

The effects of discipline-specific background knowledge on reading comprehension

by

Svetlana Chigayeva

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF ARTS

Major: English (Teaching English as a Second Language/Applied Linguistics)

Major Professor: Dan Douglas

Iowa State University

Ames, Iowa

2001

Copyright © Svetlana Chigayeva, 2001. All rights reserved.

Graduate College
Iowa State University

This is to certify that the Master's thesis of

Svetlana Chigayeva

has met the thesis requirements of Iowa State University

Signature redacted for privacy

Major Professor

Signature redacted for privacy

For the Major Program

Signature redacted for privacy

For the Graduate College

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vii
CHAPTER 1. INTRODUCTION	1
Purpose of the Study	2
Rationale	2
Research Question	3
Organization of the Study	3
CHAPTER 2. LITERATURE REVIEW	5
Reading Models	5
Schema Theory	7
Discipline-Specific BGK	8
Summary	18
CHAPTER 3. METHODOLOGY	21
Research Instruments	21
Participants	27
Procedures	29
Analysis	32
Summary	35
CHAPTER 4. RESULTS	36
Descriptive Statistics	36
Analysis of Variance across Groups and Passages	38
Item Bias	45
Passage Ease/Difficulty Perceptions	48
Passage Familiarity	49
Previous Education	50
Summary	50
CHAPTER 5. DISCUSSION AND CONCLUSION	52
Summary	52
Insights for Researchers, Teachers, and Test Developers	54
Limitations of the Study	57
Future Research	58
APPENDIX A: READING SUBTEST	59
APPENDIX B: HUMAN SUBJECTS APPROVAL	69

APPENDIX C: CONSENT FORM	70
APPENDIX D: STUDENTS' QUESTIONNAIRE	71
APPENDIX E: RATERS' QUESTIONNAIRE	74
APPENDIX F: SAMPLES OF NORMAL PROBABILITY PLOTS	78
APPENDIX G: MANOVA OUTPUT	81
APPENDIX H: TOPIC FAMILIARITY	84
APPENDIX I: PREVIOUS EDUCATION	86
BIBLIOGRAPHY	88
ACKNOWLEDGEMENTS	90

LIST OF TABLES

Table 2.1: Studies of BGK and Its Effect on Reading Comprehension	19
Table 3.1: Readability Indicators (Microsoft Word)	22
Table 3.2: Verbal Structures	24
Table 3.3: Test Items as Coded by the Raters	25
Table 3.4: Distribution of Students between the Colleges	27
Table 3.5: Distribution of Students between Broader Areas of Study	28
Table 3.6: Students' First Languages	29
Table 3.7: The Testing Sessions and the Order of Subtest Administration	30
Table 3.8: Conversion Table of Passage Scores	33
Table 3.9: Summary Table	35
Table 4.1: Distribution of Scores between Passages	36
Table 4.2: Table of Means (%) across Groups and Passages	37
Table 4.3: Logistic Regression Model for Score as a Function of Passage within Group BSSS	39
Table 4.4: Logistic Regression Model for Score as a Function of Passage within Group PST	40
Table 4.5: Logistic Regression Model for Score as a Function of Passage within Group LMS	40
Table 4.6: Logistic Regression Model for P(BSSS) Score as a Function of Group	42
Table 4.7: Logistic Regression Model for P(PST) Score as a Function of Group	43
Table 4.8: Logistic Regression Model for P(LMS) Score as a Function of Group	43
Table 4.9: Item Facility Indices across Groups for Each Passage	45
Table 4.10: Lowest Indices between the Groups (with Item Codes in Parentheses)	46
Table 4.11: Ease of the Passages	48
Table 4.12: Summary Table	51

LIST OF FIGURES

Figure 2.1: Methodology Employed by Previous Studies

9

CHAPTER 1. INRODUCTION

A large university such as Iowa State attracts international students not only of various ethnic backgrounds, but of various educational backgrounds as well. Here, students come to get their undergraduate and graduate degrees in diverse fields like engineering, agriculture, sociology, history, etc. Like most other US universities, ISU offers special English courses for international students, and the tool used to determine which class is necessary for a particular student is the English Placement Test. One part of this test is intended to measure students' reading proficiency, and this particular subtest is the interest of the current study.

It has long been suggested that reading can be either facilitated or hindered by students' background knowledge (be that cultural, pragmatic, structural, or discipline-specific knowledge). In the field of English for Specific Purposes (ESP), research into the effects of discipline-specific background knowledge started with the attempt to answer the question of the effectiveness of ESP tests in EAP settings (English for Academic Purposes). The results were somewhat contradictory, and no conclusive evidence was produced for the use of ESP tests. With the works of Hale (1988) and Clapham (1996), the picture slightly changed and new dimensions have been added to the research into the effect of discipline-specific knowledge on students' test performance. In her work, Clapham showed that there are certain levels of linguistic proficiency at which students start using their background knowledge, and more importantly, there are levels of text specificity under which the use of discipline-specific background knowledge seems unnecessary.

The Reading Subtest of the English Placement Test offered by ISU consists of passages whose content is quite general and the level of text specificity is low. However, there are still elements (such as topics, technical vocabulary, formulas, etc) that could relate the passages to specific fields of study. It will be of interest, therefore, to analyze whether students' performance is influenced by their particular discipline-specific background knowledge when they deal with the passages of the Reading Subtest of the English Placement Test.

Purposes of the Study

The broadest purpose of this study, therefore, is to examine whether students use discipline-specific background knowledge when reading passages of low specificity level, such as the passages of the ISU English Placement Test. A more specific purpose is to analyze whether there is a passage in the English Placement Test that is more specific than others, and whether dealing with it requires utilizing background knowledge. The final major purpose is to analyze whether discipline-specific background knowledge, as reflected through students' previous education and familiarity with passage topics, influences reading test performance.

Rationale

Knowing whether the reading test performance is influenced by the students' discipline-specific knowledge will help language researchers, teachers, and test developers on two significant levels. Locally, ISU test users will be able to get a better idea of the effectiveness of the test and raise the level of its use and applicability. ISU ESL instructors might possibly get a better understanding of why students in their reading courses score low on the Reading Subtest and devise effective strategies to use in reading instruction.

On a broader level, researchers might benefit from this study by getting more insights into the complexity of discipline-specific knowledge and its effect on reading test performance. They will see that testing the effects of discipline-specific background knowledge is not as straightforward as it may seem. Because it can be acquired through different ways, several tools of measuring background knowledge should be employed. Moreover, since passages within the same test can be of different specificity levels, researchers will be able to see how to assess the effect of passage specificity on reading test performance. Since the issue of methodologies will be in the center of my attention, researchers will benefit additionally by acquiring a framework for examining the methodology of their own studies and studies performed by others.

In addition to general language researchers, this study will benefit language test developers by providing insights into how "general" passages of general language tests are

and how “specific” they can get. The question of whether tests of English for General Purposes or tests of English for Specific Purposes should be used as placement tests employed by universities diverse in academic programs will be addressed later on in the study.

Research Questions

The following are the specific research questions this study is intended to answer:

- RQ1: Do students tend to perform better on the passages related to their field of study?
- RQ2: Are there test tasks that seem to be problematic for specific areas of study?
- RQ3: Do groups differ in perceptions of passage difficulty?
- RQ4: Does a reported familiarity with the topic of a passage help students perform better on the passage?
- RQ5: Does students’ reported educational history influence their reading test performance?

Organization of the Study

The sections above gave a short overview of the reasons for this study and outlined the research questions to be answered. The next chapter of this work, Literature Review, will be devoted to an extensive overview of what research has been done into the relationship between reading test performance and students’ discipline-specific background knowledge. First, it will give some information on the theories that informed the research into background knowledge and give the history of schema theory, in particular. Then, it will divide the studies of discipline-specific background knowledge into several types based on the methodology and the kind of tools used to collect data and show some aspects of the studies that could be improved in the further research.

Chapter 3, Methodology, will present the methodology used in the current study and will include parts on the subjects, tools, and analysis methods involved. Chapter 4, Results, will summarize the results of the analysis and provide some explanation of the possible causes for obtaining certain results. Finally, Chapter 5, Discussion and Conclusion, will discuss the findings in more detail by relating them to the research questions, as well as discussing possible benefits to researchers, teachers, and test developers and by focusing on

the aspects of the study that could be improved in the future. It will finally discuss the place of the study within the existing research on the topic.

The last part of the work will contain appendices with such information as the Reading Subtest passages and items, Human Subjects Approval Form, the tools used for additional data collection, and various statistical outputs.

CHAPTER 2. LITERATURE REVIEW

Much research has been done in the area of reading and its components. This chapter will focus on the particular theories and studies that have contributed to the investigation of discipline-specific background knowledge (BGK, hereafter) and its effect on reading comprehension. More specifically, the chapter will be divided into 1) Reading Models—an overview of some theories that attempt to provide a comprehensive view of the reading process and its components; 2) Schema Theory—a brief description of the theory that influenced some of the reading models, especially those that took into account BGK; 3) Discipline-Specific BGK—an in-depth review of the most important studies that looked into the relationship between discipline-specific BGK and students' performance on reading tests; and 4) Summary—major conclusions of the mentioned reading models, findings and dominant themes of the reviewed studies, and a summary table of the methodologies used in the studies.

Reading Models

The process and product of reading have long been the focus of attention for specialists of different fields: linguistics, education, psychology, artificial intelligence, etc. Even though much has been investigated, little consensus is reached on the most fundamental question: What is reading? Traditionally, there have been three major models of reading adopted by linguists and psychologists, and the major difference between these models lies in their interpretation of the process and components of reading ability.

The first model is comprised of theories claiming that the process of reading depends mainly on the text. The basic assumption of this model is that the reader begins the reading process by analyzing small units of the text, which are subsequently built into the understanding of the global meaning of the text. Theories comprising this model are most often called “bottom-up” theories, and the main proponents of this trend are Gough (1972), LaBerge and Samuels (1974), and Carver (1977).

According to Gough (1972), reading is a linear process that develops in the following way: letter decoding, word analysis, sentence processing, and understanding the text. The

theory is one-sided and provides little, if any, acknowledgement of the higher level processes. LaBerge and Samuels' model (1974) includes automaticity as its central concept. It claims that the task of reading, which consists of decoding and comprehension, requires a lot of attention. As soon as decoding becomes automated, e.g. through practice, it demands less attention, thus leaving more space for comprehension. Reading, according to this theory, consists of three stages: visual memory, phonological memory, and the response system. Although the automaticity can allow skipping some of the stages, the general assumption is that the stages are linear and hierarchical. Even though the theory suggests that the reading process can be "affected by the familiarity of the materials," the fixed sequence of stages makes the theory bottom-up (Kamil, 1986, p. 77). Carver (1977) proposes a theory of reading which has the prediction of the reading behaviour as its focus. "Reading is the process of obtaining meaning from text such that every 'thought' is immediately comprehended" (Kamil, 1986, p. 79). Carver believes that every word has to be comprehended sequentially in order for the whole text to be understood. This major claim makes the theory "bottom-up".

Unlike the bottom-up theories, theories of the second category—the top-down model—propose that readers initiate the reading process by guessing the meaning of the text due to the background knowledge they possess. Influenced by Barlett's schema theory, the top-down theories do not reject the usefulness of bottom-up processes but suggest that the latter help the reader in confirming their guesses. Two of the best-known supporters of this model are Goodman (1968) and Smith (1988).

The most significant characteristic of Goodman's theory (1968), which treats "reading as a psycholinguistic guessing game" is that it minimizes the influence of print and phonics on the reading process by emphasising the possibility of learners' reliance on existing syntactic and semantic structures. The learner's choice of strategies depends on the level of proficiency; otherwise, the higher the level of student proficiency the fewer bottom-up processes are involved and the more guessing takes place. Smith (1988) takes the ideas proposed by Goodman slightly further and says that "[r]eaders can go straight to the meaning of the text by means of prediction. Reading is not a matter of identifying word after word" (p. 285).

The third category of reading models is built as a critical response to the two models and proposes a compromise between them. According to the theorists belonging to this school, reader guesses facilitate initial decoding of the text; while decoding facilitates guessing. Or, top-down processes influence bottom-up processes, and vice versa. The best known theory of this model, the interactive model, is the one put forward by Rumelhart (1977). According to the interactive interpretation of this theory, reading is not a linear process and different levels of processing, both bottom-up and top-down, are involved in the construction of the meaning of the text at the same time.

Both top-down and interactive models of reading assume that the reader has an existing knowledge of the world and the capacity to remember it. Fundamental to these models, therefore, is the idea that there should be a specialized system that would store and retrieve information when necessary. This idea was first developed by the schema theory of reading where schemata represent a learner's structured framework of knowledge of the world.

Schema Theory

Schema theory was born as an attempt to explain why "what readers know effect what they understand" (Alderson, 2000, p. 33). It started with the work of Barlett (1932) who described several memory-related experiments aimed at analysing which parts of the text were easier to remember. It turned out that the prior knowledge the readers had influenced what they remembered of the text. Not only that, but it also influenced the understanding of the text and the processes involved in reading. Later, to account for the results of his study, Barlett proposed a "theory of remembering," the central part of which was based on the notion of "schema". According to Barlett (1932),

'Schema' refers to an active organisation of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response. That is, whenever there is any order or regularity of behaviour, a particular response is possible only because it is related to other similar responses which have been serially organized.

(as qtd. in Clapham, 1996, p.17)

After Barlett's work, numerous theories have been constructed on the basis of schema theory (e.g., Minsky's (1975) frame system theory, Schank and Abelson's (1977) artificial intelligence theories, etc.). Even though they all propose something new and different in their interpretations of schemata, the common point between them is their claim that the "state of reader's knowledge influences process, product and recall" (Alderson, 2000, p. 34). Other components common to all schemata theories are expressed in the following:

First of all, readers' knowledge is stored not in lists but rather in hierarchies. Within these hierarchies are schemata which are embedded in other schemata, and which themselves contain subschemata. These schemata vary in their levels of abstraction, and represent all sorts of knowledge, such as objects, academic topics, rules, events, routines and social situations. They represent knowledge, rather than definition, so they are not language based, but are symbolic representations of knowledge which may be used for understanding language. Schemata are not static, but fluid; they change according to the input. Schemata can be refined and the new ones can be developed by the process of accommodation, that is the modification of previous schemata in the light of new information.

(Clapham, 1996, pp. 23-24)

There have been disagreements on which levels of language processing are covered by schema theory. Generally, it is assumed to account for the higher level processing of texts; however, some researchers propose a wider application of the theory, for example to account for the verb and narrative schemata (van Dijk and Kintsch, 1983). Others, such as Adams and Collins (1979), expect schema theory to account for all levels of reading processing including both top-down and bottom-up processes. Schema theory has influenced at least some research into the effects of background knowledge on reading.

Discipline-Specific BGK

By now, numerous studies have been carried out to show the effects of background knowledge (BGK) on reading comprehension. A theoretical distinction should be made, however, as to what kind of BGK the studies have investigated. Carrell (1987) distinguishes between formal schemata, "background knowledge of the rhetorical structures of different

types of texts," and content schemata, "background knowledge of the content area of a text". She further divides BGK into "culture-specific" and "cross-cultural" knowledge (Carrell, 1987, p. 183). This section of the current chapter is centered around the studies that have looked at discipline-specific BGK, the category which could fall under the heading of "cross-cultural" content schemata of Carrell's classification. Discipline-specific BGK is assumed to include BGK developed as a result of studying a particular field of art or science.

The major emphasis in reviewing the studies will be the question of methodologies. I argue that only those studies that employ multiple tools for collecting data can be considered effective in analysing the relationship between BGK and reading performance. Figure 2.1. presents a summary of the methodologies employed by the studies under analysis.

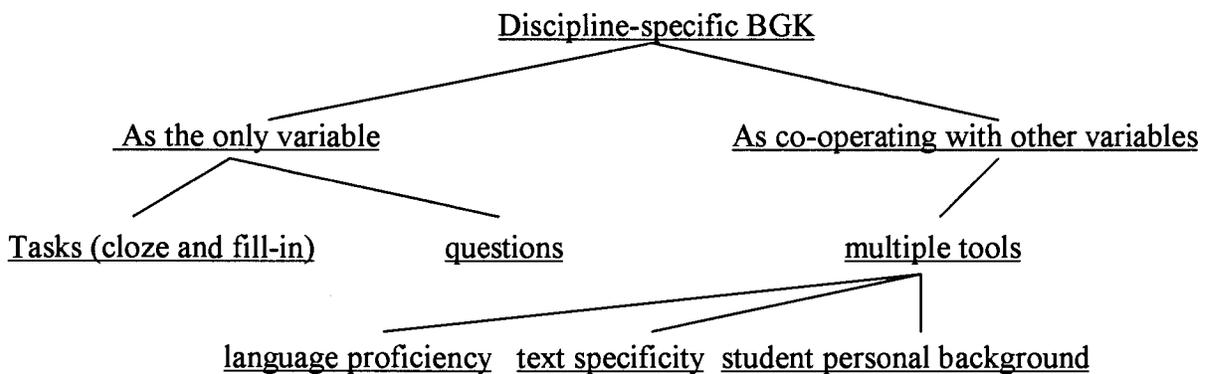


Figure 2.1: Methodologies Employed in Previous Studies

According to this schema, the studies of discipline-specific BGK can be divided into those that analyze discipline-specific BGK as the only independent variable explaining reading comprehension and those that take into account other independent variables as well. Studies belonging to the first group employ one major measurement instrument, most often consisting of tasks or questions, while studies belonging to the second group of studies employ multiple tools and analyses which are used to account for all the independent variables, such as general language proficiency, text specificity, etc.

In the following part of the chapter, I will attempt to analyze several studies in detail. For each study, I will first explain the methods used and the conclusions arrived at, then I

will switch the attention to the potential problems raised by each method. This procedure will help me determine some of the most adequate ways of measuring the effect of discipline-specific BGK on reading test performance and to shape the methodology used in this study.

Discipline-Specific BGK as the Only Analyzed Explanatory Variable

Among the earliest well-known studies in the field of discipline-specific BGK and its influence on reading comprehension were those performed by Alderson and Urquhart (1983, 1984, 1985) who conducted three studies one after another. The first two studies attempted to measure the effect of students' academic discipline on their performance on ESP reading tests. The subjects for the first study were drawn from areas of Development Administration and Finance (DAF), Engineering (ENG), Liberal Arts (LA), and Science and Mathematics (SM). They were all non-native speakers, (NNSs), of English taking a course on study skills at the University of Aston. The five texts comprising the reading test represented two DAF texts, two ENG texts, and one "general" text. The specialized texts were taken from textbooks and professional publications, while the general text was chosen from reading suggestions for American high school students. The task was gap-filling, and acceptable responses were counted as correct.

The results for the first study showed that ENG students outperformed DAF on ENG texts, while DAF students outperformed ENG on DAF texts. The other two groups, LA and SM, were not analyzed extensively due to the insufficient number of students in them. The insufficient number of participants (from 15 to 5) was the overall drawback of the study which made the calculations of significance irrelevant. Another drawback of the study was the failure of the researchers to explain the role of linguistic proficiency in the task results: even though a measure of linguistic proficiency was administered to students in the beginning of the experiment, no analysis of it was provided.

The second study by Alderson and Urquhart (1984) was a partial replication of the first one and involved more students (22 to 38 per group). In it, the first group of students included DAF and Economics (DAFE), while the other three remained the same (ENG, LA, and SM). This time, students from two universities, University of Aston and University of Lancaster, were involved. The texts used in the study were the same as in the first study, but

the tasks included short answers as well as gap-filling.

The results for the gap-filling task were such that DAFE scored higher on the DAFE texts, but ENG did not follow the expectation and did not outperform DAFE on the ENG texts. On the short answer tasks, DAFE followed the same pattern, but ENG performed on the same level as DAFE on the ENG text. Several conclusions were drawn from these results. First of all, the fact that ENG students performed differently on the two tasks indicated a low correlation between the scores on the two test tasks. Second, the fact that the ENG group did not perform better than the DAFE group on the ENG texts indicated other factors that could have played a significant role in reading comprehension. Linguistic proficiency could have been one of those factors, and even though Alderson and Urquhart possessed students' scores on such measures of general English proficiency as English Language Battery and ELTS, linguistic proficiency as such was not analysed in terms of its possible influence on reading comprehension.

In both of their studies mentioned above, Alderson and Urquhart (1983, 1984) employed gap-filling tests as their primary tool. The relevance of the latter to testing reading comprehension is questioned.

Cloze tests provided an alternative method of measuring reading comprehension. One of such experiments was carried out by Koh (1985) who was inspired by Widdowson's idea that success of reading comprehension depends on writer and reader sharing knowledge of different kinds (1985:376). He administered his test to three groups of students (60 each) attending the National University of Singapore who had come from either Chinese-medium (CM) or English-medium (EM) schools. The first group consisted of CM business students, the second of CM science students, and the last of EM science students. Thus, the overall English proficiency was assumed to be correlated to the language medium of the school that the students had previously attended, i.e. the English language proficiency level of CM students was assumed to be lower than that of EM students.

The students were administered a cloze test consisting of four passages which differed in content and specificity. Two texts were discipline-specific and came from business and science, while the other two were considered discipline-neutral since they came from areas supposedly unfamiliar to either business or science students—from politics and

history. The cloze test had no time limits, and only exact-word replacements were accepted as correct answers.

The results of the experiment showed that science EM students did better than the other students on the history and politics texts, the fact that contributed to the supposition that the group was more proficient in general English than the other two groups. Business CM students performed better than science CM students on the same texts (history and politics), which led to the conclusion that science CM students were even less proficient than business CM students. On the business text, business CM students did as well as science EM students and better than science CM students. The first finding can lead to the idea that business SM students substitute their BGK for the lack of linguistic proficiency. The second finding can mean that either science CM students were lower in English proficiency (as had been hypothesised by the researchers previously) or that they did not possess appropriate BGK to compensate for that lack of language proficiency. On the science text, science CM students performed better than business CM students.

Overall, the experiment suggested that discipline-specific BGK affects cloze-test scores; however, several drawbacks should be pointed out. First of all, as it turned out later, some business CM students had studied science, and this could have affected their performance on the science text. The students' previous educational background, therefore, should have been controlled for. Second, the so-called discipline-neutral texts can be more familiar to some students due to their field of study than to others. More specifically, the history and politics texts could have been more familiar to business students because of their liberal arts background. However, no research on the degree of text specificity was conducted. Third, even though some hypotheses were developed during the study about the possible linguistic proficiency levels of the students, no method was used to quantify the hypotheses. Finally, the use of cloze as a measure of reading comprehension has been strongly criticized by numerous linguists. Farhady (1983), for example, claims that cloze tests measure grammar and vocabulary rather than reading comprehension, and Klein-Braley (1983) suggests that cloze tests must be thoroughly examined in terms of their validity.

Shoham et al. (1987) rejected the use of cloze in the studies of the role of BGK. Instead, for their own study in Israel, they used a set of different kinds of questions. The test

was designed to investigate the relevance of the issue of students' discipline-specific BGK in EFL reading comprehension tests. It involved 185 students from three areas, Science and Technology, Biology, and Humanities and Social Sciences. The subjects were tested on three texts related to their respective fields of study. Thus, three passages were obtained from professional publications and a set of questions was developed for each one of them.

The results did not show any significant interaction between the texts and the students' disciplines. The first explanation for that might be the fact that some of the passages were not as specific as the others. The second explanation can be that students possessed different levels of linguistic proficiency which, in turn, affected their level of reading comprehension more than their BGK. Linguistic proficiency measures were provided in the form of students' scores on the national EFL placement test; however, no information is available about the time of the placement test administration, and no analysis was conducted on the possible effect of linguistic proficiency on reading comprehension.

Linguistic Proficiency as Another Variable

All of the studies above failed to take into account factors other than discipline-specific BGK when analyzing its influence on reading comprehension. Even though most of them had some indicator of students' linguistic proficiency, none of them integrated it into the analysis of reading comprehension. In contrast, the first two studies below analyzed the effect of students' general linguistic proficiency on reading comprehension but failed to analyze the interaction of discipline-specific BGK and linguistic proficiency. The study after that is the first one that filled in the gap and did what others had overlooked—it analyzed the interaction between the two independent variables.

An experiment conducted by Erickson and Molloy (1983) was a pilot study of an ESP engineering test that involved both native (NS) and non-native (NNS) speakers, all voluntarily participating in the study. Both engineering and non-engineering undergraduate students were represented in each group. The reading comprehension test did not have time limits and was based on the passage taken from the engineering book used by the university in teaching freshmen engineering students. The 38 items consisted of “language” and specifically “engineering” items.

NSs were expected to perform better than NNSs on language items, and engineering students were expected to do better than non-engineering students on engineering items. The results of the study showed that these expectations proved right: NSs did better than NNSs overall and especially on language items; engineering NSs did better than non-engineering NSs, and the same pattern was found for the NNSs. However, it was also found that engineering NSs outscored non-engineering NSs on language items. This is a puzzling finding and might cast some doubt on the validity of the "language" items used in the test. The authors hypothesized that being familiar with the content of the passage can help students to perform well even on language items.

In comparison to Erickson and Molloy (1983), Alderson and Urquhart (1985) devised their third study to include a separate measure of linguistic proficiency and used the same four groups of students of the same universities, U of Aston and U of Lancaster as in studies 1 and 2. However, since their previous studies had shown that SM (Science and Mathematics) and ENG (Engineering) students behaved in similar ways, they were combined into one group, SMENG. The test involved texts from English Language Testing Service (ELTS) modules of Social Studies, Technology, and General Academic. In addition, a reading (G1) and a listening (G2) tests of general context were used to measure students' general English language proficiency. The tests were given as part of an instructional program, i.e. the students were not aware that they were participating in an experiments. On their arrival, students were given G1, G2, and the social science module as their placement test. Six weeks later, they were administered the technology module. The general academic module was presented as an exit test. The test tasks ranged from direct/factual questions based on the texts to inferential questions and overview questions.

The results of the study were similar to the previous two in showing the relationship between discipline-specific BGK and test performance and in failing to show the stable relationship. To be more specific, the experiment showed that the LA group outperformed the other two on everything but the technology module and scored highest on G1 and G2. On the Technology module, the SMENG group performed on the same level as the Liberal Arts (LA) group, which can show that the former used their BGK to compensate for their linguistic proficiency shown to be low by G1 and G2. The background effect, however, did

not help the last group, Development Administration, Finance, and Economics (DAFE), which failed to outperform the rest on the social science module. In general, the results of the study showed that discipline-specific BGK can play an important role in test performance, however, the effect does not appear to be consistent.

Brown (1987) was the first to account for the interaction effect of linguistic proficiency and discipline-specific BGK. In his study of "variables which may contribute to the ability to read engineering tests," he administered both an ESP test and a test of general linguistic proficiency and involved students of two language backgrounds, English and Chinese. The four groups of students (29 each) were classified as engineering NSs, engineering NNSs, non-engineering NSs, and non-engineering NNSs. The engineering reading test consisted of 60 items and was a revised and validated version of the test mentioned earlier (Erickson & Molloy, 1983). The three passages in it were based on the topic classified as discipline-specific by engineering professors, and the items were again divided into "linguistic" and "engineering" items. The test of general English language proficiency consisted of a 50-item cloze test which was based on a "relatively neutral topic". It was scored by using the acceptable-word method rather than the exact-word method. Both of the procedures were timed and the students had to finish within certain limits.

The results of the study showed that the reading test was effective in distinguishing between engineering and non-engineering students, i.e. engineering students outperformed non-engineers. It was also found that non-engineering NSs did better than engineering NNSs. In the case of NNSs, the test of general English proficiency was also taken into account. Based on this, it was found that 62% of the variance in scores on the engineering test was accounted for by general language proficiency rather than by components of specific BGK.

The results of the study, overall, showed that discipline-specific BGK played a role in reading comprehension, although linguistic proficiency was found to be more important. Several points should be mentioned in regard to this study. First, the cloze test of language proficiency should have been administered to NSs for validation because it is not clear if it really measured linguistic proficiency. In addition, the students were involved on a voluntary basis, which brings the question of motivation and its influence on reading and the reliability of scores.

Other Variables

The studies in the section above took into account two independent variables when explaining students' reading comprehension. The studies below go a step further and involve several other variables that may provide more valid results. Hale (1988) analyzed student performance on passages from TOEFL, a test of General Academic English, and classified the passages as belonging to Humanities/Social Sciences and Biological/Physical sciences. This study was supposed to be more comprehensive than any other previous study due to several reasons. First of all, the study involved 32,467 test-takers, the entire test taker population of four TOEFL administrations. Second, all regions of the world were represented by the sample. Third, 21 reading passages were included in the experiment, which means that several passages per discipline were used, which in turn was supposed to improve the reliability of the study.

The passages and the test takers were classified into four major categories: humanities, social sciences, biological sciences, and physical sciences. The principal analysis, however, required that these subdivisions be combined into the humanities/social sciences group and the biological/physical sciences group. Thus, the analysis was performed on two different levels: categories and groups.

The statistical analysis included the analysis of variance for each test form. The dependent variable in each analysis was "the difference between humanities/social sciences passages and biological/physical sciences passages." The independent variables included sex, region of origin, and the comparison between student groups, "humanities/social sciences students vs. biological/physical sciences students" (Hale, 1988, p.19). The most significant effect was found for the student groups, where it was obvious that students in humanities/social sciences outperformed students in biological/physical sciences on texts related to humanities and social sciences and vice versa. The overall results of this study suggested a strong effect of the combination of students' area of study and the nature of the passages on students' reading performance.

In his conclusion, Hale suggested that passages can be different in the extent of subject specificity, and some items can discriminate against certain groups of students. These factors, as he hypothesized, can influence the results of the studies into discipline-specific

BGK and reading comprehension and should be controlled for in future studies. Clapham (1996) did exactly what Hale (1988) had suggested and conducted a study that is the most recent and probably the most comprehensive in the field of discipline-specific BKG. It was a part of International English Language Service (IELTS) validation and involved both a pilot study and the main study. The pilot study aimed at examining just one question: whether student scores were significantly better on the texts related to their disciplines. No conclusive results were obtained about the effects of BGK on student performance on the basis of this study because no background information about students themselves was known. For the main study, it was decided that several steps needed to be taken. First of all, a content analysis of the texts and items had to be carried out, and each passage classified according to its specificity; more information had to be collected about students' background; and students had to be given the same test of language proficiency in order to place them into different levels.

It was found that discipline-specific BGK affected reading comprehension, but the results depended on the specificity of the reading texts—the more specific the passage was, the higher the effect of discipline-specific BGK was. Thus, a level of passage specificity was hypothesized at which the effects of BGK could be most pronounced. The two multiple regression analyses, which had language proficiency and variables relating to BGK as the independent variables, showed that language proficiency was a more important factor in reading comprehension than discipline-specific BGK. The analysis of linguistic proficiency also pointed out that there was level of language proficiency below which students were unable to use their BGK. The overall results of Clapham's study seem to be promising and require replications.

Major Findings/Dominant Themes

The review of all the studies above shows that research methodologies differed in many ways. Some of the studies involved NSs along with NNSs (Erickson & Molloy, 1983), while others were restricted to the pool of NNSs only (Alderson & Urquhart, 1983, 1984, 1985). Many of them examined graduate as well as undergraduate students (Clapham, 1996), while others studied graduate students only (Hale, 1988). The nature of L2 teaching was also

different across the studies: while some of them were conducted in EFL settings (Shoham et al., 1987), others were done in the ESL setting (Alderson & Urquhart, 1985). Many other details were quite different across the studies; however, none of them was as important as the method involved to collect data and additional consideration of variables other than discipline-specific BGK. Table 2.1 summarizes the methodologies and the conclusions the studies arrived at.

The following general conclusions can be drawn from the studies:

1. Discipline-specific background knowledge has an effect on reading comprehension; however, the results are not consistent either across the studies or even within single studies;
2. Linguistic proficiency has a greater effect on reading comprehension than discipline-specific background knowledge; however, not every study took this factor into account;
3. Several other factors, especially text specificity and learner background, were also found to be significantly related to reading comprehension; however, replications of the studies that investigated these factors need to be performed.

Summary

In this chapter, I have provided an overview of some theories of reading, described the essence of schema theory, and analyzed the methodologies employed by the previous studies of discipline-specific BGK and its influence on reading comprehension. The next chapter will offer an extensive account of the methodology used in the present study.

Table 2.1: Studies of Discipline-Specific BGK and Its Effect on Reading Comprehension

Study	Major data collection tool/ statistical analysis	Additional factors analyzed	Additional data collection tools	Results	Drawbacks
Alderson & Urquhart (1983)	Gap-filling tasks / pairwise t-test			Students did better on the texts related to their discipline	Small number of students; no explanation of the role of linguistic proficiency
Alderson & Urquhart (1984)	Gap-filling tasks and short answers to the same texts used in the first study / pairwise test			BGK effect was not consistent; the correlation between the two kinds of tasks was low; and other factors were suggested to play an important role.	Low correlation between the two tasks; the scores of the language tests are not incorporated
Koh (1985)	Cloze test, with exact word replacements being counted as right answers / analysis of variance			Discipline-specific BGK was found to be a necessary but not the only condition for reading comprehension.	Unproved assumption of the correlation between the type of school and linguistic proficiency; text specificity is not taken into account
Alderson & Urquhart (1985)	Texts of ELTS modules accompanied by questions ranging from direct to inferential and overview questions / pairwise test	General English proficiency	Reading and listening tests of general context	The relationship between discipline-specific BGK and reading test performance was not stable. Other factors might be of significant importance	No interaction effect between language proficiency and BGK
Shoham et al. (1987)	Texts followed by comprehension, reference and vocabulary questions / three-way ANOVA	Linguistic proficiency	Scores from Nationwide EFL Placement Test	The interaction between the texts and discipline-specific BGK was not as significant as had been hypothesized .	Texts of various degree of specificity

Table 2.1: Continued

Erickson & Molloy (1983)	Texts accompanied by multiple-choice questions / analysis of variance	Language proficiency	“Language” items were included into the test along with “engineering” items. Native speakers were used as a control group.	Discipline-specific BGK facilitates reading comprehension.	No administration of the cloze test to NSs; voluntary bases of participation
Brown (1987)	Validated version of the test used in Erickson & Molloy (1983)	Linguistic proficiency	Linguistic items of the main test as well as a separate cloze test of English proficiency	Great portion of the variation in scores (62% in this study) is accounted by linguistic proficiency rather than discipline-specific BGK.	
Hale (1988)	Reading tests of TOEFL / analysis of variance	Subjects’ gender, region of origin, linguistic proficiency	TOEFL	The relationship between students’ discipline of study and the nature of the passage is found to be significant.	
Clampham (1996)	IELTS tests/ analysis of variance, multiple regression	Linguistic proficiency, student educational background, passage specificity	“General” grammar test and two questionnaires	Discipline-specific BGK as well as passage specificity and student educational background were found to effect reading comprehension.	

CHAPTER 3. METHODOLOGY

This chapter reports on the methodology used in the study and is divided into 1) Research Instruments—the English Placement Test and the two questionnaires used in the study; 2) Participants—the subjects who took the English Placement Test and the raters who analyzed the Reading Test; 3) Procedures—the progression of the study; 4) Analysis—the statistical methods used to examine the data; and 5) Summary—chapter highlights in one table.

Research Instruments

The following research tools were used in this study: the English Placement Test, the Students' Questionnaire, and the Raters' Questionnaire.

The English Placement Test (EPT)

The major tool used in this study was the English Placement Test (EPT) offered to all international students newly arriving at Iowa State University (ISU). Its major purpose is to place students with specific reading, listening, and writing needs into appropriate classes. The three major abilities that the test is intended to measure are listening, reading, and writing, and each of the three subtests of the test is devised to measure a particular skill.

The test has not been subjected to a detailed analysis, either quantitative or qualitative. No specifications are available either. It is used only at ISU and serves as an institutional way of distinguishing the students with the lowest English ability from everybody else.

The Reading Subtest

There are two existing version of the Reading Subtest; however, the department has been using only one of them for the last couple of years. Therefore, the results of this study are based only on one version of the subtest. The Reading Subtest consists of three passages and thirty questions. Passages 1 and 2 are followed by 7 questions each, while passage 3 is followed by the remaining 16 questions. Ideally, it would have been nice to have a passage

with its content specifically related to a particular area of study to investigate how each group of students performs on the passage related to their field of study. However, this is not the case with the Reading Test: the raters (see more information on them below) did not think the passages were particularly specific, though they were still able to classify them as relating more to one area rather than the others. They classified the first passage, P(PST), as belonging to the area of “Physical Science and Technology”, the second as “Business Studies and Social Sciences” (P(BSSS)), and the third passage as “Life and Medical sciences” (P(LMS)). The complete texts are provided in Appendix A.

The Reading Passages. The passages of the Reading Subtest were first analyzed for linguistic differences. This was done with the realization that the differences could influence the participants’ scores, and therefore, the results of the study. Table 3.1 presents the readability statistics calculated by “Word Count”, a Microsoft Word 2000 tool.

Table 3.1: Readability Indicators (Microsoft Word)

	P(BSSS)	P(PST)	P(LMS)
Words	463	621	479
Words per sentence	20.1	22.0	23.8
Flesch Reading Ease	46.0	45.9	47.4
Flesch-Kincaid Grade Level	11.6	11.7	12.0

The table shows that P(PST) is the longest among the three passages in terms of word counts. P(LMS) has the largest number of words per sentence, which means that its sentences are most likely to be more complex in their structure. The two indicators measure the difficulty of the texts in terms of number of words; however, the relationship between the number of words and text difficulty is not straightforward (Alderson, 2000), and a further analysis is needed. The following two indices might shed some light on the problem.

The Flesch Reading Ease scores show how easy it is to understand a passage. The formula is based on a 100-point scale, and the lower the score, the more difficult the passage is. The formula is $RE = 206.835 - (0.846 \times NSYLL) - (1.015 \times W / S)$, where NSYLL is the average number of syllables per each 100 words, and W / S is the average number of words

per sentence (Alderson, 2000, p. 71). According to the Microsoft Word explanation, most standard documents are rated 60-70 (Microsoft Word 2000). The formula, however, is intended to measure the difficulty of grade-school texts, and as pointed out by Clapham (1996), “is not finely tuned for university level texts” (p. 146). All the three passages of the Reading Test fall within the same interval of 45-48, which means that even though the passages might be difficult for American high-school students, they are most probably quite easy for our subjects, NNS graduate students (Singer & Donlan, 1980).

The Flesch-Kincaid Grade Level score is also based on a US grade-school level system and is a transformation of the Flesch Reading Ease Score. For most standard school documents, the average is 7-8 (Microsoft Word 2000). The difficulty score of our passages ranges from 11.6 to 12.0, which means that the passages are on the level of American freshman university students and are probably either slightly difficult or easy for international graduate students.

An interesting inconsistency is obvious in the numbers presented by the indices. While according to the Flesch Reading Ease scores, P(PST) appears to be the most difficult among the three passages, according to the Flesch-Kincaid Grade Level scores, P(LMS) is the most difficult. No reasonable explanation can be found for the inconsistency at this moment, and since the significance of the inconsistency is not clear either, the general conclusion is that the two passages are more difficult than the third one, P(BSSS), which was rated as the easiest one by both of the indices. A further analysis of the texts might offer more insights into the difficulty of the passages. Below is an attempt to analyze the verb phrases of the three passages and see whether the frequency of specific verbal structures could explain the difficulty of the passages.

As can be seen from Table 3.2, P(LMS) has the largest proportion of perfect verbs, perfect passive verbs, and perfect modals. According to corpus-based studies, perfect aspect is found most frequently in academic writing (Biber et al., 1999, p. 483); academic writing, in its turn, is perceived to be difficult to read. Thus, it appears that P(LMS) is most academic and possibly the most difficult passage out of the three. Table 3.2 also shows that P(BSSS) has the lowest proportion of passive verbs, and because passives are another feature of academic writing (Biber et al., 1999, p. 476), P(BSSS) might be considered the most

Table 3.2: Verbal Structures

	P(BSSS)		P(PST)		P(LMS)	
Verbs	43		56		45	
Perfect verbs	1	(2.33%)	8	(14.29%)	11	(24.44%)
Modal verbs	4	(9.30%)	5	(8.93%)	3	(6.67%)
Passive verbs	7	(16.28%)	13	(23.21%)	10	(22.22%)
Perfect + passive	0	(0%)	3	(5.37%)	6	(13.33%)
Modal + perfect	0	(0%)	0	(0%)	2	(4.44%)
Modal + passive	2	(3.57%)	1	(2.33%)	0	(0%)

“unacademic” in nature, and therefore the easiest as well. This is supported by another finding from the same table, according to which P(BSSS) has the biggest proportion of modals. Modals convey “stance-type meaning,” comments, and attitudes characteristic of conversations and fiction (Biber et al., 1999, p. 487), genres that are comparatively easy to read.

Overall, it emerges that P(LMS) is the most difficult passage, while P(BSSS) is the easiest one. The analysis of passage difficulty will be helpful in my further interpretation of the relationship between discipline-specific BGK and reading test performance. More directly, it will be used in the analysis of RQ3, which focuses on students’ perception of passage difficulty. Passage content specificity is the next aspect of the analysis here.

P(BSSS) is the most general out of the three in terms of content. It is about culture shock, stages of its development, and possible ways of dealing with it. The words used are quite general, and the way of presentation resembles a conversation. All international students coming to ISU are required to attend an introductory session on adjustments in the US; therefore, most of the test takers will be familiar with the topic of this passage.

P(PST) is more content-specific. It deals with the issue of the global greenhouse and involves vocabulary related to chemistry. In addition to chemical formulas such as CFC, CO₂, it has terms such as “methane, carbon dioxide, glaciers, etc. Even though the content is quite technical, some students might be familiar with the topic of the passage due to their previous general education. In addition, the style of writing is such that it portrays the problem in an easy way.

P(LMS) is also more content-specific than P(BSSS) and deals with the topic of the ancient wheat. Though the number of technical terms here is low (DNA is the only example),

the passage involves several analogies (such as “a molecular scrapbook”), and the style of writing resembles scientific writing.

Overall, it appears that P(BSSS) is the easiest in terms of structure and is the least specific in terms of content. P(LMS), on the other hand, is the most difficult in terms of structure and word choice, and P(PST) is the most discipline-specific.

The Test Items

Table 3.3 presents the classification of each item of the Reading Subtest based on the degree to which an item relates to the passage and requires the knowledge of the passage to be answered. The classification (a part of the Raters' Questionnaire) was completed by two independent raters, whose background will be discussed later. In cases where the raters did not agree, my opinion determined the final answer.

This classification will help me to answer the question of why some items might appear biased towards particular groups of students (RQ2). At this moment, it is important to note that 11 out of 30 items are classified as relating to only one part of the passage and

Table 3.3: Test Items as Coded by the Raters

Number in the test booklet	P(PST)	Number in the test booklet	P(LMS)	Item code	Code description
1	3	15	6	0	No relationship to passage; item can be answered without reference to passage
2	1	16	6		
3	1	17	3	1	Item relates to one specific part of the passage and requires only localized understanding of that part
4	0	18	1		
5	3	19	1		
6	0	20	1	2	Item relates to more than one part of the passage
7	6	21	1		
	P(BSSS)	22	6	3	Item relates to the entire passage
8	3	23	1	4	Item relates to one part of the passage, requires the test taker to relate information in that part to real world
9	2	24	5		
10	0	25	0		
11	3	26	1	5	Item relates to more than one part of the passage and requires the test taker to relate information in that part to real world
12	1	27	1		
13	0	28	0		
14	0	29	1	6	Item relates to the entire passage, requires the test taker to relate it to real world
		30	6		

requiring only the localized understanding of that part (item code 1), which implies that one can answer these items without a careful reading of the entire passage. 7 out of 30 items, on the other hand, are classified as having no particular relationship to the passage (item code 0), which shows that the items are not specific to the passage and do not necessarily check reading comprehension. Rather, they might be testing other aspects of test taker knowledge.

The Students' Questionnaire

The Students' Questionnaire was adapted from the questionnaire developed by Clapham (1996) and served the purpose of evaluating students' background and their perceptions of the Reading Test difficulty. The questionnaire consisted of 21 items. Items 1-5 asked students for their personal background, and the information collected was used to assess the sample in terms of gender, nationality, and first language. Item 6 targeted students' previous school education and provided some knowledge of which fields the students had studied previously. Items 8-10 targeted students' future fields of study and helped us to separate graduate students from undergraduates. Item 11 showed the reading interests of the students, and the results were used in the assessment of the relationship between levels of familiarity with a certain field and reading scores. Items 12-21 asked for the students' evaluation of the passages and helped to analyze the relationship between students' perceptions of difficulty and familiarity with the topic with their reading scores. The results of the questionnaire were used to answer RQs 3-5. The full version of the Students' Questionnaire can be found in Appendix D.

The Raters' Questionnaire

The Raters' Questionnaire was again adapted from Clapham's (1996) work, which in turn had utilized Bachman's Test Method Rating Instrument (1991). The questionnaire was used for the purposes of evaluating the Reading Test in terms of specific vocabulary, difficulty, and item type. Part I asked the raters to classify the passages as belonging to one of the three categories: BSSS, LMS, and PST. The classification was then directly used in this study. Part II was intended to provide the analysis of the passages at the level of the

rubric, vocabulary, degree of contextualization, relationship of item to passage, topic, grammar and cohesion. It was intended to provide some additional information on the passage difficulty; however, some of the results were found to be conflicting and only the item classification was used in the further analysis (RQ2). (For the full version of the Raters' Questionnaire, refer to Appendix E.)

Participants

Both students and raters were involved in the study.

The Students

The major participants of the study were 252 non-native speakers of English starting their graduate career at various colleges of Iowa State University (ISU). Since a minimum of 500 on the Paper Based TOEFL or 173 on the Computer Based TOEFL is required for admission to ISU's Graduate College, it is possible to conclude that the participants were of different English language proficiency levels, but all above a minimum threshold. They represented all the eight colleges of ISU, and the following is the table that illustrates the distribution of students between the colleges.

Table 3.4 shows that the distribution of students between the colleges is uneven, with Liberal Arts and Sciences and Engineering students prevailing over the others. This can be explained by the nature of ISU. A large-scale US Midwestern university, it is famous for its engineering programs; therefore, a big portion of the international population is attracted primarily to this college. The LAS college is comprised of numerous departments, including English, math, sociology, chemistry, history, etc.; therefore, the number of students in it is always high.

Table 3.4: Distribution of Students between the Colleges

College	N	College	N
Agriculture	23	Veterinary Medicine	6
Engineering	75	Education	7
Family Consumer Sciences	9	Design	12
Liberal Arts and Sciences	107	Business	13
Total	252		

For the sake of simplification of the analysis procedures, single colleges were grouped into larger units. Thus, the colleges of Liberal Arts and Sciences, Education, Design, and Business were grouped into the category of “Business Studies and Social Studies” (BSS), the colleges of Agriculture, Family Consumer Sciences, and Veterinary Medicine into “Life and Medical Sciences” (LMS), and the college of Engineering was put into the category of “Physical Science and Technology” (PST)¹. Table 3.5 displays the distribution of students between the broader areas of study, calculated for two interdependent samples (Total-students of both Fall and Winter testings, and Winter-students of the Winter testing).

Table 3. 5: Distribution of Students between the Broader Areas of Study

	BSSS	PST	LMS
Total	139	75	38
Winter	16	13	13

Due to the unequal distribution of students between the colleges and for further simplification, it would have been reasonable to combine LMS students with PST students (similar to Hale, 1988); however, the mean scores of the two subgroups on the Reading Test were found to be significantly different (one-tail test with $\alpha=.05$ resulted in $z=1.94$, $p<.0262$), and combining them would have altered the results of the analysis.

Since the data were collected in two testing sessions (Fall 2000 and Winter 2001) and the Winter session did not involve the Students’ Questionnaire, the amount of information available from each testing is not equal. Thus, no information is available about the origins and the native languages of the Fall subjects. The questionnaire conducted during the Winter administration of the EPT showed that the genders were distributed nearly equally (22 males and 20 females in our sample). Table 3.6 provides the distribution of students between the languages.

¹ These are the same groupings as the ones used by Clapham (1996) in her study .

Table 3.6: Students' First Languages

Language	N	Language	N
Arabic	4	Korean	2
Chinese	17	Spanish	6
Languages of India	4	Russian	4
		Others	5

The Raters

Judges were needed for the purposes of analyzing the reading subtests and the reading items of the English Placement Test; therefore, two second-year and one third-year MA students of TESL were additionally involved. They had agreed to participate on a voluntary basis out of the group of students who had attended a course in Language Assessment in Spring 2000. Out of the three raters, one (the third-year MA student) did not return the questionnaire; therefore, the answers of the other two raters were used in this study. In cases where the two raters did not agree, the opinion of the author of this study determined the final decision. Look at the following section for the description of the complications encountered during the administration and analysis of the Raters' Questionnaire.

Procedures

Fall 2000 and Winter 2001 Administrations

The major data for this study comprised the results from the two separate testing sessions conducted by the faculty and graduate students of the English Department of ISU in August of 2000 and January 2001. The data collected from the first testing will be therein coded as the Fall data and the data collected from the second testing—the Winter data. Table 3.7 that follows displays the administration order of each subtest of the EPT during each testing procedure.

Group 1 and Group 2 were administered the test on the same date but were placed in different auditoriums. The supervisors for each auditorium were given the same instructions and were trained in the same way with an attempt to standardize the testing procedures. Each auditorium had 2 proctors to help the supervisor prevent academic dishonesty and answer students' questions about filling out different forms.

Table 3.7: The Testing Sessions and the Order of Subtest Administration

	Fall testing	Min.	Winter testing	Min.
Group 1	Personal Information	10	Personal Information	10
	Writing Test	30	Writing Test	30
	Reading Test	40	Reading Test	40
	Break	10	<i>Students' Questionnaire</i>	15
	Listening Test	50	Break	10
			Listening Test	50
Group 2	Personal Information	10	Personal Information	10
	Writing Test	30	Writing Test	30
	Reading Test	40	Listening Test	50
	Break	10	Break	10
	Listening Test	50	Reading Test	40
			<i>Students' Questionnaire</i>	15
Makeup Group	Personal Information	10	Personal Information	10
	Writing Test	30	Writing Test	30
	Reading Test	40	Listening Test	50
	Break	10	Break	10
	Listening Test	50	Reading Test	40
			<i>Students' Questionnaire</i>	15

The makeup group comprised students who arrived late for the major test administration dates, and the EPT was administered to them one week after the main administration.

Administration of the Students' Questionnaire

The Winter administration differed from the Fall administration in terms of the number of parts and the changing order of the subtests. The questionnaire was administered only during the Winter data collection session because the idea that it could help to clarify certain aspects of students' reading performance occurred only after the preliminary analysis of the Fall data. The analysis showed that PST students performed comparatively well across the passages; therefore, other variables than just the students' area of study and their linguistic proficiency were possibly influencing their test performance.

By January, the Student' Questionnaire had been adapted from Claphams' (1996) work and was used as an additional data collection instrument. It had to be distributed right

after the Reading Test to ensure that the students would not forget the passages and the items. The author of the study introduced the questionnaire by reading the instructions and asking the students to sign their agreement to participate (for the Consent Form that carried both the signatures and the instructions, refer to Appendix C, and for the Human Subject Committee approval, refer to Appendix B).

The order of presenting the subtests differed from Group 1 to Group 2 during the Winter administration for the purpose of providing me with the chance to be present in each auditorium while the questionnaires were answered, explain the purposes of the study, and answer possible questions. The two groups did not have a chance to meet with each other during the break because of the timing, and, therefore, no leak of information about the last subtests was available.

The makeup test for the Winter administration followed the second order of administration, chosen at random. After the tests were administered, the scores were calculated by the University Testing Services and were transmitted to the EPT administrator. She, in turn, provided me with the information necessary for the purposes of this study—scores, college assignments, and ID numbers. The latter were removed as soon as the need for them was satisfied.

Administration of the Raters' Questionnaire

The Raters' Questionnaire was administered during late January. The raters were given a fixed set of instructions and were encouraged to ask questions when necessary. Due to the busy schedules of all the three participants, no specific introductory session was provided. This, in turn, led to several problems. First, one of the raters was not able to finish the questionnaire due to the number of questions and their specificity. Second, after the rater was dropped from the study, the classification of passages turned very controversial. The problem lay with the first passage. The first rater classified it as "PST"; however, the second classified it as "LMS". The final decision on how to classify this passage was reached after a meeting with both of the raters, during which the second rater agreed with the classification of the first rater but said that due to the confusing classification system, she was not sure whether the passage was PST or LMS. It was agreed, that the passage was PST rather than

LMS. (See Chapter 5 for future suggestions on how to eliminate this problem in further research.)

Analysis

As has been mentioned above, several data types were used in answering research questions. The total sample of subjects was comprised of students from both Fall and Winter test administrations and was used in the analysis of the interaction between a discipline of study and reading test performance (RQ1) as well as in the analysis of item difficulty for particular groups of students (RQ2). Most of the Winter students, the ones who have completed the Students' Questionnaire, were used to answer RQs 3-5.

Every research question of the study required a specific statistical analysis. To find the relationship between the students' field of study and their performance on the reading passages (RQ1), descriptive statistics for each group of students were first calculated. The statistics, however, did not provide much information because they did not show whether the groups differed significantly in dealing with the three passages. A repeated measures analysis of variance was therefore used, with student group and reading passage as the two independent variables. Using SAS repeated measures MANOVA, the results were analyzed to see whether there were significant differences between the group scores, the test scores, and the interaction of the two.

The problem with this kind of analysis was the fact that the score distribution for all three groups was not normal (see Appendix F for samples of normal probability plots), and since normality is one of the requirements for ANOVA/MANOVA, the results of the analysis could be misleading. There were at least two major reasons for the lack of normality: the easiness of the items which led to positively skewed results, and the small number of items following each passage. Even though P(LMS) had the largest number of items (16 instead of 7), the scores to this passage were still slightly skewed. To normalize the score distribution, the scores were converted from discrete to categorical variables. Table 3.8 presents the conversion of the scores for each passage.

Use of the Multinomial Logistical Regression Model became possible after the reclassification of the dependent variable of passage score. Multinomial logistic regression is employed to handle the analysis of categorical dependent variables with more than two

Table 3.8: Conversion Table of Passage Scores

P(BSSS) scores		P(PST) scores		P(LMS) scores	
Raw	Converted	Raw	Converted	Raw	Converted
1-5	1 (low)	1-5	1 (low)	1-12	1 (low)
6	2 (average)	6	2 (average)	13-14	2 (average)
7	3 (high)	7	3 (high)	15-16	3 (high)

classes. It applies maximum likelihood estimation after transforming the dependent variable into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring or not. The advantages of this models is that it does not assume linearity of relationship between independent variables and the dependent variable, does not require normally distributed variables, and in general has less stringent requirements than parametric tests (Garson, 2001).

For the purposes of the analysis, the CATMOD procedure of SAS was used, which calculated the log odds of each category of the dependent variable relative to the last category of the dependent variable. The two series of analysis were carried out. First, I analyzed the performance of each group on specific passages and, thus, had three sets of output. Second, I analyzed the performance of the three groups on the same passages and, thus, had another set of three outputs.

To see whether the groups of students had difficulty with particular items of the Reading Test (RQ2), an item analysis was carried out. Separate item analyses were conducted by the Testing Services of Iowa State University for each group of students, and item facility indices were compared across the groups. Additionally, possible reasons were hypothesized to account for the difficulty of particular items to specific groups. For this purpose, the raters' classification of items from the Raters' Questionnaire was used. (See Table 3.3 for item classification.)

To analyze the group differences between students' perception of passage difficulty (RQ3), question 12 of the Students' Questionnaire was used. It reads:

Q12: Questions to which passage do you think you answered best?

The numbers were calculated for each passage and were then compared across the groups. The Chi-square method of statistical significance was used for this purpose, with number of people considering the passage easy as the dependent variable, and the group or

area of study being the independent variable. Chi-square is a non-parametric test of statistical significance that analyzes whether two or more groups are different enough in some characteristic or aspect of their behavior. CHISQ option of the FREQ procedure of SAS was used to calculate chi-squares in this study.

To find the relationship between students' familiarity with the passage topic and their performance on that particular passage (RQ4), the answers to Questions 14, 17, and 19 of the Students' Questionnaire were primarily used. The general format for the question was the following:

Q14: Were you familiar with the problem of greenhouse effect before you read the passage?

To change the category of the independent variable from nominal to ordinal, the answer "yes" was coded as 1, while the answer "no" was coded as 0. The CATMOD procedure of SAS was used to determine whether the groups differed significantly in being familiar with passage topics. It was first analyzed whether the groups differed in being familiar with the topics of each passage. Then, it was determined whether familiarity with a passage topic correlated to the passage score.

To see whether students' previous education influenced their scores on the reading passages (RQ5), the answers to Question 6 of the Students' Questionnaire were first analyzed. Those students who pointed that they had subjects like literature, history, etc. in the last two years of their education, were given the score of 1, while those who did not point any subject from the area of BSSS in their educational background, were given the score of 0. The same was done for every other area of study. The scores were then compared within groups to see whether students' current area of study was a significant indicator of their previous education. Afterwards, the scores on the reading passage from one area were compared to students' answers to the question on previous education. It was analyzed whether having had some education in an area determined the scores on the passage related to that area. Both kinds of analysis were carried out on SAS by using the CATMOD procedure.

Summary

In this chapter, we have discussed the methodology used in the study. This included the description of the test takers and the raters, the detailed analysis of data collection tools, and the preview of statistical methods used for the major analyses. Table 3.9 summarizes all the major points discussed above, mainly the subjects, tools, and statistical analyses used.

Having previewed the statistical methods used in the study, it is time to go into the actual analyses of data and the discussion of the results. The following chapter will be divided into several sections, each one of which will summarize the results of the analyses used for each research question.

Table 3.9: Summary Table

RQ	Subjects	Tool	Statistical analysis	
1	Do students tend to perform better on the passages related to their field of study?	Total sample	Reading Subtest of the English Placement Test	Multinomial Regression analysis
2	Are there items that seem to be problematic for specific areas of study?	Total sample	Reading Subtest	Item analysis provided by the Testing Services
3	Do groups differ in passage difficulty perceptions?	Winter sample	Students' Questionnaire	Chi Square analysis
4	Does a possible familiarity with the topic of a passage help students perform better on the passage?	Winter sample	Students' Questionnaire + Reading Subtest	Multinomial Regression analysis
5	Does students' educational history influence their reading test performance?	Winter sample	Students' Questionnaire + Reading Subtest	Multinomial Regression analysis

CHAPTER 4. RESULTS

In this chapter, I will discuss the results of the multiple statistical analyses I have performed to investigate each research question stated in Chapter 2. The chapter, therefore, will be divided into the following parts: 1) Descriptive Statistics—measures of the students' performance on the passages of the Reading Subtest; 2) Analysis of Variance across Groups and Passages—multinomial logistic regression results of the relationship between passage scores and student groups; 3) Item Bias—facility indices for each item of the passages compared across groups and passages; 4) Passage Ease/Difficulty Perceptions—Chi-square results of the relationship between students' area of study and their choice of the easiest passage; 5) Topic Familiarity—multinomial logistic regression results to show whether groups differ in passage familiarity and whether familiarity helps them to perform on the passages; 6) Previous Education—multinomial logistic regression results to analyze the relationship between students' previous education and their performance on the passages; and 7) Summary—chapter highlights in one table.

Descriptive Statistics

Initial analysis of data involved calculating descriptive statistics of the passage scores. Table 4.1 presents the distribution of scores between passages based on the total sample of students. Even though no conclusive statements can be made about the difficulty of the passages based on the descriptive statistics presented here, certain conclusions still arise.

Table 4.1: Distribution of Scores between Passages

	P(BSSS)	P(PST)	P(LMS)
Number of students	252	252	252
Number of items	7	7	16
Mean raw score	6.11	5.94	13.01
Raw score SD	.93	1.05	2.26
Mean as %	87.24	84.81	81.30
SD as %	13.26	15.01	14.11
Median as %	85.71	85.71	81.25
Maximum/Minimum as %	100/42.86	100/14.29	100/31.25

First of all, the table shows that the number of items is highest in P(LMS); therefore, the raw mean score and the raw score SD comparisons are not appropriate between P(BSSS) and P(LMS) and between P(LMS) and P(PST). Some comparison could be made between P(BSSS) and P(PST), however, and the results show that on average students performed better on P(BSSS) (mean of 6.11 and SD of .93 compared to 5.94 and 1.05, accordingly).

The percentages expressing the mean scores and the SDs show that items to P(BSSS) could be easier than items to the other passages, and items to P(LMS) could be the most difficult. The median says that the most frequent percentage for P(BSSS) and P(PST) is 85.71 while for P(LMS) it is only 81.25. The maximum score shows that the highest score for each passage was 100%, while the minimum score was lowest for P(PST). Thus, on the basis of the above table, it is plausible to predict that P(BSSS) is the easiest among the three, while P(PST) and/or P(LMS) are more difficult.

Table 4.2 presents another set of descriptive statistics—means across groups and passages. This is done to show whether groups perform better on passages related to their specific fields of study. According to our hypothesis, groups will perform better on the passages associated with their fields of study. This, however, is not supported by the data presented in the table above, according to which group BSSS performs better than the others on P(PST) and P(LMS), while group PST outperforms everybody on P(BSSS). These results, however, are not conclusive because no significance test has yet been performed at this stage. Therefore, the next step is to choose the appropriate statistical method to evaluate the significance of differences between the group scores on the three passages.

Table 4.2: Table of Means (%) across Groups and Passages

N	Group	P(BSSS)	P(PST)	P(LMS)
139	BSSS	87.87	86.64	82.06
75	PST	89.14	84.57	81.25
38	LMS	81.20	78.51	78.62
252		87.24	84.81	81.30

Analyses of Variance across Groups and Passages

As has been mentioned in Chapter 3, the scores on P (PST) and P(BSSS) were not distributed normally, and the small number of items and their comparative ease are probably one of the major reasons for that. Even though scores on P(LMS) displayed a better distribution pattern, the results were still skewed.

Multiple analysis of variance was employed in the beginning, and the results of it are reported in Appendix G. However, the validity of the results of MANOVA highly depends on the normality of score distribution, and since our distributions do not satisfy the normality requirement, it was inappropriate to rely on the results of MANOVA alone. Therefore, another model, multinomial logistical regression, was used in the further analysis. Two series of analyses using CATMOD were carried out: one to see whether the scores within specific groups depend on the passage, and second to see whether the scores on specific passages depend on which college students are in. Below we will present the parts of the outputs that will help us directly to come to our conclusions.

Variance Within Groups

Group BSSS

Table 4.3 below shows the results of the analysis of scores within group BSSS. The response variable here is *score*, with three response levels (1-low, 2-average, and 3-high); the number of populations is 3 referring to the different categories of passages in the predictor variable (passages BSSS, PST, LMS).

The first part of Table 4.3, displaying the maximum likelihood analysis of variance, shows that there is statistical significance for the intercept ($p < .05$), but not for the type of passage ($p > .05$). The intercept shows that there is a significant difference across the passages in the estimated average log of the odds of a passage getting low scores. The lower part of the table provides more information, and Parameters 3 and 4 are of specific importance. Parameter 3 shows that items based on P(BSSS) are more likely than items based on other passages to be scored high, which shows that group BSSS performs significantly better on this passage, the passage related to the discipline of study. Parameter 4 shows that items to P(PST) are about as likely as all other items to be scored high.

Table 4.3: Logistic Regression Model for Score as a Function of Passage within Group BSSS

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	10.22	0.0060 <.05	
	Passage	2	5.00	0.0821 NS	
	Likelihood Ratio	2	0.44	0.8027	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.1022	0.3458	10.16	0.0014 <.05
	2	-0.3242	0.2946	1.21	0.2711 NS
Passage	3	0.3480	0.1578	4.86	0.0275 <.05
	4	0.1852	0.1391	1.77	0.1830 NS

The average percentage of correctly answered items for group BSSS on P(BSSS) is 87.87, and the percentage for P(PST) is 86.64. Finally, the percentage for P(LMS) is 82.06. Thus, items based on both P(BSSS) and P(PST) are much more likely than items based on P(LMS) to be answered well, and it is not surprising that the CATMOD results show that P(LMS) is significantly less likely to be answered best by group BSSS.

Group PST

Table 4.4 presents part of the results for the same kind of analysis using CATMOD but this time for group PST. The response variable here is again *score*, with the same three response levels; and the number of populations is three referring to the number of passages each member of the group had to read.

According to the maximum likelihood analysis of variance, both the type of passage and the intercept are highly significant ($p < .05$). The intercept shows that for group (PST), there is a significant difference across the three passages in the estimated average log of the odds of a passage getting low scoring. The results of the analysis of maximum likelihood estimates show that items based on passage 1 (P(BSSS)) are more likely than items based on other passages to be answered well (Parameter 3). According to our hypothesis, the group will perform better on the passage related to its field as well; however, Parameter 4 shows that even though group PST answers items based on P(PST) well enough, the result is not significant. Thus, our hypothesis was not fully supported for this group.

Table 4.4: Logistic Regression Model for Score as a Function of Passage within Group PST

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	12.93	0.0016	<.05
	Passage	2	11.22	0.0037	<.05
	Likelihood Ratio	2	1.39	0.5001	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.7352	0.4827	12.92	0.0003 <.05
	2	-0.6223	0.4047	2.36	0.1241 NS
Passage	3	0.7314	0.2187	11.19	0.0008 <.05
	4	0.3471	0.1966	3.12	0.0774 NS

Group LMS

Table 4.5 displays the last set of results from this series of CATMOD analyses and shows how group LMS performs on the three passages. The response variable, the response levels, and the population kinds are the same as in the two preceding analyses. Our hypothesis is that the group scores highest on the passage related to it, P(LMS). Both the maximum likelihood analysis of variance and the analysis of maximum likelihood estimates show that the performance of this group, LMS, does not depend on the kind of passage presented to the students. Moreover, the intercept shows that there is no significant

Table 4.5: Logistic Regression Model for Score as a Function of Passage within Group LMS

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	2.40	0.3017	NS
	Passage	2	0.93	0.6267	NS
	Likelihood Ratio	2	0.66	0.7173	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	0.7478	0.6761	1.22	0.2687 NS
	2	1.0451	0.6769	2.38	0.1226 NS
Passage	3	-0.0565	0.3043	0.03	0.8526 NS
	4	-0.2642	0.3115	0.72	0.3964 NS

difference across the three passages in the estimated average log of the odds of a passage getting low scores. Items to P(BSSS) are as likely as items to other passages to be scored low (Parameter 3), and items to P(PST) are as likely as items to other passages to be scored low as well (Parameter 4). Thus, the hypothesis has not been confirmed for this group; furthermore, it is clear that the group performs equally badly on all three passages. This is not surprising because the results of the preliminary analysis show that the mean percentages of the group for all the three passages are comparatively equally low (81.20, 78.51, and 78.62 for P(BSSS), P(PST), and P(LMS) comparatively).

The overall conclusion based on the three sets of this series of CATMOD analyses is that group BSSS performs better on P(BSSS), group PST does not perform significantly better on P(PST), and group LMS performs at the same level on all the three passages. This conclusion might seem surprising given the fact that subject area familiarity was hypothesized to have a positive effect on passage scores. Therefore, further analysis of why the hypothesis does not hold true across the groups should be carried out. The next step will be to analyze whether the same passage scores are significantly different across the groups.

Performance across Passages

Passage BSSS

Table 4.6 below demonstrates the results of the CATMOD procedure where the scores on P(BSSS) are analyzed across the groups. The response variable here is *score*, with the same three levels of response (1-low, 2-average, and 3-high); the number of populations is three but this time the populations correspond to the groups presented in the study (groups BSSS, PST, and LMS). The hypothesis is that group BSSS will perform better than the other two groups on the items based on this passage.

The intercept, calculated through the maximum likelihood analysis of variance, shows that there is a significant difference across the groups of students in the estimated average log of the odds of the scores being low for P(BSSS). Parameter 3 leads to the conclusion that group BSSS is as likely as other groups to get high scores on this passage, while Parameter 4 shows that group PST is as likely as other groups to perform well on this passage as well. Thus, group BSSS does not perform significantly better than the others here.

Table 4.6: Logistic Regression Model for P(BSSS) Score as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	10.13	0.0063 <.05	
	Group	2	3.77	0.1520 NS	
	Likelihood Ratio	2	8.22	0.0164	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.2885	0.4050	10.12	0.0015 <.05
	2	-0.4682	0.3435	1.86	0.1729 NS
Group	3	0.4275	0.2251	3.61	0.0576 NS
	4	0.2475	0.1999	1.53	0.2157 NS

According to the descriptive statistics (Table 4.2), students of BSSS scored an average of 87.87 and students of PST--89.14 on P(BSSS). Finally, students of LMS scored an average percentage of 81.20 on the same passage. It is not surprising then that group LMS is significantly less likely to get high scores on the items to P(BSSS).

Passage PST

The next table will display the results of the same kind of analysis for P(PST). Here the response variable, the response levels, and the kinds of populations are the same as in the previous analysis. The hypothesis states that group PST will perform significantly better on P(PST) than the other groups due to the effect of discipline-specific BGK.

According to the analysis of variance results, both the intercept and the kind of college are statistically significant ($p < .05$). The intercept tells us that there is a significant difference across the three groups in the estimated average log of the odds of the items to P(PST) being scored low. Parameter 3 of the model shows that group BSSS is more likely than all groups to get high scores on this passage. Parameter 4 shows that group PST is as likely as others to get the high score on the same passage. This confirms our previous conclusion that group PST does not perform better on the passage from the area closest to the group's discipline of study.

Table 4.7: Logistic Regression Model for P(PST) Score as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	8.97	0.0113	<.05
	Group	2	6.58	0.0372	<.05
	Likelihood Ratio	2	1.21	0.5471	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.1313	0.3977	8.09	0.0044 <.05
	2	-0.1923	0.3559	0.29	0.5889 NS
Group	3	0.5617	0.2229	6.35	0.0117 <.05
	4	0.2049	0.2127	0.93	0.3353 NS

Passage LMS

Table 4.8 presented below is the last from this series devoted to the analysis of passages and represents the results of CATMOD for the last passage (P(LMS)). The response variable, the response levels, and the populations are left without changes. The hypothesis to be analyzed is whether group LMS is the one that performs best on this passage. According to the table, neither the intercept nor the group are statistically significant ($p > .05$). The intercept shows that there is no significant difference between the three groups in the

Table 4.8: Logistic Regression Model for P(LMS) Score as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	2.03	0.3616	NS
	Group	2	2.55	0.2793	NS
	Likelihood Ratio	2	0.17	0.9191	
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-0.2862	0.3857	0.55	0.4581 NS
	2	0.2258	0.3758	0.36	0.5480 NS
Group	3	0.3029	0.2187	1.92	0.1660 NS
	4	0.0355	0.2212	0.03	0.8724 NS

estimated average log of the odds of the scores being low on P(LMS). Parameter 3 shows that group BSSS is as likely as other groups to perform well on this passage, while Parameter 4 explains that group PST is also as likely as others to get high scores on this passage.

The results do not come across as surprising because, according to Table 4.2, none of the groups performed well on the passage and the distribution of the average percentage of correct answers for groups BSSS, PST, and LMS was 82.06, 81.25, and 78.62. There are several reasons that could account for why the groups do not differ in their performance on this passage. First, the passage could be more difficult than the others, and, therefore, is more problematic. Second, the groups could have the same level of background knowledge on the topic of P(LMS). However, the last reason does not seem plausible because of the following. Group (LMS) has been shown to perform worse than the others on the two previous passages, and the same could have been expected for P(LMS). The fact that the group does not perform significantly worse on it than the other two groups is tentative evidence that shows the group's use of discipline-specific background knowledge.

Conclusions

The overall conclusions we came to based on the analyses described above are the following:

1. Group BSSS performs better on P(BSSS) than the other passages;
2. Group PST does not perform significantly better on P(PST);
3. Group LMS performs equally badly on all the three passages;
4. Groups BSSS and PST perform on the same level on P(BSSS), while group BSSS performs significantly better on P(PST) than the other groups.

Thus, it is obvious, that groups BSSS and PST do not perform best on passages related to their disciplines of study. However, group LMS, though not performing best on P(LMS) either, might still be benefiting from using discipline-specific BGK. Our next step is to analyze what could contribute to the results found in this part of analysis, and the first thing we will look at is item bias.

Item Bias

The results of two kinds of item analysis, the raters' classification of the passage items and the facility indices of the items provided by the Testing Services of ISU, will be described in this section. Table 4.9 presents facility indices of each item for every group of students. The comparison is carried out between the groups on each item of each passage, and our interest lies with the item difficulty indices shown in bold. From the three group indices for each item, the lowest one is first chosen and then compared to the second lowest index. If the lowest index differed by 4 or more points from the second lowest index, it was shown in bold in the table because it was believed to be significantly low from the other two indices.

The analysis of the table will focus on several aspects. First, we will look at which student group gets the highest number of difficulty indices shown in bold. Second, we will analyze items of which passage are of lowest facility level for each group of students. Finally, we will try to examine why some items are of particular difficulty to the specific groups of student. First, let's examine the distribution of the lowest indices between the groups and passages.

Table 4.9: Item Facility Indices across Groups for Each Passage

Number in the test booklet	P(PST)			P(LMS)			Number in the test booklet
	LMS	BSSS	PST	LMS	BSSS	PST	
1	88	87	96	79	82	89	15
2	85	92	92	82	91	86	16
3	94	93	96	91	81	77	17
4	79	91	94	62	64	61	18
5	41	61	49	79	76	79	19
6	74	88	83	79	88	90	20
7	91	97	92	76	76	79	21
		P(BSSS)		82	83	87	22
	LMS	BSSS	PST	68	76	89	23
8	68	82	89	74	86	83	24
9	94	97	96	97	98	99	25
10	71	66	79	73	76	69	26
11	65	89	77	82	77	66	27
12	88	89	94	76	76	86	28
13	97	99	96	100	96	94	29
14	88	94	96	85	85	70	30

Table 4.10 reveals that group BSSS has the least number of the lowest indices (1 out of 16); while group (LMS) has the greatest number of the lowest indices (11 out of 16). Group PST is in the middle and has 4 out of 16 lowest indices. Even though group BSSS was shown to perform statistically significantly better than the other groups on P(PST) only (Table 4.9), the trend presented in Table 4.10 is in agreement with our descriptive statistics according to which group BSSS performs better than the other two groups on P(PST) and P(LMS) and group PST performs better on P(BSSS) (Table 4.2, see Table 3.3 for the full classification of the items and for the code description.).

Table 4.10: Lowest Indices between the Groups (with Item Codes in Parentheses)

	P(BSSS)	P(PST)	P(LMS)	Total
BSSS	1 (0)	0	0	1
PST	0	0	4 (3, 1, 1, 6)	4
LMS	3 (3, 3, 0)	4 (1, 0, 3, 0)	4 (6, 1, 1, 5)	11
Total	4	4	8	16

The second step in this analysis is to analyze what kind of questions each group gets the lowest indices on. Since the BSSS group does not have many indices in bald, we will skip this group; instead we will focus on groups PST and LMS. By looking at table 4.10, it becomes obvious, that the kinds of items group PST finds most difficult are the ones coded as 1, 3, and 6, while the kinds of items group LMS finds difficult are coded as 0, 1, 3, 5, and 6. Thus, while the former has problems with three kinds of the items, the latter has problems with five kinds of items, almost every kind except for 2 and 4 (no item coded as 4 is available in the passages, and only one item coded as 2 is present).

Of the most significance are the items coded as 0. These are the items that can be answered without reading the passage because they have little, if any, relationship to the passage. These are usually vocabulary items and are worded similar to the example below.

Item 13 (P(BSSS)): The word “successive” in line 24 means

- a. getting weaker and weaker.
- b. happening at the same time.
- c. happening one after another.
- d. getting stronger and stronger.

Even though the item refers the test takers to the specific line of P(BSSS), it can be answered without referring to the passage. The vocabulary item “successive” is widely spread enough to be defined without reading the passage or part of the passage. The analysis of the item facility indices shows that group PST does not find difficulty answering questions like the one above, while group LMS does. This could imply that the first group has a stronger linguistic proficiency and a larger vocabulary, while the second group experiences lacks in vocabulary. The fact that PST students have most problems with items related to the passage shows that they are not reading carefully, instead they rely on their general English knowledge and surface reading. LMS students, however, have problems with many kinds of items which shows that they are simply of a lower linguistic proficiency than everybody else.

P(LMS) has the highest number of the lowest facility indices, compared to the other two passages. Even though it is difficult to compare the numbers across the passages because the number of items across the passages differs significantly, this is another indicator of the fact that P(LMS) could be the most difficult for the students. Moreover, no item coded as 0 was shown to have the lowest facility index for this passage, which means that students have problems with items related to the passage itself rather than with items non-related to the passage. This, in turn, means that they have to perform a more careful reading of the passage to be able to answer the questions.

Thus, the conclusion is that items appear to be more difficult for students of LMS rather than for student of the other two groups. It also appears that passages differ in the kinds of items, and items to P(LMS), for example, require more in-depth reading of the passage than the items of the other two passages. Finally, it also appears that the students of PST do not experience problems with items unrelated to the passages, but they do experience problems with items related to the passages. This is a possible indicator of them relying on their linguistic or other background to answer questions. Group LMS students, however, along with experiencing problems with passage-related items, experience problems with passage unrelated items. This could be a possible indicator of these students having linguistic proficiency lower that the other students.

The overall conclusion, therefore, is that items do have different difficulty values for each group and across the passages. This could be explained by the linguistic proficiency of

the students, passage specificity level, or passage difficulty. The next thing to analyze now is to see whether students perceive the ease of the passages similarly across the groups.

Passage Ease/Difficulty Perceptions

The cells in Table 4.11 represent the number of people (with group percentages in the parentheses) who considered a specific passage as the one they answered best. This shows the comparative ease of the passages for each group of students. According to this table, group BSSS perceives P(BSSS) and P(PST) as being equally easy compared to P(LMS), the fact that has been proved through the logistic regression results shown in Table 4.3.

Table 4.11: Ease of the Passages

	P(BSSS)	P(PST)	P(LMS)	Total
BSSS	8 (44.44%)	8 (44.44%)	2 (11.11%)	18
PST	6 (33.33%)	9 (50.00%)	3 (16.67%)	18
LMS	8 (57.14%)	3 (21.43%)	3 (21.43%)	14
Total	22	20	8	50

Group PST chooses P(PST) as the one they answered best, with P(BSSS) coming after it. This is also in accord with the findings described in Table 4.4, where it was shown that group PST does perform better on P(BSSS) and P(PST) rather than on P(LMS). Group LMS, even though they find the BSSS passage as the easiest one, still consider P(LMS) of the same difficulty level as P(PST). LMS students are the only test takers that consider P(LMS) of at least the same ease level as P(PST). For all the other colleges, the passage is of highest difficulty.

These results show a hypothetical relationship between the students' area of study and the passage that they perceive as the easiest one. Chi-squares analysis of the relationship, however, did not show a strong relationship between the area and the passage types. In fact, the chi-square is 3.2937 with 4 degrees of freedom, which amounted to the probability of more than .05. Thus, the relationship is not significant. The phi coefficient of .2567 shows that 6.6% of the dependent variable has been explained by the independent variable, which means that college type explains only 6.6% of the choice of the easy passage. Thus, 93.4% of the variance has to be explained by something else.

The numbers shown above, however, could be due to the small frequency numbers in the cells. Chi-square analysis of 3*3 tables requires that all cells but one have at least 5 observations. In our case, there are 3 cells that have three observations and 1 cell that has one observation only. Overall, even though the significant relationship has not been found between the students' area of study and their choice of the easiest passage, the raw data still suggest that such a difference exists. Moreover, they suggest that students tend to be more confident about passages that come from their field of study, even though they do not perform best on those passages.

Passage Familiarity

So far, I have looked at how students from different groups deal with passages related to their disciplines of study; I have analyzed whether there are specific items that present biggest problem to students of each group and whether there are specific passages that have most of the problematic items. I have also looked at the students' perception of the passage difficulty and tried to see whether studying a particular discipline affects the difficulty perceptions. My next step is to analyze various aspects of discipline-specific background knowledge. Students' familiarity with the topic of the passage could substitute for their background knowledge in the discipline presented in the passage. Being in one area does not exclude the possibility of being familiar with topics of other areas.

To see whether students test performance is influence by them being familiar with the topic of the passage, two sets of analyses were performed. First, I analyzed whether being familiar with the topic of the passage depends on the area of study, and then I looked at how or whether being familiar with the topic of the passage helps the students to perform better on the passage. The results of the analysis revealed no relationship between students' topic familiarity and their test performance. The tables that provide the significance tests and their results are included in Appendix H. The next step is to analyze whether previous education plays a significant role in reading comprehension exams.

Previous Education

Previous education can contribute to students' discipline-specific background knowledge, and therefore is considered to be of importance here. Three sets of analyses were performed during this stage of the study, each set of which consisted of two aspects: the first focused on whether being in a particular college is a significant indicator of one's previous education, and the second examined whether having a particular educational background helps students to perform better on the passages of the Reading Subtest. Overall, it became obvious that previous education and current discipline of study have high correlation. However, having had some education in a particular area does not significantly facilitate students' performance on the passage related to that area. The tables that contain the information related to this analysis could be found in Appendix I.

Summary

The current study has attempted to analyze several aspects of discipline-specific background knowledge and its effect on reading comprehension. First, the effect of students' discipline of study on their reading scores was analyzed. Later, variables such as topic familiarity, passage ease perception, item bias, and previous education were additionally analyzed. Table 4.12 will present a summary of the overall results. It will first restate the research questions and then explain what has been found for each one of them.

Table 4.12: Summary of the Results

RQ	Results
1 Do students tend to perform better on the passages related to their field of study?	The results are quite inconclusive. Group BSSS performs significantly well on P(BSSS); however, so does group PST. Group PST does not perform significantly better than others on P(PST); however, group BSSS does. As for group LMS, it performs significantly badly on all the passages, but when it comes to P(LMS), it does not perform worse than the others.
2 Are there items that seem to be problematic for specific areas of study?	Overall, the items appear to be more difficult for students of LMS. It also appears that P(LMS) is followed by items that are most problematic, not only for group LMS but for group PST as well.
3 Do groups differ in passage difficulty perceptions?	Even though no significant relationship has been found between the groups and their perception of passage difficulty, it is obvious that group LMS considers P(LMS) of similar difficulty as P(PST), while all the other groups consider P(LMS) as the most difficult passage.
4 Does a possible familiarity with the topic of a passage help students perform better on the passage?	It is found true for all the three passages that there is no significant interaction between students' familiarity level with the topic of the passage and their performance on it.
5 Does students' educational history influence their reading test performance?	Groups' current fields of study and their previous education are found to be of high correlation, but no significant influence of previous education on the passage scores is obvious.

CHAPTER 5. DISCUSSION AND CONCLUSION

Based on the results of the current research, this section on the conclusions will be divided into the following specific parts: 1) Summary—the overall results of the study and some possible explanations; 2) Insights for Researchers, Teachers, and Test Developers—some practical and theoretical insights into the effects of discipline-specific BGK on reading test performance; 3) Limitations of the Study—the numerous aspects of the study in need of improvement; and 4) Future Research—possible areas in future research into background knowledge and its influence on reading test performance.

Summary

This study has looked at the effect of the complex construct of discipline-specific BGK on reading test performance. It has attempted to measure some of the factors that contribute to the development of discipline-specific BGK, such as topic familiarity and previous education, and analyze them in terms of their relation to reading test performance. The study has also looked at difficulty level of the passages and tried to establish a connection between difficulty of the passage and the groups' performance on it. Table 4.12 repeated the research questions raised in the study and states the conclusions provided by the various statistical procedures.

The most significant observation arising from the results for RQ1 has to do with the performance of group LMS. The fact that it performs comparatively badly on all the three passages of the test suggests that the group is of low linguistic proficiency. However, the fact that it does not perform worse than the other groups on P(LMS), the one related to the group's field of study, provides some evidence that the group might be utilizing aspects of BGK to answer the questions.

The fact that students of BSSS and PST perform comparatively well on the passages suggests that they are of higher linguistic proficiency. However, linguistic proficiency does not help them in dealing with P(LMS), which might again indicate that the passage requires some background knowledge, inaccessible to students of BSSS and PST. The fact that BSSS

and PST groups perform similarly on the passages related to their fields of study shows that the passages are not specific enough to require the use of BGK.

According to the results for RQ2, group LMS has the highest number of items with the lowest facility indices. Because those items include both passage-related and passage-unrelated items, our previous supposition that the group is of low linguistic level is even more plausible. However, when P(LMS) is analyzed, it is obvious that group PST is having problems with its items as well. In contrast to LMS students, PST students experience difficulty with items that test passage-related knowledge only. A possible explanation for this phenomenon could be PST students' over-reliance on linguistic proficiency and practice of surface reading.

Results for RQ3 show that groups do differ in their perception of passage difficulty. For group BSSS, passages P(BSSS) and P(PST) are of equal ease, while for group PST, passage P(PST) is the easiest. Interestingly, P(LMS) is the most difficult for both of the groups, but not for group LMS, for which P(PST) and P(LMS) are of equal difficulty. These comparisons reveal that passages related to groups' fields of study are perceived to be either the easiest or at least as difficult as the other passages. Moreover, LMS students' failure to choose P(LMS) as the most difficult one shows that the group may be utilizing the background knowledge of the field

Results for the last two RQs did not add to our interpretation of the effect of discipline-specific BGK and its influence on reading test performance. Neither being familiar with the topic of the passage nor having had previous education in the area related to the passage are significant variables in the assessment of reading test performance in our study.

In general, the answer to the major question of the study is that only certain students utilize their discipline-specific BGK when dealing with reading tests. More specifically, students with lower general language ability (LMS, in our case) benefit from their BGK. One of the minor questions was to analyze the specificity level of the passages, and it appears that P(BSSS) is the least specific and least difficult in terms of grammar and lexicon, while P(LMS) and P(PST) are more specific and more difficult in terms of both structure and word choice. As for such variables as previous education and topic familiarity, no effect of these was found on students' reading test performance.

Insights for Researchers, Teachers, and Test Developers

Even though the results of the current study appear to be quite controversial, several important insights still emerge.

Insights for Researchers

Alderson (1981) was among the first researchers in ESP to question the use of ESP tests in EAP settings. He acknowledged that discipline-specific tests were better measures of students' reading comprehension than were general tests and were more suitable because various departments of diverse universities had different requirements for the students. However, he was not sure whether it was possible to compile tests with passages of specific content. More specifically, he questioned the idea of "specificity" and asked "how specific is specific." In his later work, Alderson (1988) went a step further and asked what text could be considered general.

Hale (1988) in his study of TOEFL suggested that texts of EAP tests differ in the extent of specificity. The results of our study show that the English Placement Test, a test of English for General Purposes, has comparatively easy passages with "general" topics. However, they also show that one of the passages, P(LMS), is more specific than the others, and requires some BGK. Thus, Alderson's idea of relativity of the notions 'specific vs. general texts' appears to be reasonable. Therefore, studies into the effect of BGK on reading could and probably should be performed not only in the area of ESP and EAP, but in the area of EGP as well.

Findings concerning text specificity are not the only important insights of this study. Similarly important is the finding that students of lower linguistic proficiency do utilize discipline-specific BGK. Even though a specific level at which students start using their BGK has not been measured here, the idea proposed by Clapham (1996) in her study of BGK about levels of linguistic proficiency under and above which BGK is not activated has been supported. It has also been found that students of various proficiency levels experience problems with different kinds of items. It would be interesting to analyze the factors influencing item difficulty, including factors related to BGK.

Finally, although familiarity with the topic of the passage was not found to be an advantage in answering the passage questions, it is still suggested that more research is needed in this area. Lack of significance in our results probably stemmed from the fact that the topics were considerably general. Whether there is a level of topic specificity at which familiarity of the topic plays a role in reading performance would be a question to investigate.

Insights for Teachers

Teachers, especially reading teachers, may benefit from this research by applying its results to their practice. In heterogeneous groups of students with different language proficiency levels, it is recommended that teachers use passages of content that might be familiar to the group with the lowest level of proficiency. This way, students with a higher level of linguistic ability will benefit from learning new material, while students of lower proficiency will benefit from knowing the content. Since reading involves an interactive process of decoding the language and processing the information, the second part will be easier for those who have already had some exposure to that topic.

In addition to developing effective reading materials, teachers may benefit from the information presented in this study in several other ways. One of them is using the test scores and item analysis. By looking at what items students of different levels have problems with, teachers will be able to produce specific exercises and will know what to focus on, be that reading strategies, vocabulary items, or content.

It will be of interest for the teachers to know that students of different levels of linguistic proficiency differ in their reading styles. While students of lower level may compensate for their knowledge by using careful reading techniques, students of higher proficiency level may over-rely on their linguistic knowledge and perform badly due to surface reading of the texts.

Finally, teachers will benefit just by realizing that students that compose their classes are of various educational backgrounds. They do not come to English reading courses without any knowledge; instead, they bring their experiences and interests. Materials related to their interests and backgrounds will definitely help them in learning.

Insights for Test Developers and Users

Both test developers and test users or administrators can apply the findings of the study in different ways. The ISU English Placement Test developers and administrators might be interested in the following suggestions. The test consists of passages with varying numbers of items following them; the highest number of items belongs to P(LMS), the passage that is most difficult for students of all areas. First, I suggest that passages have an equal number of items no matter what topic they are on. Second, if the passages have different numbers of items, the passage with the highest difficulty level should not be followed by significantly larger amount of items.

A substantial portion of the test items of the EPT seems to be focusing on students' vocabulary. This could be justified because the test is intended to separate students with the lowest level of reading proficiency from everybody else. However, having vocabulary questions that are sometimes easily answered without reading the passages leads to the surface reading on the part of students with higher levels of linguistic proficiency. Having a bigger variety of item types could eliminate this problem.

Finally, the EPT has few items that appear to be of particular difficulty for a great number of students, including students of higher levels of English proficiency. Items 5 and 10, for example, turned out to be very difficult even for the students of group BSSS (item difficulty indices lower than 60). This could imply that either the wording of the items is not clear to the test takers or that the item is on the material not presented in the passage. In any case, more work and revision is needed in relation to items.

Test analysts, in general, should take into account item bias, passage specificity, passage difficulty in terms of language and structure when looking into the effects of discipline-specific background knowledge on reading test performance. When developing a test, specialists need to be aware of the fact that certain kinds of items (the ones unrelated to the passage, in our case) present difficulty to some students (intermediate level of English proficiency, in our case), while other students (e.g., lower level proficiency students) have problems with all kinds of items. They also need to be aware that passages of specific discipline content could be comparatively easy for all students because of the writing style employed, while passages of less discipline specific content can be more difficult for the

readers because of the language choice. Thus, several aspects of the passages should be taken into account when compiling a test and analyzing passage difficulty and specificity levels.

Limitations of the Study

Now that I have outlined the positive features and findings of the study, it is time to turn attention to some of the negative aspects. Ironically, though a large section of the study is devoted to the review of the methodologies employed by various studies into the effect of BGK on reading test performance, the major problems of this study are methodological as well. In the center of the most important limitations lie the two classification systems employed here. First and foremost, the classification of colleges into broader areas of study may not be reliable. This is due to the fact that the available information consisted of the code of the college the students were enrolled in. The colleges at ISU are, however, so diverse, that math, chemistry, and English majors, for example, are found in one college, LAS. No records were available about the students' specific departments. Coding students according to specific departments could have been considerably more valid.

The second problem lies with the second classification system, used to assign the passages to certain areas of study. This was done on the basis of Clapham's (1996) classification which turned out to be confusing to the raters helping this study. Moreover, the passages were such that could have been classified into two categories, which according to Clapham (1996) are separate from each other. For example, P(PST) could belong to both PST and LMS, according to the classification; while P(LMS) could be both LMS and BSSS. Again, a more careful classification system should have been used.

Another problem was the amount of data. Due to the fact that the Students' Questionnaire was performed comparatively late in the progress of the study, not many responses were collected. Therefore, statistical significance test results could be misleading. A warning for future researchers should be, "Start collecting data as soon as you can!"

Because of the number of items employed after each passage of the Reading Subtest and because of the ease of the passages, the score distribution was found not to follow the normal distribution; therefore, parametric tests of statistical significance were not possible to use. Instead, alternative forms of significance tests were used, specifically non-parametric

chi-square analysis and multinomial logistic regression. Texts with larger numbers of items would be more informative.

Future Research

Based on the limitations of the current research, future researchers are advised to follow some precautions in devising the methodology of their study. First of all, it will be important to use classification systems that would describe the unique nature of the subjects and the passages. Second, it will be beneficial to predict as many variables influencing the reading test performance as possible and control for them in model construction. It will be advisable to have tools to collect information about the additional variables as soon as possible so that more data can be collected for the analysis. Finally, and most importantly, it will be necessary to prepare a measure of linguistic proficiency of the students so that to control for the level of linguistic proficiency each student possesses.

While doing this study, I came to realize that test items play a significant role in students' overall performance on the test. If I were to continue this study, I would therefore focus on issues related to item difficulty, item specificity, item bias, item clarity, etc. I would be interested in analyzing the relationship between item type and field of study, number of items and performance on the test, etc. Beyond the level of items, I would be interested in examining text features that make text specific, such as grammatical-syntactic constructions, lexicon, structure, etc, and separating out the 'most specific' features. I would like to find those levels of linguistic proficiency at which BGK starts operating, levels of BGK at which it becomes effective, etc. In a word, there is a whole universe of unexplored phenomena related to BGK and reading performance, and my exploration of it is just starting.

APPENDIX A: READING SUBTEST**Managing the Global Greenhouse**

"The world is warming. Climatic zones are shifting. Glaciers are melting. Sea level is rising. These are not hypothetical events from a science fiction movie; these changes and others are already taking place, and we expect them to accelerate over the next years as the amounts of...gases accumulating in the atmosphere through human activities increase.... A rapid and continuous warming will not only be destructive to agriculture but also lead to the widespread death of forest trees, uncertainty in water supplies and the flooding of coastal areas" (Houghton & Woodwell, 1989).

For some years now, warnings like this have been heard from leading authorities in the scientific community. According to these experts, we are leaving our children a frightening legacy: an accumulation of so-called greenhouse gases in the atmosphere and the potentially disastrous climate changes that this build-up may bring about. However, the scientific community is not speaking with one voice. Other leading scientists point out that the evidence for greenhouse warming is inconclusive and argue that predictions based on it are questionable. The scientific debate has been intense. It has also fueled the political controversy about what measures, if any, need to be taken to address the possible problem of greenhouse warming.

In the presence of scientific debate and political controversy, what is a concerned public to think about greenhouse warming? For an adequate assessment of the issue, an essential first step is to identify what is known and what is not yet known about the phenomenon.

First, there is unanimous scientific agreement that gases like carbon dioxide (CO₂), chlorofluorocarbons (CFCs), and methane (CH₄) have the potential to produce a greenhouse effect. These relatively transparent gases allow sunlight to pass through and warm the earth; however, when that heat is released by the earth in the form of infrared radiation, it is absorbed very efficiently by these gases and not allowed to escape out into space.

Also undisputed is the fact that any potential effects of greenhouse gases will be both long-term and global. Sulfur dioxide (SO₂) and nitrous oxides (NO), which are the primary causes of acid rain and photochemical smog, only remain in the atmosphere for days or weeks; their effects are local or regional rather than global. Carbon dioxide, methane, and CFCs, however, remain in the atmosphere between ten and one hundred years and clearly do not stay localized in the areas where they are originally released or in adjacent regions. They spread throughout the global atmosphere. Their long life means that their effects are likely to be felt more by our descendants than by our current generation.

45 Despite these facts, many scientists are reluctant to attribute the clear
global warming trend of the last one hundred years to the buildup of
greenhouse gases in the atmosphere. Such a conclusion, they argue, is not
justified by the present evidence, which merely shows that a slight global
warming trend has occurred at the same time as concentrations of
50 greenhouse gases in the atmosphere have been climbing. In addition, the
temperature increase of 0.5 degrees C that has been established for the last one
hundred years is at the lower end of the range of the increase predicted by
proponents of greenhouse theory. This suggests to some scientists that, at the
very least, the effects of the greenhouse gases on the global climate have been
exaggerated.

55 Some politicians and governments have used the lack of scientific
certainty on greenhouse warming as a justification for not taking immediate
action to control it. However, the majority of scientists investigating global
warming warn that it would be extremely risky to wait for greater certainty.
60 Much work remains if we are going to understand and control potential
greenhouse warming.

1. Which statement best expresses the main idea of the passage?
 - a. Scientists disagree about the seriousness of greenhouse gases, and therefore there is no cause for concern.
 - b. There are some known facts about the existence and potential danger of some gases even though their seriousness is debated.
 - c. Acid rain and photochemical smog remain in the atmosphere for days or weeks and have a local effect, but other gases remain longer and have effects beyond the region.
 - d. More research needs to be conducted to understand the significance of gases such as carbon dioxide and methane.

2. What is the meaning of "the scientific community is not speaking with one voice" in line 14-15?
 - a. Scientists do not agree about the seriousness of greenhouse gases.
 - b. The climate changes that the greenhouse gases bring will not be serious.
 - c. There is a political controversy about what should be done about greenhouse gases.
 - d. Scientists do not speak to the public about the dangers of greenhouse gases.

3. What is absorbed by gases like carbon dioxide (CO₂), chlorofluorocarbons (CFCs), and methane (CH₄).
 - a. Heat.
 - b. Sunlight.
 - c. Transparent gases.
 - d. Greenhouse gases.

4. What does "undisputed" mean in line 33? (the first line in paragraph 5)
 - a. Unproven.
 - b. Not well-understood.
 - c. Debated.
 - d. Agreed upon.

5. According to the passage, why do some scientists think that the greenhouse gases may not be a serious problem?
 - a. Because temperature increases are small, global warming may not be occurring.
 - b. Greenhouse gases may not have caused the global warming trend.
 - c. Sulfur dioxide and nitrous oxides do not spread but remain in one region.
 - d. The time that carbon dioxide, methane, and CFCs stay in the atmosphere has been exaggerated.

6. What is the meaning of the word "proponents" in line 51?
 - a. opponents.
 - b. supporters.
 - c. researchers.
 - d. reporters.

7. What is the reason that people are concerned about global warming?
 - a. Global warming can cause political controversies.
 - b. Some gases allow sunlight to pass through but do not allow heat to escape.
 - c. Scientists disagree about the causes of global warming.
 - d. Global warming may cause long term climate changes.

Culture Shock

1 Culture shock is a common experience for a person learning a second language in a second culture. Culture shock refers to phenomena ranging from mild irritability to deep psychological panic and crisis. Culture shock is associated with feelings in the learner of estrangement, anger, hostility, 5 indecision, frustration, unhappiness, sadness, loneliness, homesickness, and even physical illness. The person undergoing culture shock views his new world out of resentment, and alternates between being angry at others for not understanding him and being filled with self-pity. Edward Hall describes a hypothetical example of an American living abroad in Japan for the first time:

10 At first, things in the cities look pretty much alike. There are taxis, hotels with hot and cold running water, theaters, neon lights, even tall buildings with elevators and a few people who can speak English. But pretty soon the American discovers that underneath the familiar exterior there are vast differences. When someone says "yes" it often doesn't mean they are 15 pleased. When the American visitor makes a helpful gesture he may be rebuffed; when he tries to be friendly nothing happens. People tell him that they will do things and don't. The longer he stays, the more enigmatic the new country looks.

20 This case of an American in Japan illustrates the point that initially the person in a foreign culture is comfortable and delighted with the "exotic" surroundings. As long as he can perceptually filter his surroundings and internalize the environment in his own world view, he feels at ease. As soon as this newness wears off...he becomes disoriented.

25 It is feasible to think of culture shock as one of four successive stages of acculturation. The first stage is the period of excitement and euphoria over the newness of the surroundings. The second stage—culture shock—emerges as the individual feels the intrusion of more and more cultural differences into his or her own image of self and security. In this stage the individual relies on and seeks out the support of his or her fellow countrymen in the second culture, 30 taking solace in complaining about local customs and conditions, seeking escape from his or her predicament. The third stage is one of gradual, and at first tentative and vacillating, recovery. This stage is typified by what Larson and Smally call culture stress: some problems of acculturation are solved while other problems continue for some time. But general progress is made, slowly but 35 surely, as the person begins to accept the differences in thinking and feeling that surround him, slowly becoming more empathetic with the persons in the second culture. The fourth stage represents near or full recovery, either assimilation or adaptation, acceptance of the new culture and self-confidence in the "new" person that has developed in this culture.

8. The main idea of this passage is
- culture shock can cause deep psychological panic.
 - culture shock occurs to Americans living in Japan.
 - culture shock happens to everyone who travels abroad.
 - culture shock has predictable symptoms and characteristics.
9. According to the passage, which of these statements is false?
- People entering a new culture are excited at first.
 - People suffering from culture shock experience a variety of symptoms.
 - People entering a new culture experience the same level of culture shock.
 - People living in a new culture gradually become able to deal with the differences.
10. The word "solace" in line 30 probably means
- anger.
 - sadness.
 - comfort.
 - isolation.
11. This passage implies that
- the stages of culture shock are unpredictable.
 - most people living in a new culture will experience culture shock.
 - disliking the customs of another culture is the only reason for culture shock.
 - culture shock occurs as soon as a person notices the newness of his or her surroundings.
12. In stage three, a person in a new culture often
- begins to feel somewhat more comfortable.
 - feels the cultural differences most strongly.
 - escapes his or her problems by returning home.
 - has solved his or her problems with living in the new culture.
13. The word "successive" in line 24 means
- getting weaker and weaker.
 - happening at the same time.
 - happening one after another.
 - getting stronger and stronger.
14. The word "assimilation" in line 37 means
- rejecting the new culture.
 - rejecting one's home culture.
 - returning to one's home culture.
 - becoming comfortable in the new culture.

Go on to the next passage.

Ancient Wheat

According to two British experts who are studying what people ate in the days before recorded history, the ancient Greeks may have been baking high quality bread long before commonly thought. Researchers Terry Brown and Glynis Jones have been examining how early people's choices of food may have shaped the destiny of humanity. They are delving back to when history was preserved on a molecular scrapbook rather than on stone or parchment.

"We have discovered that it is possible to obtain small traces of DNA from preserved wheat seeds, some dating back to the earliest stages of agriculture," Brown and Jones note in their first report on their findings. This report, "New Ways with Old Wheats" suggests that Bronze Age Greeks could have begun baking high-quality bread as early as 3300 B.C. That was centuries before experts have believed the types of wheat existed that were needed to produce bread similar to modern varieties.

Brown, a molecular biologist at the University of Manchester Institute of Science and Technology, and Jones, an archaeobotanist at the University of Sheffield, hope to use modern DNA techniques to answer questions about such things as how farming spread, a central factor in the progress of ancient civilization. In their study, Brown and Jones looked at charred wheat grains discovered in Assiros, a small Bronze Age community in northern Greece. What they found was that the Bronze Age wheat had good bread-making qualities, something that historians had not thought would be possible at that time.

Because the variety of wheat used to make bread does not occur naturally, it had to be developed through crossbreeding. DNA markers in the ancient wheat suggested a husked variety known as spelt, which experts previously did not think had become known in the Mediterranean area until the 1st century. Husked wheat, which produces bread similar to modern "stone-ground" varieties, contains a hard covering that is commonly broken off with a stone grinder.

The findings also suggest wheat has been domesticated twice and in two different places, rather than only once as was believed previously. Wheat was thought to have first been domesticated in the fertile crescent, an area that crosses parts of modern day Iraq, Iran, and Jordan. However, a parallel development appears to have been going on in northern Greece. The findings could alter the way historians render the timeline and location of ancient farming methods and technology.

45 There has been an unexpected contemporary "spin-off" from the DNA research, according to Brown. Because it is sometimes difficult to differentiate between some types of ground or processed wheat, the DNA technique now will let millers test a grain shipment's molecular and genetic makeup. This will allow them, for example, to determine if more expensive durum wheat used to make pasta has been adulterated with cheaper bread wheat. "We didn't set out to solve that problem," Brown said.

15. The most important finding reported in the excerpt is
 - a. that bread-making wheat was being grown in Greece earlier than had been realized
 - b. that people in the fertile crescent were the first people who were able to grow wheat
 - c. that a husked variety of wheat was very similar to modern "stone-ground" varieties
 - d. that DNA techniques can be used to differentiate between types of processed wheat
16. Why was it possible for Brown and Jones to do this type of research?
 - a. because charred wheat grains were discovered in Assiros, Greece
 - b. because historians now know more about agriculture in the fertile crescent
 - c. because modern bread-baking technology gives insights into the past
 - d. because new DNA sampling techniques can be applied to historical research
17. The phrase "when history was preserved on a molecular scrapbook rather than on stone or parchment" in line 5 means that
 - a. ancient people saved information in scrapbooks
 - b. historians need to learn to read scrapbooks
 - c. historians can learn from modern genetic research techniques
 - d. ancient people wrote on parchment and stone
18. Where did Brown and Jones get their DNA samples?
 - a. from wheat grown in Greece
 - b. from bread baked in the Bronze Age
 - c. from ancient stone grinders
 - d. from grain shipments

19. What is the main idea of paragraph 3 (lines 17-25)?
- Brown is a molecular biologist and Jones is an archeologist
 - Brown and Jones were using modern DNA techniques
 - Brown and Jones looked at wheat grains from a community in Greece
 - Brown and Jones found that Bronze Age wheat could have made good bread.
20. The meaning of "spelt" (line 29) is
- the spelling on an ancient manuscript
 - a type of wheat with an outer covering
 - a DNA marker in ancient wheat
 - the part of the wheat broken off in a stone grinder
21. What is "the fertile crescent"?
- a place in the Mediterranean area near Greece
 - any place that is very good for growing crops
 - a region that includes modern day Iran and Iraq
 - an area where people did not bake bread
22. Learning about the food that ancient people ate interested Brown and Jones because
- they wanted to create a test that could be used for grain shipments.
 - they wanted to apply DNA techniques to improve the production of wheat.
 - they were interesting in baking the type of bread eaten in the Bronze Age.
 - they were interested in learning about the history of agriculture.
23. The meaning of "husk" in the phrase "husked variety" (line 29) is
- a stone-ground variety of wheat
 - the outer covering on wheat
 - a DNA marker in ancient wheat
 - a naturally occurring type of wheat
24. Why is it significant that crossbreeding was necessary to produce the variety of wheat identified by Brown and Jones?
- It shows that early people in Greece were not able to eat bread.
 - It shows that agricultural advances had taken place in Greece.
 - It shows that early Greeks used stone grinders to prepare their wheat.
 - It shows that wheat was an important part of people's diet in Greece.

25. The meaning of "alter" in line 40 is
- support
 - render
 - change
 - Inform
26. What is the main idea of paragraph 5 (lines 35-41)?
- Wheat was transformed from the wild and cultivated in more than one place
 - Wheat was first cultivated in the fertile crescent
 - Historians are reevaluating the farming practices in early Iran, Iraq, and Jordan.
 - Historians have already altered the timeline and location of ancient farming practices.
27. The meaning of "spin-off" in line 43 is
- surprising and modern research
 - differentiating between kinds of wheat
 - the spinning process used in DNA research
 - a result coming from a new direction
28. The meaning of "differentiate" in line 44 is
- to tell apart
 - to process
 - to make something different
 - to test
29. The meaning of "adulterated" in line 48 is
- paid for by millers sending grain shipments
 - solved by modern scientific methods
 - made impure by adding inferior ingredients
 - determined by examining DNA in wheat
30. Why is it important to know the kind of wheat that was being grown in ancient societies?
- It allows historians to predict when farmers first learned to cross-breed wheat.
 - It tells historians when early people could first eat bread.
 - It gives information about where early developments in agriculture were occurring.
 - All of the above.

End of reading test.

Instructions:

Read each of the passages in the test and then answer the questions which follow. Answer by choosing the best response and darkening the circle on your answer sheet that corresponds to it. When you finish with one passage, you should proceed to the next passage without stopping.

Example:

You read the following:

The complex phenomenon of growth has been described simply as an increase in mass of a body of matter which is usually also correlated with an increase in volume. Growth, which results primarily from the production of new protoplasm, includes variations in form—some the result of inheritance, some the result of environmental response.

Then, you answer the questions.

Growth is described as

- an increase in the varieties and inheritance of matter.
- an increase in the mass and volume of a body of matter.
- an increase in the inheritance and environmental responses of a body.
- an increase in the environmental response and mass of a body of matter.

You choose response b and then darken response b on the answer sheet:

A B C D

Growth results primarily from

- new inheritances.
- the production of varieties.
- the variations in protoplasts.
- the production of new protoplasm.

You choose response d and then darken response d on the answer sheet:

A B C D

APPENDIX B: HUMAN SUBJECTS APPROVAL

OFFICE USE ONLY		
EXPEDITED <input checked="" type="checkbox"/>	FULL COMMITTEE <input type="checkbox"/>	ID# 01-281

PI Name: Svetlana Chigayeva Title of Project: Interaction between students' fields of study and the content of the reading passages of the English Placement Test

Checklist for Attachments

The following are attached (please check):

13. Letter or written statement to subjects indicating clearly:

- a) the purpose of the research
- b) the use of any identifier codes (names, #'s), how they will be used, and when they will be removed (see item 18)
- c) an estimate of time needed for participation in the research
- d) if applicable, the location of the research activity
- e) how you will ensure confidentiality
- f) in a longitudinal study, when and how you will contact subjects later
- g) that participation is voluntary; nonparticipation will not affect evaluations of the subject

14. A copy of the consent form (if applicable)

15. Letter of approval for research from cooperating organizations or institutions (if applicable)

16. Data-gathering instruments

17. Anticipated dates for contact with subjects:

First contact	Last contact
01/03/00	
Month/Day/Year	Month/Day/Year

18. If applicable: anticipated date that identifiers will be removed from completed survey instruments and/or audio or visual tapes will be erased:

01/10/01
Month/Day/Year

19. Signature of Departmental Executive Officer _____ Date 12/4/00 Department or Administrative Unit _____

Charles [Signature]

20. Initial action by the Institutional Review Board (IRB):

Project approved _____ Date _____ Pending Further Review 12/7/00 Date _____ Project not approved _____ Date _____

No action required _____ Date _____

21. Follow-up action by the IRB:

Project approved 12-12-00 Date _____ Project not approved _____ Date _____ Project not resubmitted _____ Date _____

Patricia M. Keith _____ 12-12-00 _____ *pmkeith* _____
Name of IRB Chairperson Date Signature of IRB Chairperson

APPENDIX C: CONSENT FORM

Interaction between students' fields of study and the content of the reading passages of the English Placement Test

You are invited to be in a research study of the effects of field-specific background knowledge on student performance on the Reading Subtest of the English Placement Test offered by Iowa State University to all international students. You were selected as a possible participant because you are an international student who has just taken this test. We ask that you read this document and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Svetlana Chigayeva, an MA student of English, Iowa State University.

Background Information: The purpose of this study is to examine whether the Reading Subtest of the English Placement Test (EPT) is biased towards a particular group of students. We believe that the passages used in the subtest convey information specific to certain fields. Therefore, students coming from these fields might do better than students coming from other fields of study.

Procedure: If you agree to be in this study, we will ask you to complete the questionnaire below. The approximate time needed to complete the questionnaire is 15 minutes.

Risks and Benefits of Being in the Study: As participants, your risks will be minimized because your personal information will not be released by the researcher. The researcher will delete the information about your name and social security number as soon as possible.

While there are no direct benefits offered by this study, you will have a chance to test your reading ability and to analyze your reading skills and background knowledge.

Confidentiality: The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify you as a subject.

Voluntary Nature of the Study: If you decide not to participate in filling the questionnaire, no penalty will be imposed on you.

You may ask any questions you have now.

If you have questions later, you may contact the researcher at sveta@iastate.edu.

If you want to talk to someone other than the researchers, contact Human Subjects Review Office at 221 Beardshear Hall, 294-4566.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature _____ Date _____

Name _____

Signature of Investigator or Person Obtaining Consent

_____ Date _____

APPENDIX D: STUDENTS' QUESTIONNAIRE*

1. Age: _____
2. Gender (circle one): male female
3. Nationality: _____
4. Country of Birth: _____
5. What is your first language (the main language you speak at home): _____

PREVIOUS EDUCATION

School Education

6. Which of the following subjects did you study most of all during the last 2-3 years of your school education? Please, check all the appropriate boxes.

- | | | | | | |
|-------------------------|--------------------------|-----------------|--------------------------|-----------|--------------------------|
| English | <input type="checkbox"/> | Social Sciences | <input type="checkbox"/> | Chemistry | <input type="checkbox"/> |
| Other Foreign Languages | <input type="checkbox"/> | Biology | <input type="checkbox"/> | Religion | <input type="checkbox"/> |
| Literature | <input type="checkbox"/> | Physics | <input type="checkbox"/> | Geography | <input type="checkbox"/> |
| History | <input type="checkbox"/> | Mathematics | <input type="checkbox"/> | Economics | <input type="checkbox"/> |

Other subjects: _____

University or College Education (if applicable)

7. Please, fill in the information about the universities/colleges you might have attended before coming to Iowa State University.

Name of the university/college and the country	Major(s)	Minor(s)

FUTURE COURSE OF STUDY

8. What major(s) will you pursue at ISU? _____

* This questionnaire is an adaptation of Clapham's (1996) "Students' Questionnaire" (267-272).

9. What minor(s) will you pursue at ISU? _____
10. Level of study (circle one): graduate _____ undergraduate _____

BACKGROUND KNOWLEDGE

11. Think about the reading you do for your school and professional work or during your free time. Do you read books, magazines, academic papers or newspaper articles on any of the following subjects? Please circle 1, 2, or 3.

	Often	Sometimes	Never
Psychology	1	2	3
Chemistry	1	2	3
Natural Environment	1	2	3
Sociology/Anthropology	1	2	3
Geography	1	2	3
History	1	2	3
Biology	1	2	3

READING TEST

You have just finished the reading test. Think about the three passages of the test and answer the question related to them.

12. Questions to which passage do you think you answered best? (Circle the appropriate number.)

“Managing the Global Greenhouse”	P(BSSS)	P(LMS)
1	2	3

Reading Passage 1: “Managing the Global Greenhouse”

13. Was this passage easy or difficult? (Circle the appropriate number):

Very easy						Very difficult
1	2	3	4	5	6	

14. Were you familiar with the problem of greenhouse effect before you read the passage? Yes/No

15. If you were familiar with this problem, did this help you to answer the questions? Yes/No

Reading Passage 2: P(BSSS)

16. Was this passage easy or difficult? (Circle the appropriate number):

Very easy						Very difficult
1	2	3	4	5	6	

APPENDIX E: RATERS' QUESTIONNAIRE

I. Classify each passage as belonging to one of the three categories: business studies and social studies, life and medical sciences, or physical science and technology.

Text A _____
 Text B _____
 Text C _____

1. Business Studies and Social Science

Archeology, Art History, Business Studies, Development Studies, Economics, Education, English Language Teaching, English Literature, Estate Management, Finance, Geography, History, Languages, Law, Linguistics, Music, Philosophy, Political Science, Population Studies, Psychology, Public Administration, Sociology, Social Anthropology, Theater Studies, Theology, Urban Development

2. Life and Medical Sciences

Agriculture, Agronomy, Animal Nutrition, Bacteriology, Biology, Clinical Tropical Medicine, Community Health, Dentistry, Ecology, Environmental Science, Epidemiology, Forestry, Genetics, Immunology, Land Protection, Medicine, Microbiology, Nutrient Enrichment, Obstetrics and Gynecology, Ophthalmology, Pediatric Medicine, Parasitology, Pathology, Physiology, Plant Physiology, Toxicology, Veterinary Science, Virology, Zoology

3. Physical Science and Technology

Architecture, Chemistry, Civil Engineering, Computer Studies, Construction Management, Control Systems, Electrical Measurement, Electronics Engineering, Energy, Exploration Geophysics, Fermentation Technology, Fluid Mechanics, Geochemistry, Geology, Hydrology, Laser Spectroscopy, Mathematics, Mechanical Engineering, Metallurgy, Meteorology, Petroleum Geology, Physics, Soil Mechanics, Water Resources

II. Analyze the passages at the following levels:

Rubric

This facet relates to the instructions to test takers about how they should proceed in each part of the test. Ratings should be placed on the following scale:

RUBRIC 0 = clear for unprepared test takers
 1 = possibly unclear for unprepared test takers
 2 = unclear for unprepared test takers

Your rating _____

Propositional Content

Vocabulary (passages and items)

Answer this in relation to the specific group of test takers for whom the test is intended. In the case of English Placement Test, the test takers are ESL students who are starting their studies at Iowa State University.

NB: These facets apply not only to words but also to fixed and idiomatic expressions that may be relatively infrequent, specialized, or ambiguous.

INFREQUENT (Frequent) 0 1 2 (Infrequent)

Text A

Text B

Text C

SPECIALIZED (General) 0 1 2 (Specialized)

(e.g., technical,
jargon, slang)

Text A

Text B

Text C

AMBIGUOUS (Clear) 0 1 2 (Ambiguous)

Text A

Text B

Text C

(Ambiguity refers to the possibility of more than one reading, or interpretation, of a phrase, sentence, or text. For multiple-choice items, this could arise if the keyed response results in an ambiguous sentence, or if there is more than one possible answer.)

Degree of Contextualization

In rating this facet, consider the relevant proportion of “new” to “contextual” information. “New information” is that which is not known to the test taker and cannot be predicted from the context. “Contextual information” is that which is developed in the passage itself. Thus, a passage is “not at all contextualized” if there is a lot of new information in the passage that is not explained through definitions, examples, paraphrase, etc. The passage is “highly contextualized” if there is no new information, or if the new information is explained. If the reader has prior knowledge that will help comprehension, then the text is contextualized. If the reader does not have relevant prior knowledge, the discourse is context reduced.

Input can be contextualized in terms of two types of information: cultural and that which is topic specific. Cultural content relates to national (general) culture such as national habits, customs, and beliefs. Ratings on this facet should be as follows:

	Highly contextualized		Not at all contextualized
CULTURAL	0	1	2
CONTENT			
Text A			

Text B			
Text C			
TOPIC	0	1	2
SPECIFICITY			
Text A			
Text B			
Text C			

Relationship of Item to Passage

This facet should be rated in terms of the extent of text to which the item relates and whether the item requires the test taker to relate information in the passage to the real world (i.e., to the test takers' knowledge schemata). For purposes of this rating, "specific part" means "one sentence or several contiguous sentences". Items should be rated on the following scale:

- 0 = No relationship to passage; item can be answered without reference to the passage, or relationship of item to passage is not clear.
- 1 = Item relates to one specific part of the passage and requires only localized understanding of that part. If this is the case, write "1" even if it is possible to reach the answer by referring to more than one part of the passage.
- 2 = Item relates to more than one specific part of the passage or requires the test taker to relate one specific part to one or more others.
- 3 = Item relates to the entire passage and requires an understanding of the entire passage.
- 4 = Item relates to one specific part of the passage, requires only localized understanding of that part, and requires the test takers to relate information in that part to the real world.
- 5 = Item relates to more than one specific part of the passage or requires the test taker to relate one specific part to one or more others, and requires the test taker to relate the information in those parts to the real world.
- 6 = Item relates to the entire passage, requires an understanding of the entire passage and requires the test taker to relate information in the passage to the real world.

Please, mark every item in the test booklet.

Topic

This facet has to do with the topic, or "subject", of the text, and not whether the test taker is American, or an academic, or in a specialized area. Thus, for example, a text that has a great deal of specific American cultural content is highly specific to this category and would be rated "2", irrespective of whether a given test taker is of that background or orientation. Note, therefore, that for this facet the test taker should not be taken into account.

	Not at all specific		Highly specific
CULTURE	0	1	2
SPECIFIC			
Text A			
Text B			
Text C			

ACADMEIC SPECIFIC	0	1	2
Text A			
Text B			
Text C			
SPECIALIZED TOPIC	0	1	2
Text A			
Text B			
Text C			

Organizational Characteristics

Grammar

This relates to the complexity of sentence types and embedding, and the frequency of the passive voice.

	0 (very simple)	1	2 (very complex)
Text A			
Text B			
Text C			

Cohesion

This relates to the use of cohesive devices such as Reference, Substitution, Adversatives, Causals, Temporals, and Lexical Cohesion (Halliday).

	0 (not at all complex)	1	2 (highly complex)
Text A			
Text B			
Text C			

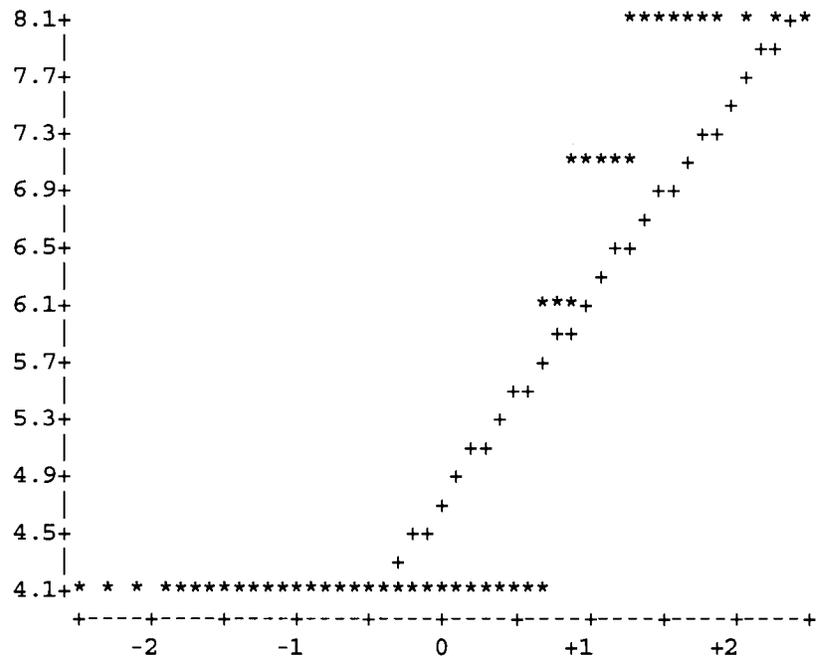
Rhetorical Organization

This facet should be rated in terms of how complex the rhetorical organization is, not on how familiar test takers are with it. In general, instruction and description should be rated 0, comparison and contrast 1, and argumentation 2.

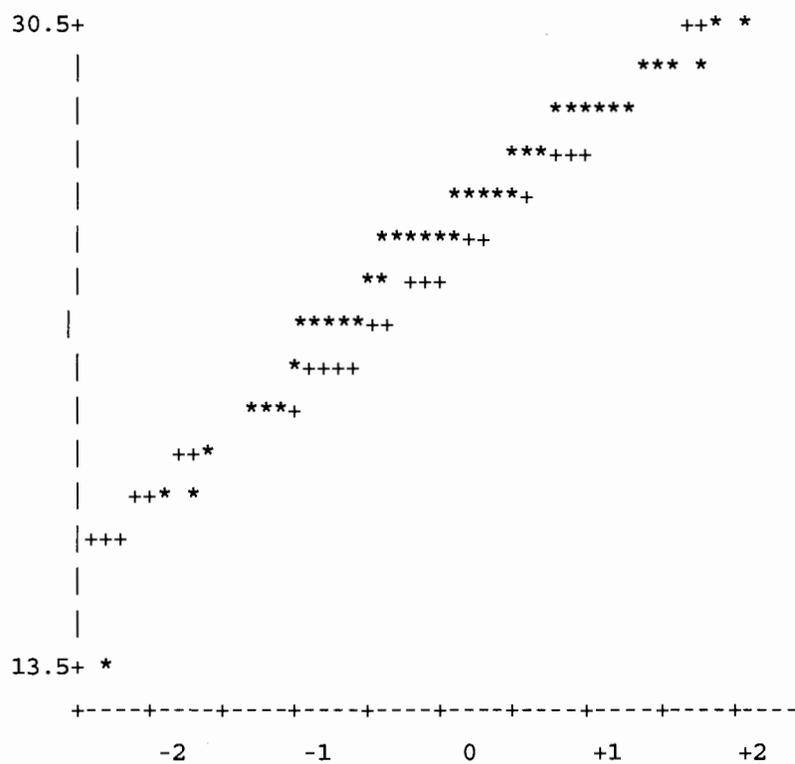
Complexity	0 (very simple)	1	2 (very complex)
Text A			
Text B			
Text C			
Number of specific types of rhetorical organization	1	2	3-more
Text A			
Text B			
Text C			

APPENDIX F: SAMPLES OF NORMAL PROBABILITY PLOTS

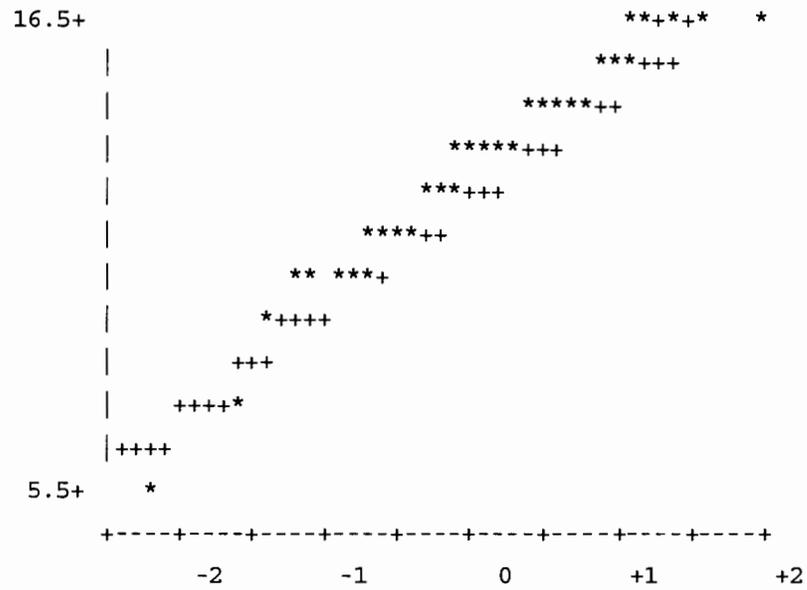
Normal Probability Plot for Group BSSS -- P(BSSS)



Normal Probability Plot for Group PST—Total of Reading Subtest



Normal Probability Plot for Group LMS—P(LMS)



APPENDIX G: MANOVA OUTPUT

The SAS System 1
 21, 2001 19:27 Wednesday, February

The GLM Procedure

Class Level Information

Class	Levels	Values
area	3	1 2 3
test	3	1 2 3

Number of observations 756

The SAS System 2
 21, 2001 19:27 Wednesday, February

The GLM Procedure

Dependent Variable: score

Source	DF	Sum of Squares	Mean Square	F Value
Model	8	8515.7762	1064.4720	5.42
Error	747	146616.2248	196.2734	
Corrected Total	755	155132.0010		

R-Square	Coeff Var	Root MSE	score Mean
0.054894	16.58930	14.00976	84.45059

Source	DF	Type I SS	Mean Square	F Value
area	2	3379.637448	1689.818724	8.61
test	2	4501.749606	2250.874803	11.47
area*test	4	634.389139	158.597285	0.81

Pr > F	Source	DF	Type III SS	Mean Square	F Value
0.0002	area	2	3379.637448	1689.818724	8.61
0.0008	test	2	2833.892946	1416.946473	7.22
0.5202	area*test	4	634.389139	158.597285	0.81

The SAS System

3

19:27 Wednesday, February

21, 2001

The GLM Procedure

t Tests (LSD) for score

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	747
Error Mean Square	196.2734
Critical Value of t	1.96314

Comparisons significant at the 0.05 level are indicated by ***.

area Comparison	Difference Between Means	95% Confidence Limits	
1 - 2	0.5356	-1.7394 2.8107	
1 - 3	6.0594	3.1527 8.9662	***
2 - 1	-0.5356	-2.8107 1.7394	
2 - 3	5.5238	2.3620 8.6856	***
3 - 1	-6.0594	-8.9662 -3.1527	***
3 - 2	-5.5238	-8.6856 -2.3620	***

The SAS System

4

19:27 Wednesday, February

21, 2001

The GLM Procedure

t Tests (LSD) for score

NOTE: This test controls the Type I comparisonwise error rate, not the

experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	747
Error Mean Square	196.2734
Critical Value of t	1.96314
Least Significant Difference	2.4502

Means with the same letter are not significantly different.

t Grouping	Mean	N	test
A	87.245	252	2
A			
A	84.807	252	1
B	81.300	252	3

The SAS System

5

19:27 Wednesday, February

21, 2001

The GLM Procedure

Level of area	Level of test	N	-----score-----	
			Mean	Std Dev
1	1	139	86.6392600	14.2035560
1	2	139	87.8725591	12.7443487
1	3	139	82.0593525	14.0931034
2	1	75	84.5714286	14.2393069
2	2	75	89.1428571	13.4586864
2	3	75	81.2500000	13.1583393
3	1	38	78.5714286	17.8088149
3	2	38	81.2030075	13.3290749
3	3	38	78.6184211	15.9596976

The SAS System

6

19:27 Wednesday, February

21, 2001

The GLM Procedure Multivariate Analysis of Variance

E = Error SSCP Matrix

score 146616.22477

APPENDIX H: TOPIC FAMILIARITY

Logistic Regression Model for BSSS Familiarity as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	1	10.00	0.0016	<.05
	Group	2	2.17	0.3384	NS
	Likelihood Ratio	0			

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-1.2137	0.33838	10.16	0.0016 <.05
Group	2	-0.2526	0.5330	0.22	0.6355 NS
	3	-0.4911	0.5867	0.70	0.4026 NS

Logistic Regression Model for P(BSSS) Score as a Function of Familiarity

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	126.99	<.0001	<.05
	Familiarity	1*			
	Likelihood Ratio	1	0.00	0.9815	

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-0.3121	0.4187	0.56	0.4561 NS
	2	-4.8035	0.4432	117.47	<.0001<.05
Familiarity	3	-0.5352	0.4187	1.63	0.2012 NS
	4	-5.3631			

Logistic Regression Model for PST Familiarity as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	1	2.78	0.0955	NS
	Group	2	2.12	0.3459	NS
	Likelihood Ratio	0	0		

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-0.5580	0.3347	2.78	0.0955 NS
Group	2	0.5580	0.4420	1.59	0.2068 NS
	3	-0.6460	0.5064	1.63	0.2021 NS

Logistic Regression Model for P(PST) Score as a Function of Familiarity

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	7.24	0.0267 <.05	
	Familiarity	2	2.48	0.2893 NS	
	Likelihood Ratio	0	0		
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	1.1168	0.5247	4.53	0.0333 <.05
	2	0.3863	0.5154	7.24	0.0071 <.05
Familiarity	3	-0.2695	0.5247	0.26	0.6075 NS
	4	-0.6931	0.5154	1.81	0.1787 NS

Logistic Regression Model for LMS Familiarity as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	1	14.36	0.0002 <.05	
	Group	2	3.58	0.1674 NS	
	Likelihood Ratio	0			
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-2.0013	0.5282	14.36	0.0002 <.05
	2	0.7068	0.7965	0.79	0.3749 NS
Group	3	0.4836	0.8000	0.37	0.5455 NS

Logistic Regression Model for P(LMS) Score as a Function of Familiarity

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	0.40	0.8173 NS	
	Familiarity	2	0.40	0.8173 NS	
	Likelihood Ratio	0			
Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	0.3180	0.5408	0.35	0.5565 NS
	2	0.0527	0.5503	0.01	0.9237 NS
Familiarity	3	0.3180	0.5408	0.35	0.5565 NS
	4	0.0527	0.5503	0.01	0.9237 NS

APPENDIX I: PREVIOUS EDUCATION**Logistic Regression Model for BSSS Previous Education as a Function of Group**

Maximum Likelihood Analysis of Variance						
	Source	DF	Chi-Square	Pr>ChiSq		
	Intercept	1	0.38	0.5384	NS	
	Group	2	10.76	0.0046	<.05	
	Likelihood Ratio	0				
Analysis of Maximum Likelihood Estimates						
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq	
Intercept	1	0.2361	0.3828	0.38	0.5384	NS
Group	2	1.7025	0.5330	10.20	0.0014	<.05
	3	-1.4686	0.5867	6.27	0.0123	<.05

Logistic Regression Model for P(BSSS) Score as a Function of Previous Education

Maximum Likelihood Analysis of Variance						
	Source	DF	Chi-Square	Pr>ChiSq		
	Intercept	2	0.05	0.9732	NS	
	Prev Educ	2	1.89	0.3879	NS	
	Likelihood Ratio	0				
Analysis of Maximum Likelihood Estimates						
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq	
Intercept	1	-0.0872	0.3904	0.05	0.8233	NS
	2	-0.0204	0.3846	0.00	0.9577	NS
Passage	3	-0.4236	0.3904	1.18	0.2778	NS
	4	-0.4904	0.3846	1.63	0.2023	NS

Logistic Regression Model for PST Previous Education as a Function of Group

Maximum Likelihood Analysis of Variance						
	Source	DF	Chi-Square	Pr>ChiSq		
	Intercept	1	0.48	0.4875	NS	
	Group	2	9.14	0.0103	<.05	
	Likelihood Ratio	0				
Analysis of Maximum Likelihood Estimates						
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq	
Intercept	1	-0.3054	0.4399	0.48	0.4875	NS
Group	2	-1.4040	0.5519	6.47	0.0110	<.05
	3	2.1795	0.7447	8.57	0.0034	<.05

Logistic Regression Model for P(PST) Score as a Function of Previous Education

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	9.02	0.0110	<.05
	Prev Educ	2	0.26	0.8764	NS
	Likelihood Ratio	0			

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	1.1168	0.5247	4.53	0.0333 <.05
	2	1.5020	0.5038	8.89	0.0029 <.05
Prev Educ	3	-0.2695	0.5247	0.26	0.6075 NS
	4	-0.2027	0.5038	0.16	0.6874 NS

Logistic Regression Model for LMS Previous Education as a Function of Group

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	1	0.02	0.8934	NS
	Group	2	10.37	0.0056	<.05
	Likelihood Ratio	0			

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-0.0608	0.4534	0.02	0.8934 NS
	2	-1.1594	0.5627	4.24	0.0394 <.05
Group	3	-1.2647	0.5916	4.57	0.0325 <.05

Logistic Regression Model for P(LMS) Score as a Function of Previous Education

Maximum Likelihood Analysis of Variance					
	Source	DF	Chi-Square	Pr>ChiSq	
	Intercept	2	1.95	0.3773	NS
	Prev Educ	2	2.93	0.2306	NS
	Likelihood Ratio	0			

Analysis of Maximum Likelihood Estimates					
Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	0.5123	0.3989	1.65	0.1991 NS
	2	0.1257	0.4261	0.09	0.7681 NS
Prev Educ	3	0.6664	0.3989	2.79	0.0948 NS
	4	0.2798	0.4261	0.43	0.5114 NS

BIBLIOGRAPHY

- Alderson, J. C. (1981). Report on the discussion on testing English for specific purposes. In J. C. Alderson & A. Hughes (Eds.), Issues in language testing. ELT Documents 111. Oxford: Pergamon Press.
- Alderson, J. C. (1988). Testing and administration in ESP. In D. Chamberlain & R. J. Baumgardner (Eds.), ESP in the classroom: Practice and evaluation. London: Modern English Publications.
- Alderson, J. C., & Urquhart, A. H. (1983). The effect of student background discipline on comprehension: A pilot study. In A. Hughes and D. Porter (Eds.), Current developments in language testing. London: Academic Press.
- Alderson, J. C., & Urquhart, A. H. (1984). Reading in a Foreign Language. London: Longman.
- Alderson, J. C., & Urquhart, A. H. (1985). The effects of students' academic discipline on their performance on ESP reading tests. Language Testing, 2, 192-204.
- Alderson, J. C. (2000). Assessing reading. Cambridge: Cambridge University Press.
- Barlett, F. C. (1932). Remembering. Cambridge: Cambridge University Press.
- Bernhardt, E. B. (1986). Reading in the foreign language. In B. H. Wing (Ed.), Listening, reading and application: Analysis and application. Middlebury, VT: Northeast Conference.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. Longman grammar of spoken and written English. Harlow, Essex: Pearson Education Limited, 1999.
- Brown, J. D. (1988). Components of engineering-English reading ability? System, 16, 193-200.
- Carrell, P. L. (1987). Readability in ESL. Reading in a Foreign Language, 4, 21-40.
- Carver, R. (1977). Toward a theory of reading comprehension and rauding. Reading Research Quarterly, 13, 8-63.
- Clapham, C. (1996). Studies in Language Testing. Cambridge: Cambridge University Press.
- Erickson, M. L. & Molly, J. (1983). ESP test development for engineering students. In J. W. Oller (Ed.), Issues in language testing research. Rowley, MS: Newbury House Publishers.

- Farhady, H. (1983). New directions for ESL proficiency testing. In J. W. Oller (Ed.), Issues in language testing research. Rowley, MS: Newbury House Publishers.
- Garson, G. D. Logistic regression. Statnotes. Retrieved February 12, 2001 from the World Wide Web: <http://www2.chass.ncsu.edu/garson/pa765/logistic.htm>
- Goodman, K. S. (1968). The psycholinguistic nature of the reading process. Detroit: Wayne State University Press.
- Gough, P. (1972). One second of reading. In J. F. Kavanaugh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge: MIT Press.
- Hale, G. A. (1988). Student major field and text content: interactive effects on reading comprehension in the Test of English as a Foreign Language. Language Testing, *5*, 49-61.
- Kamil, M. L. (1986). Reading in the native language. In B. H. Wing (Ed.), Listening, reading and application: Analysis and application. Middlebury, VT: Northeast Conference.
- Klein-Braley, C. (1983). A cloze is a cloze is a question. In J. W. Oller (Ed.), Issues in language testing research. Rowley, MS: Newbury House Publishers.
- Koh, M. Y. (1985). The role of prior knowledge in reading comprehension. Reading in a Foreign Language, *3*, 375-380.
- LaBerge, D. & Samuels, S. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, *6*, 293-323.
- Minsky, M. (1975). A framework for presenting knowledge. In P. Evinston (Ed.), The psychology of computer vision. New York, NY: Winston.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance. Hillsdale, NJ: Erlbaum.
- Schank, R. C. & Abelson, R. (1977). Scripts, plans and knowledge. In P. N. Johnson-Laird & P. C. Wason (Eds.), Thinking: Readings in cognitive science. Cambridge: Cambridge University Press.
- Shoham, M., Peretz, A. S., & Vorhaus, R. (1987). Reading comprehension tests: General or subject specific? System, *15*, 81-88.
- Smith (1971). Understanding reading. New York: Holt, Rinehart & Winston.

ACKNOWLEDGEMENTS

There are many, many people who have helped in tremendous ways to produce this piece of research.

First and foremost, I thank my dear parents, who have raised me and have provided me with love for all 23 years. Thank you for keeping in touch with me and worrying about me all the time I have been far away from you in the US. My thanks to my beloved grandmother for bringing me up and being the role woman of my life. Thank you to George for being my guiding star during my first years in the US. Thank you to my friends, and especially Aref, for understanding and helping me to cope with the stress connected with my first serious research, baring me and seeing the good in me.

Thank you to all the TESOL/AL faculty of ISU for the classes I took here. Special thanks to my major professor, Dan Douglas, for stirring my interest in the field of language assessment as well as in English for Specific Purposes, for shedding light on various questions I had and guiding me through the dark, as it often seemed, forest of research. Thank you to Susan Conrad and Cindy Myers for agreeing to be on my POS committee and helping me with the questions related to the thesis. Finally, but no less importantly, thanks to Mack Shelley, another POS committee member and my statistics instructor, for his patience in explaining statistical procedures, his help with providing examples, and his enlightening questions.

Thank you to all my classmates, especially Shannon Sauro, for keeping up my motivation to write and research by sharing their findings and their problems. Thank you to all of you! It is through valuable contributions of you that I have been able to finish the project and develop my research skills. Thank you.