four: Please answer the following question:

What type of class activities do you usually do to help students improve their lexical and grammatical knowledge of English academic writing? You may add any further comments.................

IX: Please help us better identify your students’ needs by completing the following:

Gender: Male Female

Previous courses taught 101B 101C 101D other ESL writing

Mother Tongue or L1:..........................................

If you have any questions or wish to receive the results of this questionnaire, feel free to email the researcher:

zalonazi@iastate.edu
APPENDIX E. PROTOCOL OF ESL TEACHERS’ INTERVIEWS

Greeting and explanation of the study purpose.

Questions:

1. How satisfied are you with the placement decisions of the EPT placement test?

2. To what extent do you think the test help you to plan and specify the points of focus in class?

3. What do you perceive to be a good written essay?

4. What are major problems that you see in ESL writing and how do you address them?

5. What features that you repeatedly emphasize or bring to students’ attention when grading essays?

6. How vocabulary and grammar knowledge help, in your view, to build coherent argument?

7. To what extent do you think the curriculum goals and activities meet the expectations of ESL students in the class?
# APPENDIX F. ESL LEARNERS’ QUESTIONNAIRE

**First Section:** Circle the number that corresponds with your level of agreement

<table>
<thead>
<tr>
<th>I. Perceptions of Strengths and weaknesses in Writing</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Partly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My errors in the use of vocabulary affect my overall grade of writing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2. I use appropriate range of academic words in my writing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3. I accurately use different forms of a word (e.g., <em>distinctive, distinct, politics, politicians</em>) in my writing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4. I often fail to use the appropriate word combinations or collocations, e.g., we can say <em>do your homework</em>, but we don’t say <em>make your homework.</em></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5. I use low frequent academic words (academic words that are not commonly used in every day writing) in my writing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6. I use a variety of synonyms (words of same meaning) in my writing to avoid repetition of same words.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7. I make errors in using expressions that are more common in speech than in writing, e.g., <em>so, in the same way.</em></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
8. I use a variety of antonyms (words of opposite meaning) in my writing to avoid repetition of same words.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

9. My grammatical errors affect the overall grade of my writing.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

10. I accurately use long noun phrases in my writing. For example:
   a. *The issues* were discussed thoroughly.
      (short noun phrase)
   b. *A serious of a complex thorny issue that are of core interests to all parties* were discussed thoroughly. (long noun phrase)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

11. I use prepositional phrases to create long noun phrases. For example: I can write sentences similar to that in B:

   A comprehensive analysis was conducted. (The underlined structure is a short noun phrase)

   "A comprehensive analysis [of the features [of management [of long-term budget deficits]]] was conducted." (The underlined structure is a long noun phrase)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

12. I accurately use subordination structures in my writing. For example:

   As the research is still evolving, additional validation studies are needed.
   b. While the author covered a wide range of
topics, he failed to stress some critical issues.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 13. I use post modifiers of nouns (e.g., relative clauses) to make my argument more appealing to readers.  
   A. *The study which deals with ethical principles of academic writing* has attracted a lot of attention.  
   (The underlined structure is a post modifier of the noun study) | 1 | 2 | 3 | 4 | 5 | 6 |
| 14. I use pre-modifiers of nouns (for example: adjectives) accurately in my writing:  
   A. *The newly published study provides in-depth insights of the problems in hand.* (The underlined structure is a pre-modifier of the noun study) |   |   |   |   |   |   |
| 15. I appropriately use examples and details to develop my ideas in writing. | 1 | 2 | 3 | 4 | 5 | 6 |
| 16. I alternate between synonyms (words of same meaning), antonyms (words of opposite meaning) or words of different forms (*science, scientists*) in my writing. |   |   |   |   |   |   |
| 17. I make appropriated shifts of ideas (logical sequence of sentences) in my writing. | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. In writing, I begin sentences with information familiar to my readers and end sentences with information relatively new/unexpected to readers |   |   |   |   |   |   |
19. I often have problems relevant to text organization or logical sequence of ideas.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Second Section:** Please indicate how often the following activities take place in your current writing classes.

<table>
<thead>
<tr>
<th>II. Writing Class instruction</th>
<th>Never</th>
<th>Rarely</th>
<th>sometime</th>
<th>Often</th>
<th>Very</th>
<th>from time to time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Teachers of writing inform students of possible synonyms (words of same meaning) of vocabulary items.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>21. In writing classes, teachers discuss the antonyms (words of opposite meaning) of words.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>22. In writing classes, teachers explain family members of words that are encountered in reading or class activity, e.g., policy, politics, political, politicians.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>23. In writing classes, teachers give feedback on accurate use of words and lexical expressions.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>24. In my writing classes, students analyze and compare complex nouns in academic texts. For example, the noun in sentence A is more complex than that in B: A. The newly developed technical improvements in the educational system attracted the audience attention. (complex noun). B. The improvements attracted the audience attention. (simple noun)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>25. In my writing class, there are activities to explain how to expand a simple noun phrase into a complex long noun</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
Third Section

Please answer the following question:
1. What type of class activities do you think would help you improve your grammar and vocabulary knowledge in writing courses? (For example: analysis of academic texts, vocabulary focused activities, etc.)
2. Please help us better identify your needs by completing the following:
   - Gender: Male Female
   - ISU ID name:
   - Writing class section:
   - L1 or mother tongue:

If you have any questions or wish to receive the results of this questionnaire, feel free to email the researcher:
zalonazi@iastate.edu

THANK YOU FOR HELPING US UNDERSTAND YOUR NEEDS 😊
APPENDIX G. INTERVIEW PROTOCOL WITH ESL LEARNERS

Greetings:

Thank you for agreeing to participate in this study. My name is Zaha and I am the primary researcher of this study. Just a reminder, this study aims at assessing students’ overall satisfaction with the EPT placement decisions and their improvement in writing after ESL courses. To facilitate our note-taking, I would like to audio tape our conversations today. I am the only person who have access to this taped interview and recordings will eventually be destroyed after they are transcribed. Please sign the consent form. This document s indicates that (1) all information will be held confidential, (2) your participation is voluntary and that you will be granted $5 for your participation. You may of course stop at any time if you feel uncomfortable. Thank you again for agreeing to participate.

Short introduction:

You have been selected to speak with us today because you are enrolled in ESL writing courses. I have planned this interview to last no longer than 12 minutes. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

Questions:

Q1: Which writing level you are, 101B, 101C, 101D, your major and your first language?

Q2: How satisfied are you with the placement decisions of the EPT writing test?
Q3: How much you believe your writing ability has changed after taking ESL writing course?

Q4: In which areas you believe you have improved: vocabulary, grammar, organization?

Q5. What are the major problems you have in your writing? For example, are they related to grammar, vocabulary, organization?

Q6. Do you any suggestions for improving ESL writing courses?

We came to the end of this interview, thank you so much for sharing your views.
APPENDIX H. INFORMED CONSENT FOR ESL STUDENTS’ INTERVIEWS

Title of Study: VALIDITY ARGUMENT FOR EPT WRITTEN ARGUMENTATIVE ESSAYS

Investigator: Zaha alanazi

Introduction

The purpose of this study is to identify the linguistic needs pertinent to academic writing for 101B, 101C, 101D students. Particularly the study aims at examining learners’ development and awareness of academic writing relevant features. You are being invited to participate in this study because you are currently enrolled in one of the following ESL writing courses 101B, 101C, 101D and you have finished the post self-assessment survey of writing ability.

Description of Procedures

In this follow up interview, you will be asked some questions about your writing ability and your evaluation of ESL writing courses. The interview will be about 10 minutes.

Benefits

If you decide to participate in this study, you will receive $5 in appreciation of your time.

Participant Rights

Participating in this study is completely voluntary. You may choose not to take part in the study or to stop participating at any time, for any reason, without penalty or negative consequences. If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, (515) 294-3115.

Confidentiality

To ensure confidentiality to the extent permitted by law, the following measures will be taken:
no real names or any identity-relevant information will be released. Only pseudo names will be
used to discuss the results of this interview. You are encouraged to ask questions at any time
during this study. For further information about the study feel free to email the primary
investigator at: zalonazi@iastate.edu or the major professor: Bethany Gray at:
begray@iastate.edu.

**Consent and Authorization Provisions**

By signing this form, you indicate that you voluntarily agree to participate in this study, that the
study has been explained to you, that you have been given the time to read the document, and
that your questions have been satisfactorily answered.

Participant ISU ID ………………. signature………………. Date………………….
# APPENDIX I. LEARNERS’ SUGGESTIONS OF ACTIVITIES FOR IMPROVING WRITING ABILITY

<table>
<thead>
<tr>
<th>Level</th>
<th>Suggestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>analysis of academic texts</td>
</tr>
<tr>
<td>B</td>
<td>I prefer vocabulary focused activities.</td>
</tr>
<tr>
<td>B</td>
<td>Vocabulary focused activities</td>
</tr>
<tr>
<td>D</td>
<td>analysis and comparison with standard academic writings.</td>
</tr>
<tr>
<td>C</td>
<td>Kahoot games</td>
</tr>
<tr>
<td>B</td>
<td>discuss with classmates</td>
</tr>
<tr>
<td>C</td>
<td>Frequent advice to improve my writing by proofreading</td>
</tr>
<tr>
<td>D</td>
<td>Vocabulary related activities</td>
</tr>
<tr>
<td>C</td>
<td>Quizzes</td>
</tr>
<tr>
<td>D</td>
<td>Some physical activity related to the writing, for instance: Funny/serious statements read by students</td>
</tr>
<tr>
<td>C</td>
<td>More analysis of Academic texts</td>
</tr>
<tr>
<td>C</td>
<td>Analysis of work before submitting</td>
</tr>
<tr>
<td>B</td>
<td>More paragraph practice</td>
</tr>
<tr>
<td>D</td>
<td>analysis of academic texts and writing practice</td>
</tr>
<tr>
<td>D</td>
<td>more homework activity</td>
</tr>
<tr>
<td>C</td>
<td>Having discussion and interactive class</td>
</tr>
<tr>
<td>D</td>
<td>Analysis of academic texts</td>
</tr>
<tr>
<td>D</td>
<td>Analysis of academic text</td>
</tr>
<tr>
<td>C</td>
<td>read articles and discuss, watching videos explaining it.</td>
</tr>
<tr>
<td>D</td>
<td>Analysis of academic texts</td>
</tr>
<tr>
<td>D</td>
<td>I think study a model paper or focus on useful academic grammar and words</td>
</tr>
<tr>
<td>B</td>
<td>Analysis of academic texts</td>
</tr>
<tr>
<td>C</td>
<td>Reading interesting articles</td>
</tr>
<tr>
<td>D</td>
<td>let student to imitate one example article, and give them detailed feedback</td>
</tr>
<tr>
<td>B</td>
<td>academic vocabulary</td>
</tr>
<tr>
<td>2</td>
<td>vocabulary focused activities</td>
</tr>
</tbody>
</table>
APPENDIX J. IRB APPROVAL FOR ESL SURVEY AND INTERVIEWS

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 302
Ames, Iowa 50014
515 294-4566

Date: 2/9/2017
To: Zaha M Alonazi
1103 Delaware Ave #7
Ames, IA 50014

From: Office for Responsible Research
Title: Validation study of the EPT writing test
IRB ID: 17-042

Study Review Date: 2/9/2017

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

- (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey or interview procedures with adults or observation of public behavior where
  - Information obtained is recorded in such a manner that human subjects cannot be identified directly or through identifiers linked to the subjects; or
  - Any disclosure of the human subjects’ responses outside the research could not reasonably place the subject at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

The determination of exemption means that:
- You do not need to submit an application for annual continuing review.
- You must carry out the research as described in the IRB application. Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. Changes to key personnel must also be approved. The purpose of review is to determine if the project still meets the federal criteria for exemption.

Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

Detailed information about requirements for submission of modifications can be found on the Exempt Study Modification Form. A Personnel Change Form may be submitted when the only modification involves changes in study staff. If it is determined that exemption is no longer warranted, then an Application for Approval of Research Involving Humans Form will need to be submitted and approved before proceeding with data collection.

Please note that you must submit all research involving human participants for review. Only the IRB or designees may make the determination of exemption, even if you conduct a study in the future that is exactly like this study.

Please be aware that approval from other entities may also be needed. For example, access to data from private records (e.g., student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted.

Please don’t hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.
APPENDIX K. IRB APPROVAL FOR ESL LEARNERS’ SURVEY AND INTERVIEWS

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
4400 Lincoln Way, Suite 300
Ames, Iowa 50014
515-294-4566

Date: 07/31/2018
To: Zaha M Alonazi
From: Office for Responsible Research
Title: VALIDITY ARGUMENT FOR EPT WRITTEN ARGUMENTATIVE ESSAYS
IRB ID: 18-299
Submission Type: Initial Submission
Exemption Date: 07/31/2018

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

2. Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement, survey procedures, interview procedures, or observations of public behavior, unless (i) Information obtained is recorded in such a manner that human subjects can be identified, and (ii) Any disclosure of the human subjects’ responses outside the research could reasonably place the subject at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation.

The determination of exemption means that:

• You do not need to submit an application for annual continuing review.

• You must carry out the research as described in the IRB application. Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. The purpose of review is to determine if the project still meets the federal criteria for exemption.

In addition, changes to key personnel must receive prior approval.

Detailed information about requirements for submission of modifications can be found on our website. For modifications that require prior approval, an amendment to the most recent IRB application must be submitted in IRBManager. A determination of exemption or approval from the IRB must be granted before implementing the proposed changes.

Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

IRB 03/2019