ผลกระทบของความสามารถทางภาษาและภูมิหลังความรู้ทางวิศวกรรมศาสตร์ที่มีต่อ
ความสามารถในการอ่านภาษาอังกฤษเฉพาะกิจของนักศึกษาไทย
ระดับบัณฑิตศึกษา กลวิธีในการสอบและ
เจตคติต่อแบบทดสอบ

นางณัตจิรี จาตุรพิทักษ์กุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต

สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

THE EFFECTS OF LANGUAGE ABILITY AND ENGINEERING BACKGROUND KNOWLEDGE ON ESP READING ABILITY OF THAI GRADUATE STUDENTS, THEIR TEST TAKING STRATEGIES AND ATTITUDES TOWARDS THE TEST

Mrs. Natjiree Jaturapitakkul

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in English as an International Language
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2007

Thesis Title	THE EFFECTS OF LANGUAGE ABILITY AND ENGINEERING BACKGROUND KNOWLEDGE ON ESP READING ABILITY OF THAI GRADUATE STUDENTS, THEIR TEST TAKING STRATEGIES AND ATTITUDES TOWARDS THE TEST
By	Mrs. Natjiree Jaturapitakkul
Field of Study	English as an International Language
Thesis Advisor	Professor Kanchana Prapphal, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.... Dean of the Graduate School
(Assistant Professor M.R. Kalaya Tingsabadh, Ph.D.)

THESIS COMMITTEE

.... Chairperson
(Associate Professor Supanee Chinnawongs, Ph.D.)

.... Thesis Advisor
(Professor Kanchana Prapphal, Ph.D.)

.... Member
(Assistant Professor Kulaporn Hiranburana , Ph.D.)

.... Member
(Assistant Professor Pavinee Thirakhupt, Ph.D.)

.... External Member
(Assistant Professor Somsak Boonsathorn, Ph.D.)

ณัตจิรี จาตุรพิทักษ์กุล : ผลกระทบของความสามารถทางภาษาและภูมิหลังความรู้ทางวิศวกรรมศาสตร์ที่มีต่อความสามารถในการอ่านภาษาอังกฤษเฉพาะกิจของนักศึกษาไทย ระดับบัณฑิตศึกษา กลวิธีในการสอบและเจตคติต่อแบบทดสอบ. (THE EFFECTS OF LANGUAGE ABILITY AND ENGINEERING BACKGROUNDKNOWLEDGE ON ESP READING ABILITY OF THAI GRADUATE STUDENTS, THEIR TEST TAKING STRATEGIES AND ATTITUDES TOWARDS THE TEST) อ. ที่ปรึกษา: ศ.ดร.กาญจนา ปราบพาล, 173 หน้า.

การวิจัยนี้มีวัตถุประสงค์เพื่อ (1) ศึกษาผลกระทบร่วมของความสามารถทางภาษาและภูมิหลังความรู้ทางวิศวกรรมศาสตร์ที่มีต่อความสามารถในการอ่านภาษาอังกฤษเฉพาะกิจ (2) เปรียบเทียบผลการอ่านภาษาอังกฤษเฉพาะกิจระหว่างนักศึกษาที่มีความสามารถทางภาษาอังกฤษในระดับสูงและระดับต่ำ (3) ศึกษาผลกระทบของภูมิหลังความรู้ทางวิศวกรรมศาสตร์ที่มีต่อผลของการอ่านภาษาอังกฤษเฉพาะกิจ (4) เปรียบเทียบกลวิธีในการสอบของผู้เข้าสอบในแต่ละกลุ่มที่มีต่อประเภทคำถามในแบบต่างๆ และ (5) ศึกษาเจตคติของผู้เข้าสอบที่มีต่อแบบทดสอบ ประชากรในการวิจัยนี้ได้แก่ นักศึกษาชั้นปีที่ 1 ระดับบัณฑิตศึกษา จากคณะวิศวกรรมศาสตร์และไม่ใช่คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี (มจธ.) ปีการศึกษา 2550 กลุ่มตัวอย่างประกอบด้วยนักศึกษาจำนวน 120 คนจากกลุ่มประชากร ผู้วิจัยแบ่งนักศึกษาเหล่านี้ออกเป็น 2 กลุ่มตามความสามารถทางภาษาอังกฤษ คือระดับสูงและระดับต่ำ เครื่องมือที่ใช้ในการวิจัยประกอบด้วย (1) แบบทดสอบการอ่านภาษาอังกฤษสำหรับสาขาวิศวกรรมศาสตร์ (2) แบบทดสอบภาษาอังกฤษเพื่อจัดระดับความสามารถทางภาษาของ มจธ. (3) การสัมภาษณ์ผู้เข้าสอบ และ (4) แบบสอบถามผู้เข้าสอบ ผู้วิจัยใช้การวิเคราะห์ความแปรปรวนแบบสองทาง (Two-way ANOVA) ในการวิเคราะห์ผลกระทบของตัวแปรอิสระทั้ง 2 คือความสามารถทางภาษาและภูมิหลังความรู้ทางวิศวกรรมศาสตร์ และใช้ partial Eta squared ในการวัดขนาดของผลกระทบทั้งสองตัวแปร ผลจากการสัมภาษณ์ผู้วิจัยวิเคราะห์ข้อมูลโดยการจัดกลุ่มข้อมูล และแสดงข้อมูลในรูปแบบของการแจกแจงความถี่ ในการประเมินเจตคติของผู้เข้าสอบผู้วิจัยใช้ค่าเฉลี่ยรวมของมาตรวัดเจตคติในการวิเคราะห์

ผลการศึกษาพบว่า (1) ความสามารถทางภาษาและภูมิหลังความรู้ทางวิศวกรรมศาสตร์ไม่มีผลกระทบร่วม (interaction effect) ต่อความสามารถในการอ่านภาษาอังกฤษเฉพาะกิจ (2) นักศึกษาที่มีความสามารถทางภาษาสูงและนักศึกษาที่มีความสามารถทางภาษาต่ำมีความแตกต่างอย่างมีนัยสำคัญทางสถิติ ในการทำแบบทดสอบภาษาอังกฤษเฉพาะกิจ (3) นักศึกษาที่มีภูมิหลังความรู้ทางวิศวกรรมศาสตร์และไม่มีภูมิหลังความรู้ทางวิศวกรรมศาสตร์มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ (4) โดยภาพรวมผู้เข้าสอบทั้งหมดใช้กลวิธีในการตอบคำถามประเภทต่างๆเหมือนกัน อย่างไรก็ตามผู้เข้าสอบที่มีความสามารถทางภาษาสูงใช้กลวิธีในการสอบมากกว่า และบ่อยกว่าผู้เข้าสอบที่มีความสามารถทางภาษาต่ำ และ (5) ผู้เข้าสอบมีเจตคติที่ดีต่อแบบทดสอบที่พัฒนาขึ้น ผลการวิจัยให้ความกระจ่างมากขึ้นในเรื่องการทดสอบความสามารถการอ่านภาษาอังกฤษเฉพาะกิจของนักศึกษาไทยในระดับบัณฑิตศึกษา นอกจากนี้ยังให้ข้อมูลผู้สอนภาษาอังกฤษเฉพาะกิจในการพัฒนาการเรียนการสอน และช่วยผู้สร้างแบบทดสอบพัฒนาแบบทดสอบให้มีประสิทธิภาพมากยิ่งขึ้น

สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ ลายมือชื่อนิสิต ....

ปีการศึกษา 2550 ลายมือชื่ออาจารย์ที่ปรึกษา ....

4789657920: MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
KEY WORD: LANGUAGE ABILITY/ ENGINEERING BACKGROUND KNOWLEDGE/ ENGLISH FOR SPECIFIC PURPOSES (ESP)/ ESP READING ABILITY

NATJIREE JATURAPITAKKUL: THE EFFECTS OF LANGUAGE ABILITY AND ENGINEERING BACKGROUND KNOWLEDGE ON ESP READING ABILITY OF THAI GRADUATE STUDENTS, THEIR TEST TAKING STRATEGIES AND ATTITUDES TOWARDS THE TEST. THESIS ADVISOR: PROF. KANCHANA PRAPPHAL, Ph.D., 173 pp.

The objectives of this study were (1) to investigate the interaction effect between language ability, and engineering background knowledge on ESP reading ability, (2) to compare the ESP reading performance between high and low language ability groups, (3) to explore the effect of background knowledge on ESP reading performance, (4) to compare test taking strategies used among test takers in answering different kinds of test items and (5) to study test takers' attitudes towards the test. The population was first-year graduate students from Engineering and non-Engineering faculties at King Mongkut's University of Technology Thonburi of 2007 academic year. The samples consisted of 120 students from the population. They were assigned to two groups of high and low language ability. The instruments included the developed Engineering-English Reading Test (E-ERT), the KMUTT English placement test, an introspective interview and a test takers' questionnaire. Two-way ANOVA was carried out to observe the effects of the two independent variables: language ability and engineering background knowledge. In addition, partial Eta squared was used to measure the effect size of the treatment. Frequency counts were employed to investigate test taking strategies. Descriptive statistics using mean score was carried out to reveal the test takers' attitudes towards the test.

The findings revealed that (1) there was no significant interaction effect between language ability and engineering background knowledge on ESP reading ability, (2) there was a significant difference between students with high language ability and low language ability, (3) there was a significant difference between students with engineering background knowledge and those without engineering background knowledge, (4) in general all the test takers employed the same test taking strategies in answering each kind of test item, but high language ability test takers used more strategies and used them more frequently than low language ability ones and (5) students had positive attitudes towards the developed test. The findings provide more insights into ESP reading assessment of graduate students in Thailand. In addition, English teachers as well as test writers can use the information to facilitate their ESP instructions and to develop more effective ESP reading tests for their institutions.

Field of Study: English as an International Language Student's signature: Natjiree J.......

Academic Year: 2007 Advisor's signature: Kanchana Prapphal

ACKNOWLEDGEMENTS

I am indebted to many people who provided me with guidance, support and encouragement to produce this dissertation. My deepest appreciation and gratitude is owed to Professor Dr. Kanchana Prapphal, my generous and open-minded advisor, for her valuable advice and deeply compassionate encouragement throughout the time of my doctoral study. I am grateful to the entire dissertation committees, Associate Professor Dr. Supanee Chinnawongs, Assistant Professor Dr. Kulaporn Hiranburana, Assistant Professor Dr. Pavinee Thirakhupt and Assistant Professor Dr. Somsak Boonsathorn, who devote their time to read my work, and invariably give valuable suggestions. I am highly thankful to Associate Professor Dr. Suphat Sukamolson, who gives insightful statistic lessons and valuable suggestions on statistics used in this research, and allows me to make use of his developed Classical Test Item Analysis Program for my data analysis.

This dissertation was made possible by the generous help of the subject experts in the Engineering field. I am highly indebted to their earnest support and cooperation. My special gratitude belongs to Associate Professor Dr. Richard Watson Todd, Assistant Professor Dr. Saowaluck Tepsuriwong, Dr. Kitcha Thepsiri, Dr. Pattamawan Jimarkon and Lecturer Nantana Pichaipattanasopon who spent their precious time in the instrument validation process. I would like to thank Lecturer Suvalee Chinkumtornwong, the Head of Language Studies Department, School of Liberal Arts, King Mongkut's University of Technology Thonburi, as well as the other lecturers for granting me the permission to conduct this study. I also would like to extend my special thanks to all graduate students from the Faculties of Engineering and non-Engineering who participated in the study. I particularly acknowledge for the financial support from the Commission on Higher Education, Ministry of Education, under Cooperative Research Network (CRN) Program.

I am immensely grateful to all instructors and all friends at the EIL program. My special appreciation goes to my family members for their support. Above all, I am greatly indebted to my husband, Professor Dr. Chai Jaturapitakkul for his love, understanding, support, valuable advice and everything he has assisted me throughout the time of my doctoral study. Last but not least, my special gratitude belongs to my beloved mother, Mrs. Suwannee Vajanaratana, my greatest source of inspiration.

CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Background of the Study

English for Specific Purposes (ESP) has received considerable attention as an approach for teaching English as a Second Language (ESL) and English as a Foreign Language (EFL). Seen as a highly specialized field in English language teaching, ESP courses are designed to meet learners' communication needs, whether they are discipline-specific, as in the case of English for Academic Purposes (EAP) or job-specific, as in the case of English for Occupational Purposes (EOP).

At present, the influence of ESP is growing in the areas of curriculum design and classroom materials. However, despite enthusiastic interest in this approach, little has been reported in the area of language evaluation (Erickson and Molloy, 1983). Thus there is a need for testing methodologies that will complement the advances in ESP as an approach to instruction.

In Thailand, ESP courses have been offered to learners for both academic and occupational purposes, particularly at the tertiary level. Various institutions offer ESP courses with different language content and materials, since all ESP courses and materials are customized for course participants, based on an analysis of learner needs and uses for communicating in the English language. Nevertheless, despite the myriad ESP courses, the reliability and validity of proficiency tests for evaluating language proficiency in ESP courses is debatable, particularly as regards the form of language tests. In fact, little evidence of the reliability and validity of ESP tests can be found, probably due to the focus on instruction rather than evaluation. Clearly this pedagogical imbalance in ESP should be addressed in order to complement the advance of ESP instruction.

Over the decades, the belief that language testing based on content already familiar to students is more appropriate than general English language proficiency tests (Clapham, 2000). Two reasons support this belief. Firstly, such tests would assess the linguistic abilities that the students would actually use during their studies; secondly, the students would be best able to use their second or foreign language in a familiar field, both of which would not be possible in such tests as general language proficiency tests. Furthermore, language experts such as Douglas (2000) argue that the last decade has seen the proliferation of field-specific language contexts, so that a field-specific language test is a better predictor of performance than a general purpose test. Thus in response to the need for ESP tests, it seems appropriate to pursue the development of discipline-specific language tests.

At King Mongkut's University of Technology Thonburi (KMUTT), language syllabuses are developed mainly for academic programs in Engineering, Science and multi-disciplinary technologies. At all levels in KMUTT, some form of English language course offered by the School of Liberal Arts is compulsory for students enrolled in Engineering, Science and Technology courses. However, regardless of the type of course, language ability is assessed by general English language proficiency tests. One reason for this is that it is quite difficult to provide specific English lessons and ESP tests for students from a wide range of disciplines. Another reason is the academic background of teaching staff: the majority do not have backgrounds in Engineering, Science and Technology and are thus limited in their ability to teach unfamiliar field-specific content. Not surprisingly, the learners' needs are often neglected in the syllabus and assessment is accordingly unrelated either in theme or topic to their fields of specialization. Thus learning retention rates are poor upon completion of compulsory general English language courses due to lack of exposure or opportunity to practise English, and more importantly, the irrelevance of the content they have learned to their disciplines.

In interviews conducted with 10 engineering experts from different disciplines and universities in Thailand during June-July, 2006 (see Appendices A and B), it was found that all the informants considered English language ability very important for engineering graduate students. At the time, all expressed dissatisfaction with their students' English language ability and believed that higher levels of proficiency should have been achieved upon completion of compulsory English language courses. In addition, most agreed that the English test for engineering students should be further developed to enable self assessment of actual

English language ability. Another recommendation made by the interviewees was that ESP courses should be offered whenever possible after completion of compulsory English courses, in order to reinforce the English language skills of graduate students in engineering or any particular field with guidance from field experts.

From the review of relevant literature and responses from field experts regarding needs for ESP testing, this study develops and presents an alternative for the general English language testing issue in the form of an ESP reading test for engineering students which incorporates stakeholders' interests.

While various ESP tests have been developed for engineering students, these have been customized to different learning needs for different contexts. At King Mongkut's University of Technology Thonburi (KMUTT), there are no tailored tests to meet the need for assessing English language proficiency in the engineering field. Graduate students are short of opportunities to self-assess in preparation for continuing graduate study. This is a common situation in Southeast Asia including Thailand, where the medium of instruction is the L1, or native language (Thai, in this case) while textbooks are in English. Thus adequate levels of English language proficiency could be one of the key concerns in graduate study. Therefore, the test could serve as a tool for self-assessment of English proficiency and readiness to continue higher education in engineering.

In this study, the subjects were graduate-level engineering students (from different disciplines). The majority were graduate students at King Mongkut's University of Technology Thonburi (KMUTT). The study was undertaken at KMUTT because it is a leading university renowned in Thailand for engineering education. It is hoped that by concentrating the study on one field, the familiarity with content can be better controlled as suggested by Shoham, Perets, and Vorhaus (1987). However, non-engineering graduate students were also included in this study to ensure construct validity of the test and to test one hypothesis of the study on background knowledge.It was expected that if the test had construct validity, it could distinguish engineering from non-engineering students. In addition, few studies in ESP concerned particularly with graduate students have been reported.

The results of the present study, it is hoped, will yield some interesting insights into the ESP reading ability of graduate engineering students in Thailand.

This study focused on the reading skill, since it is considered a basic skill and one most frequently used for academic study in most disciplines, particularly for second or foreign language learners (Lynch & Hudson, 1991). This is also true in Thailand, where one of the major assessment concerns is to effectively evaluate reading comprehension skills as it is considered a core skill in many areas, particularly education (Wongsothorn, 1996). Interviews conducted for this study during June-July, 2006, also revealed that engineering professors considered reading to be the most necessary language skill for graduate-level engineering instruction and learning. While most studies have been conducted on the testing of reading for general English language proficiency, there have been relatively fewer ESP reading research studies such as those by Alderson and Urquhart (1983, 1985a, 1985b), Clapham (1996), and Tan (1990). Thus a study of ESP reading comprehension skills would be beneficial for providing further insight.

When assessing reading ability, an important consideration is the variables which affect the nature of reading. Research has looked at the way readers themselves affect the reading process and product, and has investigated a number of other variables, such as background knowledge, knowledge of language, metalinguistic knowledge and metacognition, reading and test taking strategies and test takers' characteristics (Alderson, 2000). As regards ESP reading, there are a number of variables that play important roles in ESP reading performance. Several studies have found significant interaction between background knowledge and language ability (Douglas, 2000). In English for Academic Purposes, background knowledge is sometimes seen as a confounding variable to be minimized at best (Jordan, 1997). But for English for Specific Purposes, background knowledge is a necessary and integral part of the concept of specific purpose language ability. However, studies conducted by Tan (1990) and Clapham (1996) reveal that comprehension of subject-specific reading texts could be affected by language proficiency rather than background knowledge. English language ability is, therefore, found to be another key variable relating to success in the test performance.

In this study, the focus is on two interacting variables which distinguish ESP testing from general purpose testing: language ability and background knowledge (Douglas, 2000). These two variables are of interest in this research firstly for their important influence on both language use and test performance and secondly, because they may be manipulated in test design to facilitate rather than impede test takers' performance. Added to this, is the fact that there is a lack of conclusive evidence from studies concerning their impact on reading performance (Koh, 1985; Shoham et al., 1987; Tan, 1990; Jensen and Hansen, 1995), thus inviting further investigation of background and reading ability.

There is a small but growing body of knowledge on how test takers go through the process of taking language tests – the steps that they take to arrive at answers to questions. Cohen (1984: 70) mentioned that the purpose of such research has been to explore the closeness-of-fit between the tester's presumptions about what is being tested and the actual processes that the test taker goes through. The findings obtained from such research have revealed both the weaknesses in tests as well as successful and unsuccessful test taking strategies. In addition, in language assessment, there has been a growing concern for the reliability and validity of language tests. It is a relatively new undertaking to use data on test taking strategies in order to validate such tests (Cohen, 1998: 92). This study focuses effort in this direction. Test takers' strategies were investigated to identify successful strategies and validate the test purpose so that test constructors could verify the skills which test items were actually testing, thus aiding the development of tests.

A recent concern among researchers in the field of language testing has been the identification of the test takers' characteristics that influence performance on tests of English as a Foreign Language (EFL) or English as a Second Language (ESL). These test takers' characteristics include personal attributes such as age, native language, gender and educational characteristics including background knowledge, previous exposure to English as well as cognitive, psychological and social characteristics, such as learning strategies and styles, attitude and motivation, aptitude and intelligence, field dependence and independence, extroversion and introversion and anxiety, personality and risk taking (Kunnan, 1995: 1). According to the literature review in Chapter II, there are a number of variables that play important roles in language tests and affect test takers' performance. Several studies have shown the importance of attitudes as they can directly affect the test scores. Positive attitudes of test takers towards the tests were found to bring about better results and help improve performance. An investigation on test takers' attitudes towards the test is one of the aims of this study. This can provide information on how a specific purpose test tailored to meet field-related language needs is received by the test takers.

In conclusion, this study involved development of the Engineering-English Reading Test (E-ERT) as a tool for measuring students' reading ability and for eliciting the background knowledge used by test takers. In addition, test taking strategies of the test takers were investigated for their efficacy in answering different types of questions and for test validation purposes. Lastly, students' attitudes towards the test were explored to find out how the test was received by the students.

1.2 Objectives of the study

This study attempts to investigate the effects of language ability and background knowledge on ESP reading ability and the reading performance of test takers in the Engineering-English Reading Test (E-ERT). In addition, the study aims to compare test taking strategies' used by the test takers in answering different types of test items. It is anticipated that this investigation would provide qualitative information for test validation. Furthermore, the study aims to explore the test takers' attitudes towards the E-ERT. Their attitudes could project how the E-ERT would be received by the test takers who are the key stakeholders. This study, therefore, aims to:

1. investigate the interaction effects between language ability and engineering background knowledge on ESP reading ability,
2. compare the ESP reading performance between high and low language ability groups,
3. explore the effect of background knowledge on ESP reading performance,

4. compare the test taking strategies used by the four subgroups in answering different types of test items, and
5. study the test takers' attitudes towards the E-ERT.

These objectives lead to the following research questions.

1.3 Research questions

1. Is there any significant interaction effect between language ability and background knowledge on the ESP reading ability? If there is, what is its effect size?
2. Is there any significant difference between students with high language ability and those with low language ability in doing the E-ERT? If there is, what is its effect size?
3. Does background knowledge affect ESP reading performance? If it does, what is its effect size?
4. How do the test takers in the four subgroups use test taking strategies in order to answer different types of test items?
5. What are test takers' attitudes towards the E-ERT?

1.4 Statement of hypotheses

Hypothesis 1: There is a significant interaction effect between language ability and engineering background knowledge on the ESP reading ability at the .05 level.

($H_{1.1:}$ μ High w/ Eng $\neq$ μ High w/o Eng)

($H_{1.2}$: μ Low w/ Eng $\neq$ μ Low w/o Eng)

($H_{1.3}$: μ High w/ Eng $\neq$ μ Low w/ Eng)

($H_{1.4}$: μ High w/o Eng $\neq$ μ Low w/o Eng)

Hypothesis 2: Students with high language ability perform significantly better in the E-ERT than those with low language ability at the .05 level.

(H_2: μ w/ High > μ w/ Low)

Hypothesis 3: Students with engineering background knowledge can do the E-ERT better than those without background knowledge at the .05 level.
(H_3: μ w/ Eng > μ w/o Eng)

Hypothesis 4: The test takers have positive attitudes towards the E-ERT.
(H_4: Mean of attitude scale > 2.5 points from the 4-point scale in the questionnaire)

For the above hypotheses, directional hypotheses were employed because previous research studies indicate that the independent variables may have an effect on the dependent variable. As Bachman (2004) points out, in the situations where prior experience, theory or previous research suggest a direction in relationship, a stronger or directional research hypothesis could be formulated instead of the null hypothesis.

1.5 Scope of the study

1. The population and sample in this study are first year graduate students in engineering, and non-engineering students from Master's degree programs at King Mongkut's University of Technology Thonburi (KMUTT). Engineering students are from different disciplines such as Civil, Mechanical, Chemical, Computer, Industrial and Electrical engineering. Non-engineering students, or those without engineering background knowledge, include those from the School of Liberal Arts, Faculty of Industrial Education and Technology, Faculty of Bio-resources and Technology, Faculty of Science and Faculty of Energy Environment and Materials. Students from other faculties are excluded from this study.
2. The variables in the study are categorized into independent variables and a dependent variable. The former includes language ability, and engineering background knowledge. The latter is the Engineering-English Reading Test (E-ERT) scores.
3. This study focuses on the reading skill in English for Engineering. The study of other skills - speaking, listening and writing, are not included in this study.

4. The data on test taking strategies is collected by means of introspective interviews. A composite list of test taking strategies used in a multiple choice reading comprehension test is employed as a guideline in interviews.
5. An investigation of test takers' attitudes towards the E-ERT is conducted to obtain an overall, rather than a detailed picture.

1.6 Assumptions of the study

1. This study assumes that the test takers give valid data in the interviews and the test takers' questionnaire, and work to the best of their ability in taking the E-ERT. This is because before the test is administered, the researcher explains the development, the significance, and the use of the test as well as the opportunity for self-assessment to test takers. The procedures would help in making the test takers understand the purposes of the test.
2. It is assumed that test takers in different classes do not know each other because they are studying in different disciplines. Therefore, the test would be secure enough to administer several times.
3. It is assumed that gender would not have an effect on the E-ERT scores.

1.7 Definitions of terms

1. <u>Placement test</u>

The placement test refers to a validated in-house English placement test for graduate students, currently used at King Mongkut's University of Technology Thonburi (KMUTT). It consists of two main parts: grammar and reading comprehension. Each part contains fifty multiple-choice items. In this study, this test is used for placing students into high and low language ability groups.

2. <u>Language ability</u>

In this study, "language ability" refers to the ability to perform the test tasks by using language knowledge, including knowledge of vocabulary, morphology and syntax, cohesion, and registers in order to comprehend the reading texts. The test takers are categorized into high and low language ability groups based on the placement test scores.

3. High language ability group

"High language ability group" refers to a group of high language ability students whose placement test scores are at or above +1 S.D.

4. Low language ability group

"Low language ability group" refers to a group of low language ability students whose placement test scores are at or below -1 S.D.

5. Engineering background knowledge

In this study, "engineering background knowledge" refers to content knowledge, and in particular subject-specific content knowledge – the knowledge acquired mainly from university courses in the academic discipline of engineering.

6. Group with engineering background knowledge

A group with engineering background knowledge refers to graduates with a Bachelor of Engineering degree, possessing sound background knowledge of engineering.

7. Group without engineering background knowledge

A group without engineering background knowledge refers to graduates with a Bachelors degree from a non-Engineering discipline such as Arts, Education, Science or Business Administration.

8. English for Specific Purposes (ESP)

English for Specific Purposes or ESP is operationally defined as English language used especially for a purpose of language testing which is concerned with stakeholders' needs and relates in content to a particular discipline. In this study, the focus is on English for Academic Purposes (EAP), which is a branch of ESP. The ESP test is designed for engineering students at the tertiary level. In this study ESP refers to a semi-level of text specificity particularly in engineering texts.

9. Semi-level of text specificity

Semi-level of text specificity refers to a medium degree of specificity of the texts used in ESP reading. The following criteria are considered for a semi specific text in this study.

9.1 There is a core of field specific vocabulary but the vocabulary is explained in the text.

9.2 The words mostly used in the text are non-technical and sub-technical vocabulary for engineering students. The term “sub-technical” refers to the kind of lexical items with a technical as well as non-technical sense. It covers a whole range of items that are neither highly technical and specific to a certain field of knowledge nor obviously general in the sense of being everyday words which are not used in a distinctive way in specialized texts. Examples of such words are “iron, force, stress, current, tension, strength”, which have the same meaning in several disciplines.

9.3 Comprehension of the text requires knowledge of subject specific concepts, some of which are explained in the text.

9.4 The texts are not highly academic texts with a great deal of technical terminology as often seen in classrooms. They are made comprehensible by explanation in the texts.

10. <u>ESP reading ability</u>

In this study, ESP reading ability refers to scores from the Engineering-English Reading Test (E-ERT) which contains vocabulary and reading comprehension questions. All reading tasks are based on general engineering topics.

11. <u>Test taking strategies</u>

Test taking strategies refer to test taking processes that test takers have selected and of which they are conscious of, in order to answer different types of questions (vocabulary, main idea, details, inference, and engineering fact) in the E-ERT.

12. <u>Attitudes towards the test</u>

Attitudes refer to beliefs of test takers that are either positive or negative towards the E-ERT. Rating scales (from 1 – 4) are used to indicate positiveness or negativeness.

1.8 Significance of the study

The following benefits from the study are anticipated, if predicted results are obtained:

1. In terms of practical contribution, the following parties could gain the following benefits.

1.1 For graduate students, the test could be useful as an instrument for self-assessment of their actual language proficiency, particularly in an engineering discipline.

1.2 For teachers, since the test was based on the needs and suggestions of stakeholders in an engineering field, ESP instruction could be improved through the use of the target language in a specific field, an opportunity not previously available at KMUTT. The test could probably create positive washback for the design of new ESP/EAP courses which would hopefully reinforce the use of English in any academic fields.

1.3 For test developers, test design methods could be broadened. The test was developed with the more meaningful constructs of an ESP reading test, reflecting the nature of target language use in a specific field. By increasing awareness of language ability levels and background knowledge as test taker variables, these two factors could become a consideration in the design and construction of ESP tests.

2. The following theoretical contributions could be made.

2.1 This research study could reveal the interaction effects of language ability and background knowledge on graduate students' performance on ESP reading tests, particularly in the engineering field in the Thai context. This could be a resource for educational practitioners to develop more effective ESP reading tests.

2.2 In the area of language assessment, this study could contribute to the predictive validity of the test takers' reading ability. This is because the test scores obtained from the test may predict some behaviors of test takers such as the potential for success in a chosen field of study or the need to improve reading ability in English.

2.3 The obtained findings would also help to prove those inconclusive studies on the effects of language ability and background knowledge on the reading performance of students.

3. From a comparison of test taking strategies of the test takers in answering different types of test items, the following information could be obtained: 1) the researcher could check for any differences in test taking strategies of the test

takers in the four subgroups in responding to each type of test item, and 2) the data on test taking strategies could be used to validate the E-ERT.

4. From an investigation of test takers' attitudes towards the test, information is provided on how a specific purpose test, tailored to meet the needs in a chosen field is received by test takers, who are key stakeholders. If positive attitudes are revealed, similar procedures in developing and administering tests could be replicated for future research.

Overview

Chapter one provides the background of the study. It includes the objectives of the study, research questions, hypotheses, definition of terms and the significance of the study.

Chapter two presents a review of related literature in ten key areas which are 1) introduction to ESP testing, 2) ESP reading theories, 3) construct of reading, 4) background knowledge and language proficiency, 5) research in ESP reading, 6) the UCLA engineering English reading test, 7) item types by skill/ sub-skills involved, 8) test taking strategies, and 9) attitudes towards the tests.

Chapter three focuses on research methodology. The population and sample data are presented. The procedures employed in constructing the research instruments as well as the validation process are provided. Finally, data collection and data analysis are included in this chapter.

Chapter four reveals the findings of the study, which are presented according to the research questions. A discussion of each research question is presented, based on the literature review and theoretical background.

Chapter five provides a summary of the research and conclusions from the findings. The implications from the study as well as recommendations for future research are also included.

CHAPTER II

LITERATURE REVIEW

This chapter presents a review of the related literature from which the underlying concepts of this study were drawn. It covers an introduction to ESP testing, a review of ESP reading theories, construct of reading, background knowledge and language proficiency, research in ESP reading, the UCLA engineering English reading test, item types as skill/ sub-skills involved, test taking strategies, and attitudes towards the tests.

2.1 Introduction to ESP Testing

English for Specific Purposes (ESP) became an important part of English as a Second Language teaching in the 1970s and 1980s. It is a direct result of the introduction of the communicative framework (Munby, 1978). Theoretically, ESP consists of two branches: EAP (English for Academic Purposes) and EOP (English for Occupational Purposes). In other words, each branch is sub-divided according to the discipline or occupation it is concerned with. For example, EAP may be separated into English for Biology, English for Mathematics, etc. and EOP branches out into English for Pilots, English for Doctors, etc. However, in practice, Johns and Dudley-Evans (1991: 306) have noted that "For most of its history, ESP has been dominated by English for Academic Purposes…" This domination is due to EAP practitioners working in academic institutions, where research and intellectual enquiry are encouraged. In the same tradition, this study is apparently dominated by EAP as well.

Regarding ESP testing, Douglas (2000: 19) proposed a precise definition that an "LSP/ESP test is one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge….." This definition emerged from a number of concepts that form the background of LSP/ESP testing. Some of the definitions here are key features which distinguish ESP from general purpose testing.

According to Douglas (2000:2), there are two features that typically distinguish English for Specific Purposes (ESP) testing from general purpose testing. The first distinguishing feature is the *authenticity of the tasks*. This means the test tasks should share key features with the tasks that a test taker might carry out in the target language use situation. The second feature is the *interaction between language knowledge and specific content or background knowledge* which will be discussed in detail in later sections. This is perhaps the clearest defining feature of ESP testing. For general purpose testing, background knowledge is considered a confounding variable that contributes construct-irrelevant variance to the test score. On the other hand, in ESP testing, background knowledge is necessary as an integral part of the concept of specific purpose language ability.

In brief, ESP derives from the communicative framework and consists of two branches: EAP and EOP. Authenticity of tasks and interaction between language knowledge and specific content or background knowledge are key features which distinguish ESP from general purpose testing. Before the development of reading tests, literatures on the models of reading are explored in the next section.

2.2 ESP Reading Theories

The process of reading has been intensively studied, and has interested researchers in many disciplines such as anthropology, philosophy, psychology, education and linguistics. Not surprisingly, therefore, there have been many different approaches/ models. In this study, the following major models: bottom-up, top-down and interactive models, and schema theory are applied.

2.2.1 Models of Reading Comprehension

To focus on ESP reading, the literature on the models of reading process will be explored. The three primary approaches to comprehension theory which can be applied to ESP reading are as follows.

Bottom-up model: This approach is a serial process of constructing meaning by building smaller units into larger ones. Thus reading begins with the printed word: it is the recognition of graphic stimuli, combined and decoded to sound groups and words; followed by the recognition of words which are decoded to word groups representing phrases, etc. This approach was typically associated with behaviorism and with phonics approaches to the teaching of reading which argue that children need to learn to recognize letters before they can read words, and so on. In this traditional view, readers are passive decoders of sequential graphic – phonemic – syntactic – semantic systems, in that order (Alderson, 2000: 16).

Top-down model: This approach emphasizes the importance of the schemata, and the reader's contribution, to the text input. Goodman (1982 cited in Alderson, 2000: 16), for example, calls reading a psycholinguistic guessing game, in which readers guess or predict the text's meaning on the basis of minimal textual information, and maximum use of existing, activated knowledge. Smith (1971 cited in Alderson, 2000: 16) claims that non-visual information transcends the text, and includes the reader's experience with the reading process, knowledge of the context of the text, familiarity with the structures and patterns of the language and of specific text types, as well as generalized knowledge of the world and specific subject matter knowledge.

Interactive model: According to Alderson (2000: 18), the reading comprehension process can be seen as an interactive model, in which every component in the reading process can interact with any other component, be it higher up or lower down. Processing, in fact, is now thought to be parallel rather than serial (Grabe, 1991: 384). The model developed by Rumelhart (1977 cited in Alderson, 2000: 18), for example, incorporates feedback mechanisms that allow knowledge sources (linguistic as well as world knowledge) to interact with visual input. In his model, a final hypothesis about the text is synthesized from multiple knowledge sources interacting continuously and simultaneously. Stanovich (1980), on the other hand, has developed an interactive compensatory model in which the degree of interaction among components depends upon knowledge deficits in individual components, where interaction occurs to compensate for deficits. Thus, readers with poor word recognition skills may use top-down knowledge to compensate.

In summary, the reading process can be described by three models: bottom-up, top-down, and interactive models. In bottom-up, reading is a linear process progressing from graphic symbols to meaning responses. In top-down and interactive models of reading, it is presumed that humans depend on memory or previous knowledge of some kind when they interpret written cues. Without previous knowledge they would not be able to formulate any inferences or hypotheses about what was coming next. Fundamental to all these models, therefore, must be some system of storing and retrieving past knowledge. The group of theories which attempt to account for this come under the general umbrella term of “schema theory” which will be elaborated in the following section.

2.2.2 Schema Theory

According to **Schema Theory,** Clapham (2000) explained that knowledge is stored not in lists, but in hierarchies. Within these hierarchies are schemata which are embedded in other schemata, and which themselves contain sub-schemata. These schemata vary in their levels of abstraction, and represent all types of knowledge, such as objects, academic topics, rules, events, routines and social situations. They represent knowledge, rather than definitions, 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 new ones can be developed by the process of accommodation; that is, the modification of previous schemata in the light of new information.

In addition, Alderson (2000) stated that the development of schema theory has attempted to account for the consistent finding that what readers know affects what they understand. Schemata are seen as interlocking mental structures representing reader’s knowledge. When readers process text, they integrate the new information from the text into their pre-existing schemata. More than that, their schemata influence how they recognize information as well as how they store it.

According to Carrell (1987), types of knowledge or schemata may be distinguished into formal schemata and content schemata. By the former, is meant knowledge of language and linguistic conventions, including knowledge of how texts are organized, and what the main features of particular genres are. By the latter is meant, essentially, knowledge of the world, including the subject matter of the text.

In addition, content schemata also include background knowledge and subject-matter knowledge. Background knowledge may or may not be relevant to the content of a particular text. However when knowledge is directly relevant to text content and topic, it is subject matter knowledge.

Research on the usefulness of the notion of schema theory for second language reading (Grabe, 1991: 390) revealed that activating content information plays a major role in comprehension and recall of information from a test. Carrell (1988 cited in Grabe, 1991: 390) has also argued that lack of schema activation is a major source of processing difficulty with second language readers. This has been verified not only through culture-specific text comparisons but also in discipline-specific comparisons of readers with familiar and less familiar background knowledge (Alderson & Urquhart, 1983). In addition, other studies on schema theory have argued that a high degree of background knowledge can overcome linguistic deficiencies (e.g. Hudson, 1982). The major implication to be drawn from this research is that students need to activate prior knowledge of a topic before they begin to read. For students who do not have sufficient prior knowledge, they should be given at least minimal background knowledge from which to interpret the text.

In conclusion, the concept of schema theory can help us understand that the knowledge of readers affects what they understand. The reader brings a set of schemata to bear on a reading process. These relate to the lexical system, the syntactic system and the semantic system. Generally, schemata can be adjusted to accommodate new information, but if the reader's schemata are inadequate because of a lack of appropriate background knowledge, then, comprehension breaks down.

2.3 Construct of ESP Reading

By reviewing the literature on the theory of reading (Alderson, 2000; Chalhoub-Deville, 1999 and Clapham, 1996), components of specific purpose language ability and communicative language ability (Douglas, 2000; and North & Schneider, 1998 based on Bachman and Palmer 1996 framework), and considering the purpose of this research study, two key viewpoints on construct of reading in an ESP context are presented.

1. Language ability
2. Background knowledge

Language ability is the model proposed by Bachman (1990), who defines it as involving two components: language knowledge and strategic competence or metacognitive strategies. This combination of language knowledge and strategic competence provides language users with the ability to create and interpret discourse, either in responding to tasks on language tests or in non-test language use.

Language knowledge can be thought of as a domain of information in the memory that is available for use by the metacognitive strategies in creating and interpreting discourse in language use (Bachman and Palmer, 1996). Language knowledge in this study includes grammatical knowledge (referred as linguistic competence for North & Schneider, 1998), textual knowledge and sociolinguistic knowledge.

Grammatical knowledge involves producing or comprehending formally accurate utterances or sentences. This includes knowledge of vocabulary, morphology and syntax. To improve reading ability, test takers should apply grammatical knowledge as described by North and Schneider (1998). That is, test takers understand texts which contain complex and unfamiliar language, have a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms, and understand grammatical patterns and vocabulary ordinarily encountered in academic/ professional reading.

Textual knowledge includes the knowledge of the conventions for joining utterances to form a text, which includes knowledge of cohesion and rhetorical organization. Cohesion comprises ways of explicitly marking semantic relationships such as reference, sub-situation, ellipsis, conjunction, and lexical cohesion (Halliday and Hasan, 1976) as well as conventions such as those governing the ordering of old and new information in discourse. Rhetorical organization pertains to the overall conceptual structure of a text and is related to the effect of the text on the language user. Conventions of rhetorical organization include common methods of development such as narration, description, comparison, classification, and process analysis (McCrimman, 1984 cited in Bachman, 1990: 88). To apply textual knowledge on reading, test takers should be able to understand the grammatical or lexical relationship between different sentences or between parts of a sentence.

Sociolinguistic knowledge enables us to create or interpret language that is appropriate to a particular language use setting. This includes knowledge of the conventions that determine the appropriate use of dialects or varieties, register, idiomatic expression, and cultural references. To make use of sociolinguistic knowledge in reading tests, test takers need to understand many socio-linguistic and cultural references.

Another component of language ability is **strategic competence**. Strategic competence serves as a link between the external context, of the specific purpose language use situation, and the internal knowledge that forms the wherewithal for communication (Douglas, 2000: 33). It plays a central role in communicative performance in assessing the situation, setting goals with respect to the situation, planning the response by deciding what elements of knowledge – both background knowledge and language knowledge – will be needed for meeting the goal, and controlling the execution of the plan by retrieving and organizing the language elements. To successfully apply strategic competence in reading, North and Schneider (1998) suggested that the test takers should adapt style and speed of reading to different texts and purposes, and read with a large degree of independence, using appropriate reference sources selectively.

The second viewpoint on reading construct that this literature review deals with is **background knowledge** (referred as discourse competence in North & Schneider, 1998). According to Douglas (2000), background knowledge gives a central role to the cognitive construct of discourse domains which are frames of reference based on past experience, which we use to make sense of current input and predict forthcoming content. To focus on discourse competence in reading, test takers need to distinguish in detail the various parts of the treatment of a theme and understand their interrelations, separate the main ideas and details from lesser ones, recognize the line of argument in the treatment of the issue presented, though not necessarily in detail (North & Schneider, 1998).

Apart from the aforementioned viewpoints, **vocabulary knowledge** which is a part of language knowledge should be stressed since it plays a role in ESP reading comprehension. From a summary of the interview conducted with 10 engineering experts (see Appendix B), vocabulary was mentioned to be assessed in both terms of linguistic factor and engineering factor. This is because experts feel that there is a common sense relationship between vocabulary and comprehension. To clarify, messages are composed of ideas, and ideas are expressed in words. In addition, most theorists and researchers in education have assumed that vocabulary knowledge and reading comprehension are closely related, and numerous studies have shown the strong correlation between the two (Baker, 1995; Nagy, 1988; Nelson-Herber, 1986 cited in Smith (1997). Based on previous research studies and the need of stakeholders, vocabulary knowledge of test takers in this study is measured in terms of language and engineering knowledge.

The following table summarizes the components of reading construct in this research study.

Table 2.1: Components of Reading Construct in ESP Context (Adapted from Douglas, 2000 and North & Schneider, 1998 which is based on Bachman and Palmer's 1996 framework)

---

1. Language Ability

Language Knowledge

Grammatical knowledge (Linguistic competence)

Knowledge of vocabulary

Knowledge of morphology and syntax

In reading, the test takers should understand texts which contain complex and unfamiliar language, have a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms, understand grammatical patterns and vocabulary ordinarily encountered in academic/ professional reading.

Textual knowledge

Knowledge of cohesion

Knowledge of rhetorical or conversational organization

In reading, test takers should be able to understand the grammatical or lexical relationship between different sentences or between parts of a sentence.

Sociolinguistic knowledge

Knowledge of dialects/ varieties

Knowledge of registers

Knowledge of idiomatic expressions

Knowledge of cultural references

In reading, the test takers should understand many socio-linguistic and cultural references.

Strategic Competence

Assessment

Goal setting

Planning

Control of execution

In reading, the test takers should be able to adapt style and speed of reading to different texts and purposes, and can read with a large degree of independence, using appropriate reference sources selectively.

2. Background Knowledge

Discourse domains

Frames of reference based on past experience which we use to make sense of current input and make predictions about that which is to come.

In reading, test takers should distinguish in detail the various parts of the treatment of a theme and understand their interrelations, separate the main ideas and details from lesser ones, recognize the line of argument in the treatment of the issue presented, though not necessarily in detail.

It can be concluded that a test designer needs to define carefully what the construct is, and to what extent the test includes, or should exclude. This will relate to the purpose for designing the test. Regarding this study, two main components, language ability and background knowledge are included. Since these two variables are the key concerns in this study and there is a relationship between them, the details will be explained in the next section.

2.4 Background Knowledge and Language Proficiency

The role of background knowledge in reading comprehension has been long mentioned. Background knowledge plays a key part in the reading process. That is, reading is seen as an active process of constructing meaning by connecting old knowledge with new information encountered in a text. New information is learned and remembered best when it is integrated with relevant background knowledge.

There is extensive terminology to describe different kinds of knowledge. Consistency in the use of these terms is a recognized problem; subtle and dramatic differences exist between different people's definitions of the same term (Alexander, Schallert, & Hare, 1991; Dochy & Alexander, 1995). The terms "background knowledge" and "prior knowledge" are generally used interchangeably. In addition, schema is sometimes used as a general umbrella term related to background knowledge and prior knowledge. Many scholars, for example, Stevens (1980: 151) defines background knowledge quite simply as "…what one already knows about a subject..." Biemans and Simons' (1996: 6) definition of background knowledge is slightly more complex, "…[background knowledge is] all knowledge learners have when entering a learning environment that is potentially relevant for acquiring new knowledge." Dochy & Alexander (1995) provide a more elaborate definition, describing prior knowledge as the whole of a person's knowledge, including explicit and tacit knowledge, metacognitive and conceptual knowledge. This definition is quite similar to Schallert's definition (1982) cited in Strangman and Hall (2004). Thus, while scholars' definitions of these two terms are often worded differently, they typically describe the same basic concept. In this research project, the word "background knowledge" is used through all chapters for reasons of consistency and congruence with the research's title. It is operationally defined as content knowledge, and in particular with subject-specific knowledge which is acquired mainly from school, specifically in the academic discipline of engineering.

According to Pearson, Roehler, Dole, and Duffy (1991), good readers use their background knowledge to make predictions, visualize, ask questions to monitor comprehension, draw inferences, confirm hypothesis, determine what is important in text, and demonstrate to others that they have understood what they have read. It is very useful for readers to make use of background knowledge in order to comprehend the text for purposes of learning or testing. There are three kinds of background knowledge involved in reading comprehension. The first is specific knowledge about the topic of the text. The second is general world knowledge about social relationships and causal structures. The third is knowledge about text organization or genre. These are what the readers have to encounter in the text when they read.

The effect of background knowledge on reading comprehension has always been a central theme in EFL reading research. Over the past decades, there have been several studies into the effect of background knowledge on ESP test performance (Clapham (1996, 2000), Alderson and Urquhart (1983, 1985a, 1985b), Koh (1985), Tan (1990), Shoham, Peretz, and Vorhaus (1987), Johnson (1981) and Floyd and Carrell (1987). The details of these studies will be presented in the next section. To summarise, all of these studies concern the effect of background knowledge on ESP reading test performance. However, it is difficult to distinguish between background knowledge and language proficiency/ ability when interpreting test results. Therefore, language proficiency is reported in relation to background knowledge. There are a few studies which suggest that under certain conditions, background knowledge does not influence language test performance to any significant degree, but language ability plays a role instead. On the other hand, several other studies have found significant interactions between background knowledge and language test performance. It appears that, under some conditions at least, background knowledge makes a difference to language test performance.

It can be concluded that background knowledge plays a key part in reading comprehension. Background knowledge and language proficiency have long been investigated in terms of their effect on the test performance in ESP tests. There are a number of advocates for background knowledge and at the same time, there are also a number of researchers who believe that language proficiency plays a more important role. Previous research studies on the effects of background knowledge are reviewed in the next section.

2.5 Research in ESP Reading

Since this research aims at investigating the effects of language ability and engineering background knowledge on ESP reading ability, the discussion in this section will deal mainly with previous research on ESP reading in engineering.

From a review of previous research on ESP reading in engineering, it was found that the purpose of most studies was two-fold, to develop ESP reading tests particularly for engineering students (Brown, 1981, 1988, 2005; Erickson & Molloy, 1983) and to study other variables (e.g. background knowledge, language proficiency, text familiarity, level of skill processing involved in reading, reading topics and reading tasks) which affect reading ability of mixed-discipline students including engineering students (Alderson & Urquhart, 1983, 1985a, and 1985b; Clapham, 1996, 2000; Hudson, 1993; Shoham, Peretz, and Vorhaus, 1987).

However, since this study focused on the effect of background knowledge and language ability on ESP reading in engineering, discussion will centre on only these two variables.

Most studies on the effects of background knowledge and language ability on ESP reading performance were undertaken with mixed academic disciplines students including engineering students. This is probably because different research questions and research designs which determine different samplings or subjects were involved in each study. In addition, instruments used in those studies particularly the ESP reading tests were developed. These tests consisted of several reading modules (related to the subjects' fields of study) so that a comparison of the results could be made among different disciplines such as Science and Technology, Life and Medical Science, Business Studies and Social Science (Clapham 1996, 2000). At the moment, there have been few studies concerning reading test construction in the field of engineering, particularly to investigate background knowledge and language ability in ESP reading performance. In a review of such literature where comparisons between different disciplines is made, it was found that reading tests in science and technology, rather than engineering modules were administered. In view of the fact that engineering is an applied science represented in this broad academic context, and furthermore, that engineering groups are known to exhibit test behaviour very similar to students from science and technology (Alderson & Urquhart, 1985a), the review of the relevant literature in this thesis will focus on the broad areas of science and technology. Further justification for the narrow focus of the literature review in this thesis is that results obtained from reading research based on science and technology disciplines approximate those from engineering. Lastly, a literature review restricted to only engineering would be too specific and specialized. It is hoped that by reviewing studies in these fields, a clearer picture of ESP reading in engineering will emerge.

The effect of background knowledge on reading comprehension has always been a central theme in EFL reading research. Over the past decades, there have been several studies examining the effect of background knowledge on ESP test performance.

Nieh (2000) conducted a study at the Southern Taiwan University of Technology on two groups of students: one comprised engineering majors and the other, business. All were at similar levels in general English reading proficiency. The students were asked to complete a multiple-choice cloze test as a test of reading comprehension. The reading comprehension test consisted of two parts: one was concerned with engineering and the other with business. It was used to diagnose the subjects' reading comprehension proficiency in these two fields. It was found that there were highly significant differences in the reading comprehension tests between engineering and business major students. Students with more background knowledge concerning the content of reading tests had better understanding of the reading comprehension than those who had insufficient and inappropriate background knowledge. Therefore, these studies confirm that background knowledge plays an important role in reading comprehension, particularly in a foreign language.

Articles by Alderson and Urquhart (1983, 1985a, and 1985b), aroused considerable interest and led to several follow-up studies. These articles described three studies carried out with students attending English classes in Britain in preparation for studying in different disciplines at British universities. The students varied in language ability. In each study, Alderson and Urquhart compared student scores from reading texts related to their own field of study with those from texts related to other subject areas. In the first and second studies, students were tested with gap-filling passages related to three content areas. In the third study, three groups of students from different disciplines – business and economics, science and engineering, and liberal arts – took the Social Studies and Technology Modules of the reading test. The students' scores on the modules were contrary to expectations that students majoring in a discipline would perform better in their subject area than those outside it. It was found that science and engineering students taking the Technology module test did better than the business and economics students as well as liberal arts students who took the same test. However, when they took the test in the liberal arts, they achieved scores similar to liberal arts students although their language proficiency was lower. And when they took the test in the social studies module, there was no significant difference between their scores and those of the business and economic students. Alderson and Urquhart concluded that background knowledge had some effect on test scores, but not consistently so. They suggested that future studies should take account of linguistic proficiency and other factors as well.

Koh (1985) had somewhat similar results when she conducted her study at Singapore University with three groups of students – two in Science and one in Business Studies. Analysis of variance was used to estimate the effect of background knowledge on cloze test results. It was found that though there was an interaction between the student group and the test, the students did not always do best in their own subject areas. The business students, for example, had the highest scores on the science text. Nevertheless, it turned out that half of these students had studied science previously, so it could be that background knowledge was affecting their scores. The group with the highest language proficiency, one of the science groups, did consistently better than the other two groups for all the tests which dealt with topics on business, history, politics and science. She concluded that background knowledge did affect test scores but that ignorance of the subject matter could be compensated for by high linguistic proficiency.

Another study where the findings were not consistent across disciplines was conducted by Shoham, Peretz, and Vorhaus (1987). They reported a study conducted at Ben-Gurion University in Israel. The study was designed to investigate the relevance of students' discipline background on tests of reading comprehension in EFL. This study was conducted with students from three faculties: Science and Technology, Biology, and Humanities and Social Science. The students were tested on three texts related to their respective fields of study. It was found that students of Science and Technology obtained the highest mean on the entire test as well as the highest mean on the individual test passages (except the biology-related passage where they scored only slightly lower than the biology students) especially in the scientific texts. While students in the science and technology group performed better with scientific texts, the humanities and social science students did not do better on the test in their own subject areas, a result which contradicted the researcher's prediction. To explain this, Clapham (1996:10) pointed out that there seemed to be a tendency for science and technology students to perform better than other students at science-based tests, but to perform better or as well as the humanities students on humanities-based ones. Similarly, Alderson (2000:62) pointed out that non-specialist texts in arts and humanities, and to some extent in the social sciences, would be easier to process by more readers of equivalent educational background than scientific texts, on the assumption that more people would have read fiction, popular journals, and advertisements than technical or scientific texts.

Another interesting study was conducted by Moy (1975 cited in Shoham et al., 1987). She investigated the effect of content familiarity and English proficiency on reading cloze scores of students at the Chinese University of Hong Kong. The students were grouped into three broad categories: arts, sciences and social sciences. It was found that there was a significant interaction between the academic filed of subject and content matter of test passages. The social science majors had the highest mean scores on a history passage on which it was predicted that the Arts majors would do best. However, although science majors in this study obtained the highest mean scores on the scientific passage, they also had the highest mean scores on the other passages. Moy suggested that the importance and the difficulty of controlling for passage difficulty across content areas as well as for language proficiency across faculties should be taken into consideration.

Other different results were obtained by Clapham (1996). In her study of the effect of background knowledge on IELTS reading comprehension, Clapham (1996) conducted a large scale study with 842 non-native English speakers from different disciplines including engineering. Most of them were about to start undergraduate or postgraduate studies at English medium universities. She found that both students' field of study and familiarity with the subject area were significantly related to test scores. However, there was a stronger relationship between the level of English language proficiency and the test scores. Interestingly, students did not appear to be affected by background knowledge until they achieved scores of over 60% on the grammar test which all subjects needed to take. Similarly, students with grammar scores of over 80% appeared to make less use of background knowledge than the intermediate students. She commented that the low proficiency students could not take advantage of their background knowledge while the high proficiency students did not need to rely on the background knowledge as their language proficiency had already facilitated their understanding. This led her to conclude that test takers at intermediate or threshold level language competence will benefit the most from their background knowledge. Those with low language competence as well as those with high language competence could not benefit as much.

Based on the aforementioned studies, some points need to be discussed and justified as follows:

First, it should be noted that most of the studies concerning the effects of background knowledge and language ability were conducted with sample groups from mixed-academic disciplines, and the results were compared across different disciplines. However, in many studies the results turned out to be inconclusive and inconsistent. One possible reasons for this could be that in situations where students from different disciplines are grouped in broad categories, such as science and engineering or humanities and liberal arts, the construction and administration of different content-related tests would not have been justified and/or easy to control. The text selection based on the levels of specificity and levels of text difficulty in different disciplines may vary and could be factors affecting the results of the study. Therefore, this research replicated select procedures from previous studies but used a single background discipline with reading passages which were more specifically related to students' specialized field of study as suggested by Shoham et al. (1987). In this study, engineering students were selected to perform a reading test in engineering. In this way, familiarity with content or background knowledge could be better controlled.

Second, it should be noted that Alderson and Urquhart (1983, 1985a, 1985b), Koh (1985) and Moy (1975 cited in Shoham et al., 1987) used the cloze test in their studies. It is questionable whether the cloze format is the best way to assess the kind of reading comprehension required for reading academic textbooks and professional journals. The filling-in of specific blanks is better able to predict knowledge of vocabulary, syntax, and grammar than comprehension, understanding of the content and author's purpose and viewpoint. Therefore, in this study, cloze or gap-filling test items were excluded. A reading comprehension test in engineering with a multiple choice format was employed.

Third, based on the findings from the aforementioned studies, there is a tendency for students of science and technology to perform better than other students on their field-related texts and obtain the highest scores on the entire test. This fact is in agreement with Shoham et al. (1987) who explained the fact that science and technology students did well on an entire test possibly because the entrance requirements of science and technology are higher than those of liberal arts, humanities or social science. The higher scores of science and technology students could, therefore, reflect a higher level of overall competence. Another explanation for the better performance of science and technology students on the entire test might be a result of the particular comprehension strategies developed by this group of students. Science and technology students are taught how to comprehend a process, hypothesize, and experiment, etc. whereas liberal arts or humanities and social science students usually read in order to determine an author's main idea, supporting ideas, and general viewpoint. This difference in the approach to the reading material might result in the development of more effective reading comprehension strategies. Nonetheless, language proficiency is also another factor which affects success of students in reading comprehension tests. In some cases (Koh, 1985), business studies students with high language proficiency could perform better than science and technology students with low language proficiency on the scientific text.

Based on these studies, the researcher planned a follow-up study with some points of consideration in mind. The research design was partially changed. The focus was on a single background discipline with reading passages that were more specifically related to the engineering area for better control of background knowledge. A reason to focus on engineering instead of other disciplines is that at the researcher's workplace (KMUTT), engineering students constitute the majority of graduate students. Thus it was important to develop and trial a reading test for engineering. In addition, engineering students differ in their language ability. However, non-engineering groups such as liberal arts and industrial education were also included in this study to ensure construct validity of the reading test. Since the medium of instruction for liberal arts students is English, they were assigned to a high language ability group while industrial education students were grouped into a low language ability group. It is hoped that through this research, inconclusive issues from several previous studies in the same area can be better clarified, particularly with regard to ESP reading ability in engineering.

It can be concluded that background knowledge and language proficiency have long been investigated in terms of their effects on ESP test performance. There are a number of advocates for background knowledge and at the same time, there are also a number of researchers who believe that language proficiency plays a more important role. In the aforementioned studies, the results are inconclusive in some academic disciplines such as liberal arts and business compared to engineering or science and technology. However, a closer examination of engineering or science and technology students, in general, reveals a tendency for them to perform well in reading tests both within and outside their field of study, depending on their language ability. This will be considered when constructing

Engineering English reading tests. The next part in this chapter will present a review on the UCLA engineering English reading test.

2.6 The UCLA Engineering English Reading Test

To develop an ESP test to assess graduate-level engineering students (non-native speakers of English) at KMUTT, tests concerning specific-field in engineering were reviewed. However, since there is a limited number of a research study in this area (Brown, 1988, 2005; Erickson and Molloy, 1983), the test developed by a team at UCLA which is relatively close and fits to this research study was selected as a guideline for constructing the E-ERT. The following are details about the test development of the UCLA test.

Brown (1988: 193) reported that in November 1979, a project was begun at UCLA to explore, develop and evaluate a methodology for creating ESP tests. The initial interest focused on EST for engineers because of the large number of foreign students in the UCLA Engineering Department (one of the largest departments on campus). Part of the rationale was the assumption that nonnative English-speaking (NNS) students in an academic setting must perform on a par with native English-speaking (NS) students in order to succeed in a university. To determine how NNS students must perform, normative data on NS students were needed. Consequently, the test development project included piloting test items first on NS students and then on NNS students, all of whom were in undergraduate courses at UCLA.

The overall purpose of the UCLA study was 1) to investigate the suggested ten item types as reading skills to be tested and possible components of engineering reading ability and 2) to investigate whether or not the developed test can significantly distinguish between engineering and non engineering students for both native and non-native English-speaking students.

All of the subjects in the study were university graduate students at UCLA. They were classified into four groups: American engineers, Americans in TESL, Chinese engineers and Chinese in TEFL.

The main instrument used in the study was the UCLA engineering English reading test. It was designed to assess non-native English speakers in their engineering-English reading ability. In order to obtain appropriate item types for the engineering English reading test, the researchers at UCLA consulted with engineering professors, and examined the literature on ESP theory and linguistic theory. Ten different types of items were suggested by the content specialists. They were grouped into linguistic factors and engineering factors based on the literature and input provided from engineering professors. The following figure illustrates the two categories of test items for engineering students at UCLA.

I. Linguistic factor
- A. Cohesion
  1. Reference items
  2. Substitution items
  3. Lexical cohesion items
  4. Conjunction items
- B. Non-technical vocabulary items

II. Engineering factor
- A. Fact items
- B. Inference items
- C. Lexis
  1. Subtechnical vocabulary items
  2. Technical vocabulary items
- D. Scientific rhetorical function items

Figure 2.1: Two categories of test items for engineering students at UCLA

After successfully identifying the test items, the draft of the UCLA engineering reading test was developed based on material taken from *Introduction to Ceramics*, the textbook used in Engineering 146A on the following topics: (1) the mechanics of deformable bodies, (2) refractories and (3) thermodynamic analysis of heat pumps. These were topics felt by the engineering professors to be common to most engineering students. In addition, the engineering professor was asked to select several passages which met the following criteria (Erickson and Molloy, 1983: 284-285):

1. The passages should be without illustrations, diagrams, or mathematical formulas in order that no information other than that provided by the reading passage could be used by the test takers in responding to questions.

2. The passage should approximate a self-contained unit with respect to content. It would be chosen from mid-chapter so that remarks made in either introductory or concluding sections would be excluded. It was felt that introductory and concluding sections would not contain sufficient information for the reader in terms of either content or use as a reference point from which to test.

Once the passages were identified, three passages from 600 to 800 words in length were selected. In addition, the level of passage's difficulty was determined on the basis of three readability formulas (Flesch, 1948; Lorge, 1959; and Fry, 1977 cited in Erickson and Molloy, 1983). The test items were then written according to the 10 aforementioned test items based on linguistic and engineering factors. Consequently, the UCLA engineering reading test contained 60 items with two hours and thirty minutes for time allotment. The test was administered to the Chinese subjects under controlled classroom conditions. The engineers and TEFL students were tested separately, but under controlled classroom conditions within the time limits.

The results showed that there was a significant difference in performance on the test between native and non-native students as well as a significant difference in performance on the test between engineering and non-engineering majors. There were no significant interactions. In addition, concerning an investigation on 10 item types in the test, it was revealed that some items such as "inference", "sub-technical" and "technical" items are more efficient than others. In contrast, some items like "reference", "conjunction", and "non-technical" items appear to be less efficient than others.

By having the UCLA test as a guideline and making use of the previous results of the study, some stages of the test development and some item types were followed in order to construct the E-ERT. The details of the test development of the E-ERT are presented in Chapter 3. The next part presents item types as skills or sub-skills involved in the E-ERT.

2.7 Item Types as Skills/ Sub-skills Involved

Despite the fact that psychologists and educators have been conducting research on various aspects of reading skills for more than a century (Alderson, 1990a, 1990b, 2000; Alderson & Urquhart, 1983; Dewey, 1935; Johnston, 1983; Robinson, 1966; Singer & Ruddell, 1976; Smith, 1970; and Thorndike, 1917a, b, c), there are still controversies surrounding the exact nature of skill, or skills, that are involved in reading comprehension either in L1 or L2. In general, however, studies that have addressed the nature of reading skill(s) seem to have subscribed to one of two views. The first maintains that reading skill is a unitary, holistic, and indivisible skill which cannot be split into different sub-skills (Alavi, 2002; Alderson, 1990a, b; Andich & Godfrey, 1979; and Rost, 1993 cited in Ghahraki and Sharifian, 2005). The second viewpoint claims that reading skill consists of various sub-skills (Bloom, 1956; Davis, 1968; Mirhassani & Khosravi, 2002; and Munby, 1978 cited in Ghahraki and Sharifian, 2005).

According to Stein and Glenn (1979 cited in Ghahraki and Sharifian, 2005) and Downing (1982 cited in Ghahraki and Sharifian, 2005), skilled readers often simultaneously use particular sub-skills of a reading skill, and over the years, these originally distinct sub-skills, become fused, without being activated separately. Hughes (1989) refers to "macro skills" and "micro skills" of reading comprehension. The distinction between these two levels of sub-skills is not made explicit. However, it appears that the term "macro skills" refers to understanding the general ideas in the text such as information, gist and argument, while "micro skills" refers to recognizing and interpreting the linguistic features of the text such as referents, word meanings, and discourse indicators. Hughes maintains that micro skills should be taught not as an end in themselves but as a means of improving macro skills. In addition, Brown (2004a) claims that macro and micro skills are a crucial consideration in the assessment of reading ability. He proposes a list of macro and micro skills as the spectrum of possibilities for objectives in the assessment of reading comprehension. For example, Macro skills

- Infer context that is not explicit by using background knowledge.
- From described events, ideas, etc., infer links and connections between events, deduce causes and effects, and detect such relations as main idea, supporting idea, new information, given information, generalization and exemplification.
- Distinguish between literal and implied meaning.
- Develop and use a battery of reading strategies, such as scanning and skimming, detecting discourse markers, guessing the meaning of words from context, and activating schemata for the interpretation of texts.

Micro skills

- Recognize a core of words, and interpret word order patterns and their significance.
- Recognize grammatical word classes (nouns, verbs, etc.), systems (e.g. tense, agreement, pluralization), patterns, rules, and elliptical forms.
- Recognize that a particular meaning may be expressed in different grammatical forms.
- Recognize cohesive devices in written discourse and their role in signaling the relationship between and among clauses.

All of these skills are normally applied while reading for purposes of both learning and testing.

Concerning the assessment of reading, there are numbers of studies constructing reading tests in order to examine students' reading ability or comprehension (e.g. Alderson and Urquhart, 1985; Brown, 1988, 2005; Clapham, 1996; Erickson & Molloy, 1983; Hudson, 1993; Koh,1985; Shoham, Peretz, and Vorhaus, 1987; and Tan, 1990). Generally, a reading test consists of two or three subtests concerned with grammar, vocabulary and reading comprehension. Test items as skills or sub-skills are then included and defined before test construction.

In this research project, several types of test items are included in the reading test based on a literature review, previous research and stakeholders' needs. The test consists of two subtests: vocabulary and reading comprehension in engineering.

For vocabulary subtests, the following sub-skills are included:

- Sub-technical terms: Lexical items with technical as well as non-technical senses i.e. iron, force, stress and tension.
- Non-technical vocabulary: General vocabulary used in engineering text i.e. act, apply, calculate, illustrate and indicate.

For the reading comprehension subtests, the items include the following:

Sub-skills:

- Main idea: A gist or overall idea of the text.
- Specific detail: A phrase or clause taken directly from the text.
- Inference: A statement not taken directly from the text, but inferred from the text by the reader.
- Fact: a fact that needs background knowledge in engineering to answer a question

From the aforementioned definition of item types in this study, each item type is explored to investigate different test taking strategies that students use to answer such test items.

In conclusion, in order to construct the reading test, it is important to include item types as the construct of the test. These item types are seen as skills or sub-skills involved when students perform the test. They should be clearly defined right from the beginning. A study on how students answer different types of test items will be explored in relation to test taking strategies. The test taking strategies are reviewed in the next section.

2.8 Test Taking Strategies

Language use strategies are mental operations or processes that learners consciously select when accomplishing language tasks. These strategies also constitute test taking strategies when they are applied to tasks in language tests. For the purpose of this research study, test taking strategies will be viewed as test taking processes that the test takers have selected and of which they are conscious. In other words, the notion of strategy implies an element of selection.

According to Cohen (1998), it is best not to assume that any test taking strategy is a good or a poor choice for a given task. That evaluation depends on how individual test takers – with their particular cognitive style profile and degree of cognitive flexibility, their language knowledge, and their repertoire of test taking strategies – employ the strategies at a given moment on a given task. Some test takers may get by with using a limited number of strategies, and using them well for the most part. Others may be aware of an extensive number of strategies but may use few, if any of them, effectively. In addition, the strategy, while successful for one test taker, may not work well for another test taker. Since this study focuses on a reading test with a multiple choice format, a review of previous studies on test taking strategies on reading tests particularly with multiple choice items will be discussed.

Strategies for taking reading tests

In considering strategies for answering reading tests, there is some research focusing on the multiple choice format.

In a study conducted by Gordon (1987 cited in Cohen, 1998), 30 tenth-grade EFL students – 15 high proficiency readers and 15 low proficiency readers, were asked to verbalize thoughts while finding answers to open-ended and multiple choice questions. The researcher found that answers to test questions did not necessarily reflect comprehension of the text. Both types of reading comprehension questions were regarded by the respondents as "mini" reading comprehension tests. With respect to test taking strategies, the low proficiency students tended to process information at the local (sentence/ word) level, not relating isolated bits of information to the whole text. They used individual word-centered strategies like matching words in alternatives to test, copying words from the text, translating word for word, or formulating global impressions of text content on the basis of key words or isolated lexical items in the text or test questions. The high proficiency students, on the other hand, were seen to comprehend the text at a global level-predicting information accurately in context and using lexical and structural knowledge to cope with linguistic difficulties.

In another study by Larson (1981 cited in Cohen, 1984), older respondents were involved. 40 college ESL respondents were required to give retrospective verbal reports to provide insights about test taking strategies. The students were requested to describe how they arrived at answers to a 10-item multiple choice test based on a 400-word reading passage. Seventeen students met with the author of the test in groups of two or three within 24 hours after the test, while 23 students met in groups of five or six 4 days after taking the test. The researcher found that the respondents used the following strategies: 1) they stopped reading alternatives when they got to the one that seemed correct to them, 2) they matched material from the passage with material in the item stem and in the alternatives (e.g. when the answer was in the same sentence as the material used to write the stem), and 3) they preferred a surface-structure reading of the test items to one that called for more in-depth reading and inference.

Another study of test taking strategies among non-native speakers (Anderson, Bachman, Perkins & Cohen, 1991) revealed that respondents used certain strategies differently, depending on the type of question they were responding to. For example, the strategies of “trying to match the stem with the text” and “guessing” were reported more frequently for inference questions than for direct statement and main idea question types. The strategy of “paraphrasing” was reported to occur more in responding to direct statement items than with inference and main idea question types.

What makes the study of Anderson et al. more interesting is the triangulation of data sources. That is, in the follow-up phase of the research, data from the participants’ retrospective think-aloud protocols of their reading and test taking strategies were combined with data from a content analysis and an item analysis to obtain a truly convergent measure of test validation. The content analysis of the reading comprehension passages and questions was comprised of the test designer’s analysis and one based on an outside taxonomy, and the item performance data included item difficulty and discrimination. This study marked perhaps the first time that both think-aloud protocols and more commonly used types of information on test content and test performance were combined in the same study in order to examine the validation of the test in a convergent manner.

Emerging from these various studies on multiple choice tests of reading comprehension is a series of strategies that respondents may utilize at one point or another in order to arrive at answers to the test questions. The following table presents a composite list of some of the more salient test taking strategies appearing in one or more of studies mentioned previously as suggested by Cohen (1998: 103).

Table 2.2: Strategies for Taking a Multiple Choice Reading Comprehension Test

1. Read the text passage first and make a mental note of where different kinds of information are located.
2. Read the questions a second time for clarification.
3. Return to the text passage to look for the answer.
4. Find the portion of the text that the question refers to and then look for clues to the answer.
5. Look for answers to questions in chronological order in the text.
6. Read the questions first so that the reading of the text is directed at finding answers to those questions.
7. Try to produce your own answer to the question before you look at the options that are provided in the test.
8. Use the process of elimination i.e. select a choice not because you are sure that it is the correct answer, but because the other choices do not seem reasonable, because they seem similar or overlapping, or because their meaning is not clear to you.
9. Choose an option that seems to deviate from the others, is special, is different or conspicuous.
10. Select a choice that is longer/shorter than the others.
11. Take advantage of clues appearing in other items in order to answer the item under consideration.
12. Take into consideration the position of the option among the choices (first, second, etc.).
13. Select the option because it appears to have a word or phrase from the passage in it – possibly a key word.
14. Select the option because it appears to have a word or phrase that also appears in the question.
15. Postpone dealing with an item or selecting a given option until later.
16. Make an educated guess i.e. use background knowledge or extra textual knowledge in making the guess.
17. Budget your time wisely on the test.
18. Change your responses as appropriate i.e. you may discover new clues in another item.

To answer research question 4, the composite list above will be employed as a guideline in this study.

In conclusion, test taking strategies are the processes that test takers make use of in order to give acceptable answers to questions and tasks. For language assessment, the obtained data on test taking strategies can be used to validate the test. It can provide some insight into what items are actually testing, aiding us in making decisions about which items to include and which to throw out. Many studies on the relationship between the test tasks particularly in the multiple choice format and the strategies used were reported as empirical evidence of strategies used by test takers. These strategies were used as a guideline for the present study. Another objective of this study was to investigate students' attitudes towards the test. A review of attitudes is presented in the next part.

2.9 Attitudes towards the Test

There are several kinds of factors affecting test takers' performance. As Bachman (1990) points out, communicative language ability, test method and test takers' characteristics are three important categories influencing test takers' performance. The third category, test taker's characteristics, is an interesting factor to be investigated since it can directly affect test scores. In addition, a recent concern among researchers in the field of language testing has been the identification of the test takers' characteristics that influence performance on tests of English as a Foreign Language (EFL) or English as a Second Language (ESL). Kunnan (1995) briefly explains that these test takers' characteristics include personal attributes such as age, native language, gender, educational characteristics such as background knowledge, previous exposure to English as well as cognitive, psychological and social characteristics including learning strategies and styles, attitude and motivation, aptitude and intelligence, field dependence and independence, extroversion and introversion and anxiety, personality and risk taking.

Moreover, Saville (2000) identifies test takers' characteristics as sets of background factors that are hypothesized to affect second language acquisition and second language test performance. The factors can be grouped as strategic and socio-psychological and can be further classified into: (a) strategic factors: cognitive strategies; metacognitive strategies; and communication strategies, and (b) socio-psychological factors of attitude, anxiety, motivation and effort.

Therefore, one of the important factors that can affect test takers' performance is test takers' characteristics which consist of several aspects, one important one being attitude. This is one component of academic success and can predict test takers' behavior. Baron and Byres (1994) cited in Whittemore (2004) say that attitudes are positive or negative evaluations of things ranging from a solid object to ideas about oneself that are stored in memory, and related to one another. Furthermore, Eagly and Chaiken (1998) cited in Sjoberg (2005) define attitudes as a psychological tendency that is expressed by evaluating a particular entity with some degree of favor and disfavor.

Attitude is an interesting topic for research in several areas. Ten thousand references per year is an astounding number (Sjoberg, 2005). Evidence from numerous studies has shown the importance of attitudes. In addition, Gardner et al. (1976 cited in Kunnan, 1995) and Lambert (1963, 1967 cited in Kunnan, 1995) point out that attitudes and motivation can influence and achieve successful second language achievement. Furthermore, Rand (1997) suggests ways to conduct tests that involve the demonstration of language skills. He mentions the importance of attitudes towards the test and suggests that positive attitudes of test takers towards testing should be created. This is because it can bring about better test results that both teachers and students desire. Currently, students are afraid of tests because they view them as unfair, difficult, stressful and irrelevant to the course material studied. With positive tests, classroom motivation can be increased throughout the course, which in turn will lead to improved performance.

Bachman and Palmer (1996) also suggest ways to create positive attitudes in the test-takers, one of which is inclusion in all phases of test development. Test developers should involve test takers in the design and development of the test by collecting information from them about their perceptions of the test and test tasks. Bailey (1999) also points out that if the researchers feel that test-takers are involved in this way, they will perceive tests as more interactive and authentic, and will therefore be more motivated, which could lead to enhanced preparation and hence to better performance.

It can be concluded that one important factor that can cause different test performance is test takers' characteristics. An attitude is one of test takers' characteristics which can cause and affect successful foreign and second language achievement. Positive attitudes towards testing can lead to better test results and can help improve student performance. Some ways to create students' positive attitudes towards testing are to involve them in the design and development of the test. Then, they will perceive tests as more interactive and authentic, resulting in more motivated and better performance.

In conclusion, this study aims to investigate the effects of language ability and engineering background knowledge on ESP reading ability. Furthermore, a study on how test takers use different test taking strategies in answering different types of test items is conducted. Moreover, test takers' attitudes towards the test are examined. In order to fulfill those purposes, developing a reliable and valid reading test for measuring non-native English speakers in engineering-English reading ability is needed. A group of engineering students is thus tested and their performance compared to that of non-engineering students.

Summary

Chapter two presents a review of related literature that provides the underlying concepts of this study. The review includes an introduction to ESP testing, a review of ESP reading theories, constructs of reading, background knowledge and language proficiency, research in ESP reading, the UCLA engineering English reading test, item types as skill/ sub-skills involved, test taking strategies, and attitudes towards the tests. They are then employed as the basis for instrument development, data collection and analysis and interpretation of the findings. Chapter three presents the research methodology of this study.

CHAPTER III

RESEARCH METHODOLOGY

Chapter 3 deals with the research methodology regarding the population and sample, the development and validation of the research instruments, data collection and data analysis.

3.1 Population and Sample

3.1.1 Population

The population consisted of of 359 first-year graduate students at King Mongkut's University of Technology Thonburi (KMUTT) in the Master's degree program at the faculty of engineering. They represented different disciplines such as chemical engineering, civil engineering, computer engineering, environmental engineering, electrical engineering, and mechanical engineering. 236 students from other schools or faculties such as liberal arts, industrial education and technology, energy environment and materials, science, and bioresource and Technology also participated as subjects in the non-engineering group to ensure construct validity of the test and to prove one hypothesis of the study on background knowledge. It was anticipated that if the test had construct validity, it could distinguish those engineering from non-engineering students. The population was homogeneous in terms of nationality, all being Thai students in the same university.

3.1.2 Sample

The subjects were 120 students from the academic year 2007 and were chosen by means of simple random sampling with randomized block design to control extraneous variables and reduce errors in the study. Half of them were engineering students from different disciplines and the other half were non-engineering students. They were assigned to two groups of high and low language ability on the basis of their placement test scores. The high language ability group consisted of students whose placement test scores were at or above 1 standard deviation, while low language ability group referred to students whose placement test scores were at or below -1 standard deviation. The following figure illustrates the sampling design of the main study.

	High language ability(H)	*Low language ability(L)*
Engineering (E)	N = 30	N = 30
Non-engineering (NE)	N = 30	N = 30

Figure 3.1: Sampling design of the main study

Because the purpose of this study was to compare reading performance among subgroups (i.e. engineering and non-engineering groups, high language ability and low language ability groups) and two-way ANOVA was used for data analysis, the following considerations were taken into account to determine the sample size and sampling design.

- According to Tabachnick and Fidell (2001), N >= 30 is the number proposed for the sample size required in ANOVA. In addition, equal sample size in each cell is encouraged to provide complete data.
- According to Sudman (1976) and Kish (1965), an adjustment in the sample size may be needed to accommodate a comparative analysis of subgroups (e.g., such as an evaluation of program participants with non-participants). They suggested that a minimum of 100 elements is needed for each major group or subgroup in the sample and for each minor subgroup, a sample of 20 to 50 elements is necessary.

Based on these considerations and subjects' availability, the sampling design of this study consisted of equal numbers of 30 students as a minimum number in each subgroup, which is considered to be appropriate for the analysis that was planned.

In the pilot study, research tools were administered to 40 students during September 18-26, 2007 outside class hours in 4 different administrations. In the main study, the research tools were administered with 120 students during January – February 2008 outside class hours in 14 different administrations. The reasons for several administrations outside class were: 1) to avoid disrupting the teaching and learning process in class hours of two required English courses, 2) to accommodate students from different disciplines with different schedules and 3) because a testing room was reserved and available all day but could seat only about 10 students at a time. The subjects who took part in the pilot study were not included in the main study.

3.2 Research Instruments

Four instruments were employed in this research. They were the Engineering-English Reading Test (E-ERT), the KMUTT English placement test, introspective interviews and a questionnaire for test takers.

3.2.1 The Engineering-English Reading Test (E-ERT)

The Engineering-English Reading Test (E-ERT) was developed to assess students' ESP reading performance and elicit the use of engineering background knowledge of students. The E-ERT was designed to be a semi level of text specificity to avoid bias between engineering and non-engineering groups. If it was a highly specific text, non-engineering students could fail from the very beginning. On the other hand, if it was a very low specific text, the test was like a general English test and could not distinguish those who are engineering students and those who are not. The E-ERT consisted of general engineering texts with 40 multiple-choice items, and two main parts (vocabulary and reading comprehension). Each question contained four alternatives as a well amount of option when assessing reading as suggested by Kehoe (1995) and Alderson (2000: 204). Even though a number of criticisms have been leveled at the multiple choice format, it was chosen as the preferred model in this study for the following reasons.

First, multiple choice format is still a favorable alternative for language testing. Apart from the easy and reliable scoring, there are also a number of advantages that are worthy of consideration. For example, Ballantyne (2004) specified that multiple choice items can test a wide range of issues in a short time and the assessment by the multiple choice format is not affected by a writing student's ability. In addition, Chatterji (2003: 181) stated that multiple-choice items are more versatile than the other formats in their ability to tap a wide variety of cognitive levels, ranging from simple recall to more in-depth interpretation, application, complex generalization, or problem-solving skills. In this study, the multiple-choice format could help in tapping various abilities the test aims to measure i.e. the ability to understand main and specific details, the ability to draw inferences and the ability to identify the fact that uses background knowledge.

Second, from previous studies on Engineering-English reading tests such as those by Erickson & Molloy (1983), Brown (1988, 2005) and high stakes tests for multidisciplines such as IELTS (Clapham: 1996, 2000), the multiple choice format was preferred to other formats. The multiple choice format is easy to administer and to arrive at scores with reliable results as it does not require an expert opinion in rating the test performance. Therefore, the multiple choice format was employed in this study.

Before developing the test, the target language use and the construct of the E-ERT were identified based on ESP theory as mentioned in Chapter 2. Then, previous research in ESP reading tests was reviewed for an idea on test development. The following are details about the development of the E-ERT.

The Development of the E-ERT

To develop an ESP test to assess graduate-level engineering students (non-native speakers of English) at KMUTT, tests concerning specific fields in engineering were reviewed. Based on the literature review from pages 32 - 34, the test developed by a team at UCLA was selected as a guideline for constructing the E-ERT. The following are reasons to justify the use of the UCLA test.

1. As regards the objectives of the UCLA test project, they are very close and agree with the objectives of this research study. To clarify, one of the objectives at UCLA aimed at investigating the effect of background knowledge on engineering reading ability by the use of engineering reading test. The test was proved that it could distinguish between

engineering and non-engineering students. Therefore, it might be possible to apply some of the test development stages of the UCLA test for test construction of the E-ERT.

2. The UCLA test was designed to assess non-native English speakers in their engineering-English reading ability which fits with the purpose of the E-ERT development. In addition, the target group of the UCLA test was graduate engineering students which are similar to the sample of this study.
3. The UCLA test construct with the identified 10 item types are good components to assess engineering reading comprehension because they were guided by experts in an engineering field. In addition, the results on item types could contribute and be beneficial to the development of the E-ERT.

As regards the research presented in this thesis, the objectives of the research are three-fold. One is to investigate the effects of language ability and background knowledge (specific discipline in engineering) on ESP reading ability of Thai graduate students. Another is to examine test taking strategies on how students answer different types of test items. The third is to investigate test takers' attitudes towards the ESP test. By considering these objectives and conducting interviews with engineering experts, a focus on the two categories of item types emerged in a plan to include vocabulary (both sub-technical and non-technical) and reading comprehension in the field of engineering. It was assumed that these two categories would directly represent specific background knowledge in the engineering discipline and possibly permit students to use different test taking strategies while performing the test.

By adapting the test construct of the project at UCLA and modifying the perspectives of the experts interviewed for the components of reading construct in this study, the test construct of the E-ERT was partially changed; some were deleted and others were added for various reasons.

The decision to delete was made for the following reasons. Firstly, some ineffective items (i.e. reference and conjunction items) were reported in the study at

UCLA. Secondly, other linguistic components were excluded from this study since they were coincidently used in a reading comprehension process and were indirectly assessed during the reading comprehension so it was not necessary to assess linguistic knowledge separately. Lastly, testing purely linguistic knowledge might not have given clear result in terms of differentiating high from low language ability or engineering from non-engineering backgrounds.

Some parts of the E-ERT were added as desirable items (i.e. main idea and specific details items) because graduate-level students should be tested in reading as suggested in the GRE: Graduate Record Examination General Test (Brownstein, Weiner & Green, 1999) as well as recommended by the engineering experts. Furthermore, a section on vocabulary was added as another element of reading. Based on the interviews conducted with 10 engineering experts, vocabulary was mentioned to be assessed in both terms of linguistic factor and engineering factor due to its important role in reading comprehension. Readers cannot understand what they are reading without knowing what most of the words mean. In this study, sub-technical terms and non-technical vocabulary were included in this part because following research by Mudraya (2006), which revealed that the most frequent words in a specialist corpus of the engineering lexis for students are sub-technical and non-technical. Therefore, this test would have face validity. Both sub-technical terms and non-technical vocabulary in engineering were randomly selected based on a corpus-based study on Engineering English by Mudraya (2006). As a result, the following figure presents the engineering-English reading construct of this study.

I. Vocabulary
- A. Sub-technical terms items
- B. Non-technical vocabulary items

II. Reading comprehension
- A. Main idea items
- B. Specific details items
- C. Inference items
- D. Fact items

Figure 3.2: The engineering-English reading construct of this study

Based on this reading construct, the abilities to be measured in the E-ERT were consequently proposed as follows:

1. Ability to recognize the meanings of sub-technical terms used in engineering
2. Ability to identify the meaning of non-technical vocabulary in context
3. Ability to skim the texts for main ideas
4. Ability to scan the texts for specific information
5. Ability to draw inferences based on information in the text
6. Ability to identify the facts based on engineering background knowledge

Then, the table of specifications was constructed for an estimation of what was contained for a whole section in the test (see Appendix C). 45 multiple-choice items were constructed. The test consisted of 2 parts as follows:

Part I (a), (b): Vocabulary	20	questions
Part II: Reading comprehension	25	questions

Then, the draft of the E-ERT was developed. The test design criteria which required selecting engineering sub-technical and non-technical vocabulary, topics, texts, and considerations of length and difficulty of passages were taken into account for test development.

In terms of vocabulary, ten sub-technical terms and ten non-technical vocabularies in engineering were randomly selected based on a corpus-based study on Engineering English done by Mudraya (2006). These terms and vocabulary were in the one hundred most frequent word families in the student engineering word list. The definitions of sub-technical terms were defined by the use of the pocket illustrated dictionary of engineering terms (Timings and Twigg, 2001), www.reference.com, and http://en.wikipedia.org .

By topics, various topics concerning up-to-date and interesting technology, and fundamental concepts for all engineering students were chosen based on recommendations of engineering experts. All selected topics were (1) theory of relativity, (2) engineering projects, (3) vector mechanics for engineers and (4) an experiment on idea feasibility testing. These were topics felt by the engineering professors to be common to most engineering students.

In regard to text selection, from the interview with engineering experts, the researcher randomly selected several passages based on science and technology, and general engineering topics from several sources i.e. the book by Andreas (2005), the Encyclopedia Britannica (1993), the online-journal, Scientific America, (April 2006 and August 2006) and the textbook by Beer and Johnston (1977). Consequently, the professor was asked to select three passages to be included in a reading comprehension part which met the given criteria. Table 3.1 presents the criteria for text selection for the E-ERT.

Table 3.1 The Criteria for Text Selection for the E-ERT.

Criteria	Details
1. Authenticity of text type	One of the qualities of ESP test concerns about authenticity. Passages or texts should be selected from authentic text types such as article, experiment, technical and basic science text. These could help promote a positive affective response to the test task and can thus help test takers perform at their best.
2. Level of text difficulty	The three passages should be selected based on the engineering professor's judgement that they are appropriate level for first year graduate engineering students. In addition, a provision of text readability indices for all passages can help the expert make a selection.
3. Level of text specificity	The three passages should be selected with a semi-level of ESP text specificity which concerns a core of field specific vocabulary explaining in the text, non-technical and sub-technical words mostly used in the text, required knowledge of subject specific concepts in comprehension of the text and absence of highly academic texts (see a full definition of semi-level of text specificity on page 11).

It is noted that all selected passages included in the E-ERT were not modified or adapted. Then originality and authenticity of the text were sustained.

In terms of reading passage length, based on the UCLA test, passages from 600 to 800 words in length were selected and students were allowed 2 hours and 30 minutes to complete the test. In this study, each was approximately about 500 to 650 words which was considered adequate for testing engineering reading comprehension and fitted to the time allotment (90 minutes).

Concerning the difficulty level of passages, text readability indices such as Flesch-Kincaid Reading Ease, Flesch-Kincaid Grade Level and Gunning-Fog Index were used to determine the level of difficulty of the passages. These kinds of indices are an indicator of how easy or difficult a text is to read. Child (2006) describes each type of index as follows:

Flesch-Kincaid reading ease measures readability between 0-100. Ideally, the text should be around the 60 to 80 mark on this scale. The nearer 100 the text scores, the easier it is to read. The Flesch score is usually relatively low for technical documentation.

Flesch-Kincaid grade level gives a number that corresponds to the grade a person will need to have reached to understand it. For example, a Grade level score of 8 means that an eighth grader will understand the text. Ideally, the text should be around the 6 to 7 mark on this scale. The lower the score, the more readable the text.

Gunning-Fog index is a measure of text readability. It represents the approximate reading age of the text - the age someone will need to be to understand what they are reading. Ideally, the text should be between 11 and 15 on this scale. The lower the score, the more readable the text. Any number returned over the value of 22 can be taken to be just 22, and is roughly equivalent to graduate level.

In this study, all passages were chosen as they were considered the equivalent of graduate level texts. Therefore, it could be assured that the levels of text difficulty for all passages were more or less the same. Table 3.2 illustrates the text readability indices of all passages in this study.

Table 3.2: Text Readability Indices of All Passages in this Study

Topics / Indices	Building	Mechanics	Experiment
Flesch-Kincaid Reading Ease	41	31	58
Flesch-Kincaid grade level	13	13	12
Gunning-Fog index	20	21	21

The draft of E-ERT which was later used in the pilot study is presented in Appendix D. Table 3.3 presents the objectives and numbers of the test items.

Table 3.3: Objectives and Number of Test Items

Objectives	Item number
Part I a)	
1) Recognize the meanings of sub-technical terms used in engineering.	1-10
Part I (b)	11-20
2) Identify the meaning of non-technical vocabulary in context	
Part II	
3) Skim the texts for main ideas.	21,29,37
4) Scan the texts for specific information.	22,23,24,30,31,32,33,38,39, 40,43
5) Draw inferences based on information in the text.	25,26,34,41,44,45
6) Identify the facts based on engineering background knowledge.	27,28,35,36,42

After the draft version of the test had been developed, a priori validation was carried out by having 3 lecturers and 2 engineering experts evaluate the test. The evaluation form is provided in Appendix E. H, M and L were used to identify the degree to which the item (question) measured the ability indicated in the objectives.

H = High degree of congruence with the objective

M = Moderate degree of congruence with the objective

L = Low degree of congruence with the objective

The priori instrument validation result is shown in Appendix F. It shows that all except questions no.8, 10, 11, 15, 16, 24, 30 and 34 obtained the degree of congruence with the objectives less than 75%. Regarding the appropriateness of the content, all raters rated 'yes' for all parts. Similar to the overall evaluation, every rater rated 'yes' for all except for the question concerning time allotment. One rater commented that ninety minutes provided might not be enough since the texts were lengthy. Consequently, the data as well as the comments and suggestions from the lecturers and the engineering experts were used in modifying the test. The test was then revised and used for the pilot study.

The purpose of the pilot study was to test the research instruments and the procedures of test administration for the improvement of the main study. The subjects who participated in the pilot study were excluded from the main study. The researcher applied the Classical Test Item Analysis program, Version 8 (2007) initially developed by Sukamolson to conduct the item analysis. The program suggests .20-.80 for the item difficulty index, .20 or more for the item discrimination index and for point-biserial correlation. Table 3.4 presents the reliability estimate and item analysis indices calculated for the data in the pilot study.

Table 3.4: Reliability Estimate and Item Analysis Indices for Pilot Study

Description	Data
Reliability estimate (KR-20)	.861
Difficulty index	.541
Discrimination index	.449
Point-biserial correlation	.369

The reliability estimate used in the study was Kuder-Richardson 20 (KR-20) as it is considered to be the appropriate index of test reliability for multiple-choice examinations (Tulane University, 2006). The KR-20 is a measure of internal consistency reliability which accounts for the number of test items, the students' performance on every test item and the variance for the set of students' test scores. The index ranges from 0.00 to 1.00. A value that is close to 1.00 is desirable, reflecting that the test is measuring what it intends to measure. The recommended level for the reliability estimate of scores is at least .70 (Fraenkel and Wallen, 2000). The calculated KR-20 of the E-ERT in the pilot study was .861.

Item Analysis

Item analysis is an important phase in the development of a test. In this phase, statistical methods are used to identify any test items that are not working well. The two most common statistics reported in an item analysis are the item difficulty index and the item discrimination index (Brown, 2005).

Item Difficulty Index

The item difficulty index is a measure of the proportion of test takers who answer the item correctly. For this reason, it is frequently called the *p-value*. It can range between 0.0 and 1.0, with a higher value indicating that a greater proportion of test takers responded to the item correctly, and it is thus an easier item. In the pilot study of the E-ERT, the minimum and maximum item difficulty indices were .00 and .90 respectively. The mean of the item difficulty index was .541 which falls in the recommended range of .20 - .80.

Item Discrimination Index

The item discrimination index is a measure of how well an item is able to distinguish between test takers who are knowledgeable and those who are not. The Discrimination Index (d) is computed using the performance of the equally-sized high and low scoring groups on the test. The range of this index is +1 to -1. The recommended discrimination index value is the level which is .20 or more (Professional Testing, 2005). A discrimination index value below 0.0 suggests that an item is discriminating negatively which means the most knowledgeable test takers are getting the item wrong and the least knowledgeable test takers are getting the item right. In the pilot study, the mean of the item discrimination index (d) was .449 which is higher than the benchmarked value of .20, suggesting the adequate discriminating index of the test.

Another type of discrimination index which is quite common is point-biserial correlation. The point-biserial correlation looks at the relationship between an examinee's performance on the given item (correct or incorrect) and the examinee's score on the overall test. The recommended level is higher than .20. The statistical analysis shows that the mean of the point-biserial correlation coefficient in the pilot study was .369, indicating an acceptable level of discrimination index. The information of the item analysis of the pilot study is presented in Appendix G.

Item Review

An item review was carried out. The test used in the pilot study consisted of 45 items, allowing 5 items to be dropped out for the final 40 items in the main study. The criteria used in considering which items were to be discarded are:

1. Items which are very easy or very difficult were then deleted. Therefore, items no. 16 and 24 were taken out.
2. Items with no or a negative discriminating effect are to be dropped out. Therefore, items no. 26, 36 and 43 were taken out from the test.

However, some items that did not have discriminating value higher than .20 or did not fall in the suggested difficulty index of .20-.80 were retained in the test because taking them out could affect the balance of the abilities intended for measuring in the test. In addition, some items with distractors that had no responses or with positive discrimination index and positive point-biserial index were revised. Furthermore, the sample size of 40 subjects in the pilot study was not a substantial number that could generate completely stable results. Item analysis was, therefore, used as a guideline in the item review.

After the item review, the final version of the E-ERT was obtained and is presented in Appendix H.

To put in a nutshell, the test development of the E-ERT had gone through several stages. It started from defining the TLU domain and test construct and ended with obtaining the final version of the E-ERT for the main study. The following figure presents a summary flowchart of the E-ERT development.

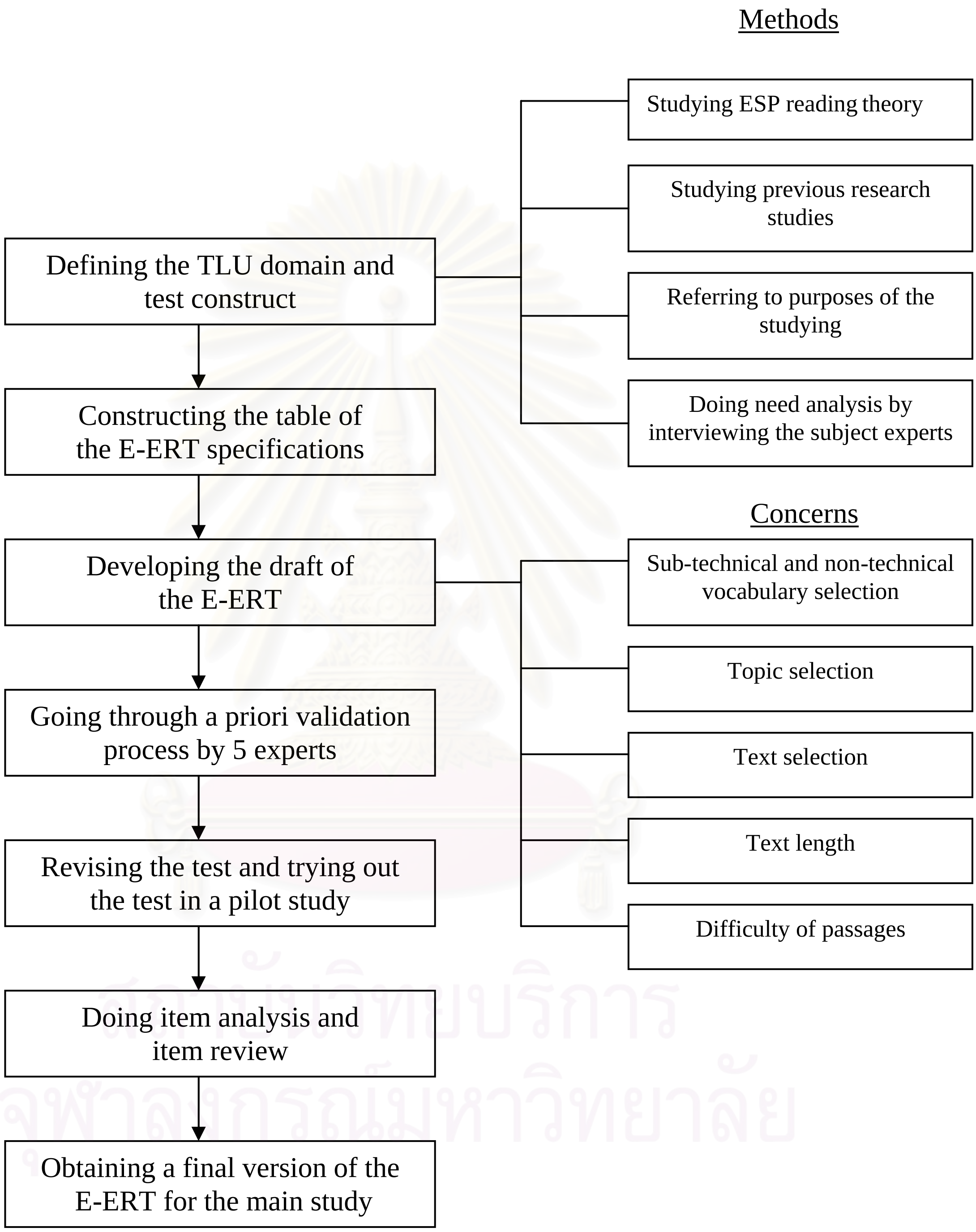


Figure 3.3: A summary flowchart of the E-ERT development

3.2.2 The KMUTT English placement test currently served as a validated in-house English test

This English test is currently used for placing students into appropriate groups either in Remedial English course for graduate students (a course for those who gain scores lower 50) or the In-sessional English course for graduate students (a course for those who gain scores higher than 50 but not exceeding 75) at KMUTT. It consists of two main parts: grammar and reading comprehension. Each part contains fifty multiple-choice items. In this study, this test was used to classify students into proficient and non-proficient groups. The criteria were set at or above 1 standard deviation for the high language ability group and at or above -1 standard deviation for the low language ability group. The reliability coefficient of the test was .774.

3.2.3 Introspective semi-structured interviews

To explore test taking strategies of the test takers in answering each kind of test items, introspective semi-structured interviews were conducted with all test takers in the pilot study and forty test takers (ten students from each subgroup) from the main study were randomly selected to be interviewed immediately after taking the test. Therefore, their memory on the test was still fresh. The questions in the interviews were delivered to the test takers concerning the test taking strategies used in answering each kind of test item i.e. sub-technical terms, non-technical vocabulary, main ideas, specific details, inferences, and engineering facts. According to the literature review in Chapter 2, a composite list of test taking strategies used in a multiple choice reading comprehension test suggested by Cohen (1998) was applied for the interviews. Those strategies in the list were grouped for interview questions according to the appropriate use for each type of test items. Some additional test taking strategies came up from the interviews in the pilot study. The researcher applied these in the main study. The interviews in the pilot study were conducted as a guideline for the main study. Some questions were later added to help students clarify their ideas and to improve the interviews in the main study. The interview questions are presented in Appendix I.

The interviews were conducted with an individual student or in a small group of two or three students. To avoid misinterpretation of data, students were given the opportunity to express themselves in Thai. The average time for each interview was 15-20 minutes. All interviews were audio taped with the students' consent.

3.2.4 A 4-Likert-scale questionnaire for test takers

To check the test takers' background knowledge and to collect data on their attitude towards the E-ERT, a questionnaire was developed. The questionnaire consisted of 2 parts.

Part 1: Checking test takers' background knowledge concerning topic familiarity

Part 2: The test takers' attitude towards the test

There were 11 areas of concern about which the test takers were asked to give responses. These were the key characteristics of the E-ERT, such as clarity of instructions and presentation, the number of questions, time allotment, the level of difficulty, the perceived usefulness, and so on. At the end of Part 2, there was a space provided for the test takers to give comments on the E-ERT's merits and areas needing improvement.

A 4-point attitude scale was employed in the questionnaire. The four options, '1) strongly disagree, 2) disagree, 3) agree and 4) strongly agree' were used to minimize the 'central tendency bias' by avoiding the middle option of "neither agree nor disagree". Central tendency bias occurs when respondents try to avoid using extreme response categories (Wikimedia, 2006). An even-point such as a 4-point or 6-point scale requires the respondents to exercise their discretion and can reduce the chance of respondents simply giving neutral responses without consulting the questions asked. In this study, the criteria were set so that the mean of attitude scale was greater than 2.5 points from the 4-point scale. This criterion was based on a previous study conducted by Nisa Vongpadungkiat (2006) who conducted a similar study on test takers' attitudes towards the ESP test.

Before administering the questionnaire, all content was validated and received a high rating from the three experts in terms of congruence with the objectives and the appropriateness of questions to the key characteristics of the E-ERT. The questionnaire is presented in Appendices J and K. The appropriate reliability estimate for the attitude questionnaire is the Cronbach Alpha Reliability Estimate as it can account for a weighted response, i.e. 1 to 4 in this case. The Cronbach Alpha Reliability Estimate computed for the questionnaire from the pilot study was .754, indicating an acceptable level of reliability estimate.

3.3 Data Collection

After the development and validation of the instrument, the pilot study was conducted during the period September 18-26, 2007. Forty first-year graduate students from the Faculty of Engineering and non-engineering Faculties were grouped into high and low language ability. Data was collected from the E-ERT, introspective interviews and questionnaires. The researcher firstly explained the objectives and significance of the study and reasons for using the tools. Then, the E-ERT was administered outside class hours, but under controlled classroom conditions. The engineering students and non-engineering students were tested separately, but under very similar conditions. They were allowed approximately ninety minutes to finish the E-ERT. After taking the test, they were asked to complete a questionnaire to provide information about their background knowledge and their attitudes towards the E-ERT. Finally, all were interviewed about the test taking strategies they used to answer different types of test items.

For the main study, a similar procedure to the pilot study was followed. The main study was conducted through 14 different test administrations during the period January 14 – February 8, 2008. One hundred and twenty students participated in the test. The test was administered outside the class, but under controlled classroom conditions. The test paper in the main study contained 40 questions modified from the pilot study results. The test time was 90 minutes. After taking the test, the students were asked to complete a questionnaire. Then, forty subjects (10 from high language ability engineering group, 10 from low language ability engineering group, 10 from high language ability non-engineering group, and 10 from low language ability non-engineering group) were randomly selected for the interviews. The 40 subjects in the pilot study were excluded from the main study.

3.4 Data Analysis

To answer the research questions, the following data analysis procedures were employed.

1. With regard to the first research question, “Is there any significant interaction effect between language ability and engineering background knowledge on the ESP reading ability? If there is, what is its effect size?” the following analysis was conducted.

The means and marginals means of the sample was obtained for two-way ANOVA to investigate the validity of the test. After that, two-way ANOVA was carried out to test if, on average, the means of the two groups or levels of subjects were significantly different. This could reveal the interaction effects between language ability and engineering background knowledge on the E-ERT scores. In addition, partial Eta squared was used to measure the effect sizes of the treatment.

2. The second research question was “Is there any significant difference between students with high language ability and those with low language ability in doing the E-ERT? If there is, what is its effect size?"

In response to this question, a two-way ANOVA was carried out concerning the main effect of language ability. Furthermore, partial Eta squared was computed to measure the effect size of the treatment.

3. The third research question was “Does background knowledge affect ESP reading performance? If it does, what is its effect size?”

In response, a two-way ANOVA was carried out to observe the main effect of background knowledge. Furthermore, partial Eta squared was computed to measure the effect size of the treatment.

4. For the fourth research question, "How do the test takers in the four subgroups use test taking strategies in order to answer different types of test items?"

In response, introspective interviews with 40 representatives of the whole population (10 from high language ability engineering group, 10 from low language ability engineering group, 10 from high language ability non-engineering group, and 10 from low language ability non-engineering group) were conducted to investigate their test taking strategies. The data obtained was listed, categorized, tabulated and presented using frequencies.

5. The last research question was, "What are test takers' attitudes towards the E-ERT?"

In response, descriptive statistics i.e. mean score and grand mean score of the attitude scale were computed to investigate the attitudes of the test takers towards the E-ERT.

The significant level for all statistical tests was set at 0.05. In addition, important basic assumptions of ANOVA such as normality, homogeneity of variances and absence of outliers were checked if they met the requirements. The details on all statistical procedures used in this study are presented in Appendix L.

3.5 Checking the Assumptions of Two-way Analysis of Variance (ANOVA)

With reference to Appendix L from pages 156 -157, the following assumptions had to be met before statistical analysis was conducted. Each of them is discussed in the following section based on the data collected.

1. Normality of sampling distributions

The distribution of data could be checked using Normal Probability Plot or Normal Q-Q plots under the SPSS function Analyze/Descriptive Statistics/Explore. The pattern of dots close to the diagonal line of expected values indicates a normally distributed data (Penkae Siriwan, 2003). The Normal Q-Q plots for all variables are presented in Appendix O. The dots in all plots are close to the diagonal lines, indicating that the data has a normal distribution. Therefore, the assumption of having the same underlying distribution for all variables is met.

2. Homogeneity of variance

Homogeneity of variance could be tested by using Levene’s test. It can be calculated by using SPSS program. The check on this assumption was presented in Appendix O. In Levene’s table, focusing on “Based on Mean” row, the observed value is higher than the set value (.05), indicating that variances are equal. Then, the assumption of homogeneity of variance is met.

3. Absence of outliers

Outliers can be detected by boxplot. This was carried out and is presented in Appendix O. In a boxplot, the median of the dataset is indicated by the black center line. The red box or what is known as Inter-Quartile Range (IQR) represents the middle 50% of the dataset. The lines extending from the upper and lower line of IQR are the extreme values that are within 1.5 times IQR. Any points that go beyond the lines are outliers and are generally represented by asterisks. In Appendix O, no asterisk is presented, indicating that there is no outlier in the data.

Summary

Chapter three presents the research methodology of the study. The data of the population and sample are presented. The procedures employed in the development of the research instruments are described. The steps taken in data collection and data analysis are also illustrated. Chapter four presents the findings of the study and the discussions of the results.

CHAPTER IV

RESULTS AND DISCUSSIONS

This chapter presents the findings of the study and discussion of the results. It is divided into five main parts. The first part deals with the descriptive statistics of the data collected in the main study. The second part concerns two-way ANOVA and partial Eta squared analysis to answer the first to third research questions. The third part presents frequencies that are used to answer the fourth research question. The answer to the last research question dealing with attitude scales of the test takers is reported in the last part. Discussions based on research findings are made at the end of each part.

4.1 Descriptive Statistics

Descriptive statistics of the E-ERT scores (dependent variable) obtained from the main study are computed. The mean, median, standard deviation, minimum value, maximum value and range are listed in Table 4.1.

Table 4.1: Descriptive Statistics of Main Study

Mean	21.30
Median	20
SD	6.373
Minimum	9
Maximum	36
Range	27

N = 120

From Table 4.1, by the mean of the E-ERT (21.30 out of 40), it can be seen that slightly more than 50% of the students could do the test, which means the test is not so difficult. The median is 20.00. It is noteworthy that the mean and median are close to each other, which is one of the characteristics of data with a normal distribution (Sirichai Pongwichai, 2006). The check on the assumption of normal distribution of data is provided in Appendix N. It indicates that data of all variables are normally distributed. According to Bachman (2004), two-third or 68% of the scores would fall in the range of ± 1 standard deviation for data with a normal distribution. The standard deviation of the E-ERT scores is 6.373, suggesting that 68% of the test scores falls in the range of 21.30 ± 6.373 or 14.927 to 27.673. The range is 27 with the maximum and minimum scores of 36 and 9 respectively. There is no missing value in the study. The total number of the subjects was 120. The reliability estimate and item analysis indices for the E-ERT in the main study were calculated. Table 4.2 illustrates the results.

Table 4.2: Reliability Estimate and Item Analysis Indices for Main Study

Description	Data
Reliability estimate (KR-20)	.800
Difficulty index	.534
Discrimination index	.395
Point-biserial correlation	.334

The reliability estimate (KR-20) is .800, meeting the required level of at least .70 (Fraenkel and Wallen, 2000) and indicating rather high degree of reliability. For the item difficulty index, the recommended level mentioned in the previous section is between 0.20-0.80. The mean of the item difficulty index of .534 denotes the appropriate difficulty level of the E-ERT (not too difficult or too easy). The mean of the item discrimination index of .395, indicating reasonably good items (Ebel, 1979 cited in Brown, 2005) is higher than the generally required level of .20 or more. The mean of the point-biserial correlation which is another indicator of the discrimination effect is .334. Even though the magnitude of reliability estimate and all indices of item analysis slightly decrease from the pilot study in which the KR-20, the mean of difficulty index, discrimination index and point-biserial correlation were .861, .541, .449 and .369 respectively, it is noteworthy that those magnitudes and indices in the main study still meet the requirement.

4.2 Results and Discussion

The results of the main study are presented based on five research questions mentioned in Chapter 1 corresponding to four research hypotheses. It is noted that basic assumptions of ANOVA were tested before doing data analysis. It was found that normality, homogeneity of variances and absence of outliers met the assumptions (see Appendix O).

Research question 1: Is there any significant interaction effect between language ability and background knowledge on the ESP reading ability? If there is, what is its effect size?

Hypothesis 1: There is a significant interaction effect between language ability and engineering background knowledge on the ESP reading ability at the .05 level.

Statistical hypotheses: ($H_{1.1:}$ μ High w/ Eng $\neq$ μ High w/o Eng)

($H_{1.2}$: μ Low w/ Eng $\neq$ μ Low w/o Eng)

($H_{1.3}$: μ High w/ Eng $\neq$ μ Low w/ Eng)

($H_{1.4}$: μ High w/o Eng $\neq$ μ Low w/o Eng)

As a prelude to other statistical tests, the sample means and marginals are shown in Figure 4.1 for two-way ANOVA design used here to investigate the construct validity of the test.

Major	***Ability*** *High language ability*	*Low language ability*	
Engineering	$\overline{X}$ = 28.3 n = 30	$\overline{X}$ = 18.8 n = 30	$\overline{X}$ = 23.6 n = 60
Non-engineering	$\overline{X}$ = 23.0 n = 30	$\overline{X}$ = 15.1 n = 30	$\overline{X}$ = 19.1 n = 60
	$\overline{X}$ = 25.7 n = 60	$\overline{X}$ = 16.9 n = 60	Grand mean = 21.3 N = 120

Figure 4.1 : ANOVA sample means and marginals

The data shows that the mean scores for both groups of high language ability students were higher than those for low language ability students and the average performance for engineering students was higher than that for non-engineering students. This confirms that the test had considerable construct validity as mean scores are congruent with performance by high and low language ability as well as engineering and non-engineering backgrounds.

As regards the effects of language ability and engineering background knowledge on the E-ERT scores, the effects of ability (refers to language ability) and major (refers to engineering and non-engineering background) are presented in Table 4.3 below.

Table 4.3: Results of Two-way ANOVA

Source	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3	971.089	58.672	.000	.603
Intercept	1	54442.800	3289.367	.000	.966
ABILITY	1	2288.133	138.246	.000	.544
MAJOR	1	607.500	36.704	.000	.240
ABILITY * MAJOR	1	17.633	1.065	.304	.009
Error	116	16.551			
Total	120				

$p \leq .05$

From Table 4.3, by focusing on an interaction effect, the obtained F (1.065) is smaller than the critical value (3.92). This means that there was no significant interaction effect between language ability and engineering background knowledge on the E-ERT scores, $F(1, 116) = 1.065$, $p > 0.05$, $\eta_p^2 = 0.009$. In other words, language ability differences could be seen in both engineering and non-engineering groups. Therefore, the alternative hypothesis that there is a significant interaction effect between the two variables on the E-ERT scores was rejected. The following is a discussion based on this research finding.

Discussion for Research Question 1

The finding relating to the first research question indicates that there was no significant interaction effect between language ability and engineering background knowledge on the E-ERT scores. The reasons which probably underlie this finding are as follows:

First, language ability based on the criteria set in this study may have a weak relationship with general engineering background knowledge. The criteria was set at or above +1 standard deviation for high language ability students and at or below -1 standard deviation for low language ability students. -1S.D. - +1S.D. may not constitute a large enough gap to distinguish high and low language ability students.

Based on this criterion, it is possible to conclude that students might not have been as proficient as expected or had very low language ability. As a result, no interaction effect between language ability and engineering background knowledge could be found in this study.

Second, the E-ERT may not differentiate real engineering background knowledge since the texts used in the study were not highly specific or hard ESP. From the beginning, the test was developed with a semi-level of text specificity to avoid bias between engineering and non-engineering students. Thus, this may have affected the mean score of the engineering group, which was not greatly different from that of the non-engineering group. The mean of the engineering students was 23.6 and that of the non-engineering students was 19.1. This finding agrees with Douglas (2000) who suggested that the degree of specificity plays a role in students' reading comprehension. If passages were sufficiently and highly specific, students could do better at a test in their own subject area. This is a reason why the mean of the engineering students was not as high as expected.

Third, the comprehension of specific texts might be compensated by transferring test taking strategies to ESP reading. Some students, particularly proficient ones may have transferred test taking or reading strategies of general English reading to ESP reading. This conclusion is supported by data in Tables 4.4 - 4.10 on pages 74 - 78. The tables present frequencies of test taking strategies used in answering different kinds of item types. The results show that all groups of students tried to make use of several strategies and used them with different levels of frequency in order to arrive at answers for each test item type. These strategies could have helped facilitate the comprehension of a text requiring subject-specific knowledge, even for non-engineering students. As a result, in ESP reading, the assessed difference between engineering and non-engineering students was not clear.

Research question 2: Is there any significant difference between students with high language ability and students with low language ability in doing the Engineering-English Reading Test (E-ERT)? If there is, what is its effect size?

Hypothesis 2: Students with high language ability performed significantly better in the E-ERT than those with low language ability at the .05 level.

Statistical hypothesis: (H_2: μ w/ High > μ w/ Low)

With regard to main effect of *ability* in Table 4.3 which describes Two-way ANOVA, the obtained F (138.246) is greater than the critical value (3.92). This means that there was a significant main effect between students with high language ability and low language ability, F (1, 116) = 138.246, $p < 0.05$, $\eta_p^2 = .544$. The high language ability group outperformed the low language ability group. The mean of the former was 25.7 and that of the latter was 16.9.

According to Hopkins (2002), the effect size is large (0.544) which means high language ability students and low language ability students were very different in performing the E-ERT. The following is a discussion based on this research finding.

Discussion for Research Question 2

The second research question concerning language ability was found to support the language factor that can affect ESP test scores. The findings reveal that the test significantly distinguishes between high language ability students and low language ability students. However, the two high language ability groups, even non-engineering students performed better than the two low language ability groups, even engineering students. This may indicate that overall knowledge of English has a stronger effect to success on this ESP test (as its effect size is huge) than knowledge of the special English of engineering. This result agrees with that of Clapham (1996) and Tan (1990), who found that comprehension of a discipline-related text could be predicted both by knowledge of the subject area and by language level, but that language level was the better predictor. One possible explanation for this result, as forwarded by Clapham (2000) to explain the gap between high and low level language ability is that low language ability students cannot take advantage of their background knowledge because they are too concerned with bottom-up skills such as decoding the text. In contrast, high proficiency students are able to make maximum use of their linguistic skills, so that they do not have to rely so heavily on their background knowledge. In other words, readers with high language ability were so proficient that their language knowledge could compensate for a lack of background knowledge.

Research question 3: Does background knowledge affect ESP reading performance? If it does, what is its effect size?

Hypothesis 3: Students with engineering background knowledge can do the E-ERT better than those without background knowledge at the .05 level.

Statistical hypothesis: (H_3: μ w/ Eng > μ w/o Eng)

To confirm this hypothesis, Table 4.3 presenting Two-way ANOVA results is referred to again. With regard to the main effect of *major*, the obtained F (36.704) is greater than the critical value (3.92). This means that there was a significant main effect between students with engineering background knowledge and those without. $F (1, 116) = 36.704$, $p < 0.05$, $\eta_p^2 = 0.240$. The engineering students performed better than the non-engineering students. The mean of the former was 23.6 and that of the latter was 19.1.

The partial Eta squared value (0.240) indicates that this is a relatively medium effect (Hopkins, 2002), which means engineering and non-engineering students are quite different in performing the E-ERT. The following is a discussion based on this research finding.

Discussion for Research Question 3

The third research question concerns background knowledge, and the findings show that the test significantly distinguishes between engineering students and non-engineering students even though there was not much mean difference between the two groups. The reasons underlying this finding might be attributed to the level of text specificity and the amount of test takers' engineering background knowledge.

The level of text specificity

The Engineering-English Reading test was intentionally developed with a semi level of text specificity to avoid bias between engineering and non-engineering groups. The texts were also selected from several sources such as the Internet, magazines, textbooks and encyclopedia so the language used in the texts was not highly specialized. Even though the test contained subject-specific vocabulary, the vocabulary was explained in the text. So, the E-ERT could facilitate reading comprehension for both engineering and non-engineering students. However, because of the semi specificity of the E-ERT, the mean score of the engineering group was not greatly different from that of the non-engineering group. This suggests that the degree of specificity plays a role in students' reading comprehension (Douglas, 2000). If passages were sufficiently and highly specific, engineering students could do better at a test in their own subject area. This also supports Clapham's (1996) findings regarding the stronger effect on test performance of the more field specific texts. Highly specific texts would have a significant background knowledge effect even among the most highly proficient test takers. However, further research directed specifically to this issue is needed.

The amount of test takers' engineering background knowledge.

Concerning the amount of engineering background knowledge for both engineering and non-engineering groups, it was proved from the test takers' questionnaire that all engineering students with high and low language abilities had enough knowledge of engineering. Thus high language ability engineering students did not experience much difficulty answering the E-ERT since their language ability enabled effective use of their background knowledge and they were familiar with the test content. However, low language ability engineering students, seemed to have difficulty doing the test, possibly because their lower language proficiency prevented them from making full use of their engineering background knowledge. Regarding the non-engineering group, the questionnaire revealed that some high language ability students (i.e. those from the faculties of science, and energy and materials) might have experience in reading science and technology texts or they might be familiar with some terms which are similar to engineering terms. As a result, they could benefit more from using background knowledge than other non-engineering students from other faculties or schools.

Research question 4: How do the test takers in the four subgroups use test taking strategies in order to answer different types of test items?

To answer this research question, qualitative analysis was conducted. The data from the interviews of 40 test takers (10 from each group) was listed, categorized, tabulated and presented in form of frequency counts. The results are presented by comparing test taking strategies used among the four subgroups in answering each kind of test items namely sub-technical term items, non-technical term items, main idea items, specific detail items, inference items and fact items. It is noted that some test takers might use several strategies in answering one test item and one test taker might use different strategies from others. The four subgroups referred in the findings are as follows:

HE = High language ability engineering students
HNE = High language ability non-engineering students
LE = Low language ability engineering students
LNE = Low language ability non-engineering students

Table 4.4: Frequencies of Test Taking Strategies Used in Answering Sub-technical Term Items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Translating a whole sentence to get an idea	8	4	7	7	**26**
2. Focusing on key words or context clues in a sentence	8	8	2	8	**26**
3. Being familiar with the terms, then answering without reading options	3	0	1	0	**4**
4. Eliminating inconsistent options	4	0	3	1	**8**
5. Eliminating unknown options	0	1	3	1	**5**
6. Making use of information from other items	0	2	0	1	**3**
7. Choosing an answer with familiarity on it	0	0	1	5	**6**
Total	23	15	17	23	**78**

Table 4.4 presents frequencies of the test taking strategies used in answering sub-technical term items among the four subgroups. It illustrates that the most frequent strategies used for all groups are "translating a sentence" and "focusing on key words". When comparing among four groups, high language ability engineering mostly used these two strategies while high language ability non-engineering used translating strategy least. Next, Table 4.5 presents frequencies of test taking strategies used in answering non-technical term items.

Table 4.5: Frequencies of Test Taking Strategies Used in Answering Non-technical Term Items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Translating the word itself and then choosing the nearest synonym	4	9	5	5	**23**
2. Focusing on key words or context clues around	9	10	9	9	**37**
3. Recognizing the meaning of the words tested	8	8	2	0	**18**
4. Guessing from the given options i.e. grouping words	2	0	3	1	**6**
5. Having a quick look at all words tested to gather information for overall idea of the text	2	0	0	0	**2**
6. Identifying word form of options and then choose the one parallel with the word tested	0	1	0	0	**1**
Total	25	28	19	15	**87**

Table 4.5 illustrates that “focusing on key words or context clues around” is the most frequent strategy used in answering non-technical term items for all four groups. Among four groups, high language ability non-engineering tends to mostly use “translating the word itself and then choosing the nearest synonym” and “focusing on key words” strategies. In addition, it is obvious that both groups of high language ability students are better in “recognizing the meaning of the tested word” than two groups of low language ability. Table 4.6 presents frequencies of test taking strategies used in answering main idea items.

Table 4.6: Frequencies of Test Taking Strategies Used in Answering Main Idea Items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Skimming for general idea of the text	10	10	10	8	**38**
2. Reading every word	0	0	0	2	**2**
3. Reading only the first and the last paragraphs	2	3	0	2	**7**
4. Looking for the frequent words used in a passage	1	0	0	0	**1**
5. Working on main idea item before doing other items	8	4	5	10	**27**
6. Working on main idea item after doing other items	2	6	5	0	**13**
Total	23	23	20	22	**88**

Table 4.6 shows that "skimming for general idea of the text" is an outstanding strategy which is mostly used in answering main idea items for all groups. Among four groups, high language ability engineering and non-engineering, and low language ability engineering make full use for skimming for main idea. In addition, high language ability engineering and low language ability non-engineering prefer to work on main idea items before doing other types of items. Table 4.7 presents frequencies of test taking strategies used in answering specific detail items.

Table 4.7: Frequencies of Test Taking Strategies Used in Answering Specific Detail Items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Scanning	10	10	10	8	**38**
2. Focusing on key words in a question, then reading for details in a specified paragraph	9	10	9	10	**38**
3. Eliminate options that are not mentioned in a passage	2	0	0	2	**4**
4. Using background knowledge	1	0	0	0	**1**
Total	22	20	19	20	**81**

Table 4.7 illustrates that for all groups, “scanning” and “focusing on key words in a question” are two strategies that were mainly used in answering specific details items. However, the group of high language ability, non-engineering students made full use of both strategies. Table 4.8 presents frequencies of test taking strategies used in answering inference items.

Table 4.8: Frequencies of Test Taking Strategies Used in Answering Inference Items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Reading for details, then analyzing the information before answering	10	9	6	5	**30**
2. Eliminating inconsistent and incredible options	9	5	3	5	**22**
3. Returning to the passage and check with the options	2	6	6	3	**17**
Total	21	20	15	13	**69**

Table 4.8 indicates that “reading for details, then analyzing the information before answering” is the most frequent strategy used in answering inference items. High language ability engineering students made full use of this technique. In addition, they were better in taking advantage of “eliminating inconsistent and incredible options” than the other three groups. Next, Table 4.9 presents frequencies of test taking strategies used in answering fact items.

Table 4.9: Frequencies of Test Taking Strategies Used in Answering Fact items

Strategies Used	HE (N=10)	HNE (N=10)	LE (N=10)	LNE (N=10)	Total (N=40)
1. Using engineering background knowledge	10	1	8	0	**19**
2. Using the world knowledge	7	6	5	5	**23**
3. Returning to the passage and check with the options	0	9	2	3	**14**
Total	17	16	15	8	**56**

Table 4.9 shows that “using world knowledge” is the most frequently used strategy in answering fact items. However, high language ability engineering

students benefited more from using this strategy than other students. In addition, both groups of engineering make effective use for using engineering background knowledge in answering fact items. The following is a discussion based on this research finding.

Discussion for Research Question 4

The findings concerning test taking strategies reveal that in general all the four subgroups employed the same test taking strategies in answering each kind of test item. It also shows that the strategies they used to answer confirm the abilities that are supposed to be measured in each test item type. This indicates that the E-ERT has construct validity.

However, the significance of the finding is the frequency of strategies used among the four groups. This depended on what type of test items the students were responding to. Table 18 presents frequencies of all test taking strategies used in answering all test item types.

Table 4.10: Frequencies of All Test Taking Strategies Used in Answering All Test Item types

Item types	HE	HNE	LE	LNE	Total
Sub-technical term items	23	15	17	23	**78**
Non-technical term items	25	28	19	15	**87**
Main idea items	23	23	20	22	**88**
Specific detail items	22	20	19	20	**81**
Inference items	21	20	15	13	**69**
Fact items	17	16	15	8	**56**
Total	**131**	**122**	**105**	**101**	**461**

Table 4.10 illustrates the overall picture, revealing that two groups with high language ability were better at using test taking strategies and used more of them than low language ability groups. This finding agrees with Prapphal's (1994) study that more proficient students were able to transfer some general English reading strategies to academic English reading. The reason underlying this finding is that

high language ability students with their proficient English, high degree of cognitive flexibility and their repertoire of test taking strategies allows them to employ more strategies in order to answer questions in the test. On the contrary, low language ability students with their non-proficient English may lack flexibility in choosing other test taking strategies and be less trained to employ the strategies at a given moment on a given task. According to Cohen (1998), some test takers may use a limited number of strategies but use them well for the most part. This may apply to low language ability students.

The next point that is drawn from the finding is that several test taking strategies are frequently used by engineering students, especially those with high language ability. This might be because the thinking process and comprehension strategies developed by this group of students facilitate more use of test taking strategies. Engineering students study in Applied Science then are taught how to comprehend processes, hypotheses, theories, experiments, to be logical and to think systematically. They are trained to apply a variety of thinking processes to solve problems. In a situation of taking a test, they may similarly use logical thinking as well as apply several strategies in order to arrive at satisfactory answers. In addition, this test was specifically developed for engineering students. Possibly, the context of familiar terminology and content would have evoked the use of learned strategies normally applied in response to such of test item types.

The last point from the finding is that the difference in frequencies of test taking strategies used depends on types of test items. By concentrating on the total frequencies for each type of test item, the finding indicates that the more specific the test item is, the fewer strategies the test takers would use. A closer examination of each kind of item type reveals that some broad questions such as “non-technical term items” and “main idea items” demand more frequent use of strategies than specific questions such as “fact items” and “inference items”.

Research question 5: What are test takers' attitudes towards the E-ERT?

Hypothesis 4: The test takers have positive attitudes towards the E-ERT.

(H_4: Mean of attitude scale > 2.5 points from the 4-point scale in the questionnaire)

The mean score of the attitude scale for each item and the grand mean score in the test takers' questionnaire were calculated to answer this research question. Table 4.11 shows the results.

Table 4.11: Results from the Test Takers' Questionnaire

Item	Statement	Scale				Mean
		4	3	2	1	
1	Instructions of the test are clear.	73	46	1	-	***3.60***
2	Typeface and size of letters in the test are appropriate.	75	42	3	-	***3.60***
3	Graph, layouts and pictures in the test are clear.	51	55	14	-	***3.31***
4	Graphs, layout and pictures in the test are appropriate.	36	67	17	-	***3.16***
5	The number of questions in the test is appropriate.	39	72	9	-	***3.25***
6	Time allotment for the test is appropriate.	34	80	6	-	***3.23***
7	Level of difficulty of the test is appropriate for graduate students.	20	91	9	-	***3.09***
8	I am satisfied with the test in general.	19	87	14	-	***3.04***
9	The test is useful for my English language learning and development.	27	71	22	-	***3.04***
10	The content of the test is similar to the English reading test for Engineering in real situations.	35	62	23	-	***3.10***
11	The format of the test is appropriate for English reading test for Engineering in real situations.	24	74	22	-	***3.02***
Grand mean score		***3.22***				

N = 120

Note**: 4** = Strongly agree, **3** = Agree, **2** = Disagree, **1** = Strongly disagree

Table 4.11 shows that the mean score for each item are all higher than 3, producing a grand mean score of 3.22. All means are greater than the set criterion of 2.5. Therefore, hypothesis 4 which states that the test takers have positive attitude towards the E-ERT is accepted. In addition, the Cronbach Alpha Reliability Estimate was .806, indicating a satisfactory level of reliability estimate.

At the end of the questionnaire, a section with an open-ended question asking for the comments on the merit and the area of improvement was provided. The following are the comments given by the test takers. When the same or similar comment is given by more than one person, the number in bracket is added to indicate the number of students giving that comment.

Merits:

1. The test is easy to understand and good to assess fundamental knowledge in English and Engineering. (8)
2. The amount of test items is just right. (7)
3. The test provides opportunity to practice English reading. (7)
4. The level of test difficulty is appropriate for graduate students. (6)
5. I have gained new knowledge on vocabulary and what I have read from passages. (4)
6. The test content and test format are good especially the colorful pictures that make the test more interesting. (4)
7. Pictures and graphs help illustrate ideas in the text. (4)
8. The test adequately covers varieties of test item types i.e. sub-technical terms, non-technical terms, main ideas, and specific details. (3)
9. Vocabulary and content in the test are consistent with fundamental knowledge in engineering. (3)
10. The test instructions are clear and detailed. (3)
11. The questions are clear and vocabulary generally used in the test is not too difficult. (2)
12. The test arranges in sequence from easy to difficult items. (2)
13. The test includes both easy and difficult items. (2)
14. The test in part I (a) is ideal for engineering students to test their engineering knowledge. (2)

15. Sub-technical terms used in the test are not too difficult and could be seen in general engineering texts.
16. The test is useful for developing English reading skills in engineering.
17. It is advantageous to have both sub-technical terms and non-technical terms in the test.
18. The test allows test takers to analyze vocabulary and passages.
19. The font size is right.
20. The given alternatives are good.
21. The time allotment is appropriate for low language ability students.

Areas of improvement:

1. The passages are lengthy. (9)
2. The test should cover content for engineering in different disciplines i.e. chemical engineering and computer engineering. (5)
3. Passage 3 is lengthy and not easy to understand due to difficulty to relate the pictures and content. (5)
4. Questions and content in the test are highly specific in engineering. It is difficult for non-engineering students though they understand the questions and contexts. (3)
5. Discussion part should be added in passage 3 to make it more understandable. (2)
6. The engineering experiment should be short. (2)
7. The test should have more items arranged in sequence from easy to difficult items.
8. The passages should be taken from text books because sentences are not complex and contain easier vocabulary.
9. The pictures in passage 1 should illustrate more information.
10. The researcher should study more engineering lab sheets in order to apply knowledge of these in a test.
11. Some passages are too difficult to understand and some items are too easy.

12. The test should have more item types (i.e. matching, or true/false) to make the E-ERT more interesting.
13. Color bar graphs and pictures are expensive and inappropriate in real situations of testing. Bar graphs with lines should be replaced to save cost.
14. Unlike passages 1 and 3, pictures should be added in passage 2.
15. In passage 2, more space should be inserted between lines to prevent dizziness.
16. The lengthy passage should be either reduced or the same length retained with the amount of test items decreased.
17. More interesting topics should be considered to include in the test.
18. Some passages contain technical terms so they are too difficult to understand.

However, it was observed that the majority of the test takers tended to skip this part. Only 56 out of the 120 test takers (46.7 %) answered the open-ended question. The following is a discussion based on this research finding.

Discussion for Research Question 5

The results indicate that the test takers have a positive attitude towards the E-ERT. The mean score for each question is clearly higher than 3 points, well above 2.5 points on the 4-point scale in the questionnaire, which is used as the benchmark in the hypothesis testing.

As Brown (2004b) suggests, students' motivation needs to be taken into consideration when planning studies involving tests, and in interpreting the scores of any test. From the researcher's observation, it was rewarding to see that the test takers were quite motivated to take the test after the researcher explained about the development, the significance and the use of the test as well as the opportunity to self-assess English language ability. Furthermore, when this information was communicated to the test takers, they were enthusiastic about taking the test and appeared ready to invest their best effort. In addition, after taking the test, they seemed eager to know their test scores, which was taken as an indication of their intention and willingness for test taking.

In this study, the strength of the E-ERT lies in the way the test was developed and informed by subject experts. The fact that test construct was based on interviews with ten engineering experts and previous research on engineering lexis in English and the development of ESP reading tests highlight its value for the parties concerned, particularly the test takers. More importantly, what was in the test takers' best interest was the opportunity for self-assessment of their English reading ability, particularly for engineering students.

This research question aims to find out whether the procedures followed in this study could motivate test takers and generate positive attitudes towards the test. The findings show that the steps employed could and did help to bring about the positive attitudes by the test takers towards the test, which in turn, reinforced the interpretation of the findings and test scores.

Summary

This chapter reports the results of the findings. Descriptive statistics for the data are presented. Two-way ANOVA and partial Eta squared were employed to answer the first three research questions. Frequency counts were conducted to reveal test taking strategies used among the test takers. The mean score of the attitude scale was calculated to indicate the test takers' attitudes towards the test. Each part ends with discussions based on the findings and the literature review.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Chapter Five presents the research summary and the summary of the findings in the first part. The conclusions including implications for language testing are given in the second part. Subsequently, the recommendations for future research are provided in the last part.

5.1 Research Summary

This study concerned the investigation of the two independent variables, language ability and engineering background knowledge on ESP reading ability. Furthermore, the test taking strategies of the test takers in response to different types of test items were explored and compared. In addition, test takers' attitudes towards the test were examined.

The focus of this study was the development of the Engineering English Reading Test (E-ERT) tailored for first year graduate students at King Mongkut's University of Technology Thonburi (KMUTT). At present, there are no tailored tests to meet the need for assessing English language proficiency in the engineering field at KMUTT. Graduate students lack the opportunity to assess and prepare themselves before continuing graduate-level study. The development of the E-ERT could help fulfill students' needs. Before the test was developed, the target language use (TLU) domain of the test was identified from ESP reading theory, previous research in ESP reading, the purposes of the study and interviews from ten engineering experts. The test went through a validation process and was piloted before it was employed in the test administration of the main study. Test takers' attitudes towards the E-ERT were examined after taking the test.

Two independent variables, language ability and specific purpose background knowledge were selected to study their effects on the E-ERT scores. The reasons for selecting these independent variables were as follows. First, these variables could have an important influence on both language use and test performance. Second, it was possible to design a language test in which these variables facilitated rather than impeded test takers' performance. Third, there were many inconclusive studies on the effects of language ability and background knowledge on the reading performance of students. It was therefore considered worthwhile to investigate these issues on language ability and background knowledge. Therefore, the effects of these two variables on test takers' ESP reading ability were investigated.

Investigation of test taking strategies of the test takers was another focus of this study. Test takers' strategies were investigated to discover successful strategies and to validate the test purpose. This provided a means of determining items for inclusion in the test as it helped to reveal what the test items were actually testing. Consequently, there were five research questions in this study:

1. Is there any significant interaction effect between language ability and background knowledge on the ESP reading ability? If there is, what is its effect size?
2. Is there any significant difference between students with high language ability and those with low language ability in doing the E-ERT? If there is, what is its effect size?
3. Does background knowledge affect ESP reading performance? If it does, what is its effect size?
4. How do the test takers in the four subgroups use test taking strategies in order to answer different types of test items?
5. What are test takers' attitudes towards the E-ERT?

The population was 359 first-year graduate students in a Master's degree program at the Faculty of Engineering at King Mongkut's University of Technology Thonburi (KMUTT). 236 students from other schools or faculties including liberal arts, industrial education and technology, energy environment and materials, science, and bioresource and technology were also part of the population and participated as the non-engineering group to ensure construct validity of the test and provide proof for one hypothesis of the study on background knowledge. The sample were 120 students from the population in semester 2, academic year 2007. Both groups of engineering and non-engineering students were classified into two groups of high and low language ability according to their English placement test scores.

Research instruments consisted of the E-ERT, the KMUTT English placement test, an introspective interview and a test takers' questionnaire. The E-ERT was used to assess students' ESP reading performance and to elicit the use of background knowledge of the students. The KMUTT English placement test was used to classify students into two groups of high and low language ability. The interviews were conducted to explore test taking strategies of the students in answering different kinds of test items. The questionnaire was used to check test takers' background knowledge and to collect data on test takers' attitude towards the test. Regarding statistical procedure, two-way ANOVA was carried out to observe the effects of the two independent variables: language ability and engineering background knowledge. In addition, partial Eta squared was used to measure the effect size of the treatment. For the study of test taking strategies, content analysis in the form of frequency counts was employed. To examine the test takers' attitudes towards the test, the mean score of the attitude scales and the grand mean were calculated.

5.2 Summary of the Findings

Concerning the first research question, the results obtained from two-way ANOVA indicated that there was no significant interaction effect between language ability and engineering background knowledge on the E-ERT scores, $F (1, 116) = 1.065$, $p > 0.05$, $\eta_p^2 = 0.009$. In other words, language ability differences were found both in engineering and non-engineering groups. Therefore, the hypothesis that there is significant interaction effect between the two variables on the E-ERT scores was not supported by the findings.

Regarding the second research question, two-way ANOVA revealed that there was a significant main effect between students with high language ability and low language ability, $F (1, 116) = 138.246$, $p < 0.05$, $\eta_p^2 = .544$. The high language ability group outperformed the low language ability group. The mean of the former was 25.7 and that of the latter was 16.9. The effect size was quite large (0.544), indicating that high language ability students performed very differently from low language ability students on the E-ERT test.

As for the third research question, there was a significant main effect between students with engineering background knowledge and those without, $F(1, 116) = 36.704$, $p < 0.05$, $\eta_p^2 = 0.240$. The engineering students performed significantly better than the non-engineering students. The mean of the former was 23.6 and that of the latter was 19.1. The partial Eta squared value (0.240) indicated a relatively medium effect. Engineering and non-engineering students were quite different in performance on the E-ERT.

For the fourth research question, findings obtained from the frequency counts reveal that in general, all the four subgroups employed the same test taking strategies to answer each kind of test item. It was also found that the strategies used corresponded with the abilities that each test item type purported to measure. However, the frequency of strategies used to respond to test items was different from that anticipated for the four groups. The frequency was associated with other factors and with type of test item. The two high language ability groups employed more test taking strategies than low language ability groups and were more successful at this. In addition, several test taking strategies were frequently used by engineering students especially the high language ability ones.

The last research finding reveals that the test takers had positive attitudes towards the E-ERT, using 2.5 points out of a 4-point scale on the test takers' questionnaire as a benchmark. The mean score computed from the test takers' responses was higher than 3 for all items, producing a grand mean score of 3.22.

5.3 Conclusions

Both language ability and specific purpose background knowledge are key features which distinguish ESP testing from general purpose testing. In addition, both of them play important roles and are key variables in ESP reading test performance. This study attempted to investigate the interaction effects of language ability and engineering background knowledge on ESP reading ability. It compared the ESP reading performance between high and low language ability groups, and explored the effect of background knowledge on ESP reading performance. In addition, test taking strategies of the test takers were investigated to validate the test purposes. Test takers' attitudes towards the developed test were also explored.

The findings indicated no significant interaction effect of language ability and engineering background knowledge on the E-ERT scores. However, each variable was found to be significantly different. The results provided additional empirical evidence to the research conducted by Alderson and Urquhart (1983, 1985a, and 1985b), Koh (1985) and Nieh (2000) who found that background knowledge plays an important role and has an effect on test scores. The findings, however, highlight that overall knowledge of English is more important for success on the E-ERT than knowledge of the special English of engineering. This finding was in line with those from the studies conducted by Tan (1990) and Clapham (1996) where comprehension of a discipline-related text was found to be predictable from both knowledge of the subject area and language level, with language level being the better predictor. As Clapham (1996) points out, the effect of background knowledge depends on the specificity of the reading passages. The more specific the text, the greater the effect of background knowledge. Since this study concerned engineering English, which contains a semi-level of text specificity, highly technical words were not used in the research instruments. As a result, the mean difference between engineering and non engineering students was small. However, background knowledge was found to be a significant factor and could contribute to ESP assessment.

As regards test taking strategies, generally speaking, all test takers from the four different subgroups employed the same test taking strategies to answer each kind of test item. However, the difference that could be found from the finding is in the frequency level of strategies used. High language ability engineering students seemed to employ more strategies and use them more frequently than other groups of students. This was related to what type of test items they were responding to, apart from several other factors. As Cohen (1998) points out, some factors such as test takers' particular cognitive style profile and degree of cognitive flexibility, their language ability, and their repertoire of test taking strategies affects the choice of strategies used in response to a given test task. As a result, individual test takers employ different strategies with different frequencies on a given test.

The findings on attitude reveal that the test takers had positive attitudes towards the E-ERT. The main reason for their interest in taking the test was for self-evaluation of English reading ability as well as to assess their fundamental knowledge in engineering. The endeavor to stimulate the motivation of the test takers and to generate positive attitudes towards the test was successful.

5.4 Implications of the Study

The implications from the findings of this study are presented as follows:

1. For test writers who would like to develop an ESP reading test for their institutions, several factors must be taken into consideration. First of all, the target language use (TLU) domain and constructs of ESP reading tests must be carefully defined as suggested by Douglas (2000). In this study, the TLU and constructs of the test were informed by several methods such as ESP reading theory and previous research studies, referring to the purposes of this research, as well as doing need analysis by interviewing the subject experts. The methods employed in this study could provide useful alternatives for ESP test developers. Secondly, the nature of the input to the test must be sufficiently field specific so that students can achieve optimum performance in their own subject area. Finally, the degree of specificity of input materials and the level of test difficulty must be worked out through consultations with subject experts.
2. The E-ERT received a high rating from the subject experts in terms of congruence with the objectives and the appropriateness of content. It can be useful for the parties concerned as outlined below.

   2.1 For graduate students, the test could be useful as an instrument for self-assessment of their actual language proficiency particularly in an engineering discipline.

   2.2 For universities, since the test was developed from data gathered in the field, the E-ERT could equip universities with target language use in English for Engineering, data which has hitherto not been available at KMUTT.

   2.3 For teachers who would like to use ESP tests to assess their students’ language ability, it confirms that there is a need to ensure that students have adequate knowledge in their subject area so as to optimize their use of knowledge associated with language ability for answering questions.

3. Concerning the washback effect on ESP teaching, since the findings reveal that language ability has a greater effect on test performance than background knowledge, instruction should then stress language knowledge which includes knowledge of vocabulary, morphology and syntax, cohesion and rhetorical structures used in the field. In addition, this study established that "focusing on key words" was the most frequent strategy used by all test takers. Thus instruction should be provided on vocabulary, especially with regard to identifying key words and recognizing terms used in the field, as these strategies are important for reading comprehension enhancement. Furthermore, engineering background knowledge is found to be a significant factor in reading comprehension and could contribute to ESP instruction. Therefore, activities, tasks and design of the tasks should be based on ESP engineering context so that all language components involved can be worked out in a meaningful way.
4. As regards theoretical contribution, the findings provide further insights into the roles of background knowledge and language ability in undertaking an ESP test. However, in this case, language ability has greater effects on test performance than background knowledge. Even though no interaction effect between these two variables was found, language ability differences could be seen in both engineering and non-engineering groups.
5. The findings reveal that all groups of test takers employed the same test taking strategies in answering different types of test items. Moreover high language ability students employed a wider range of strategies and used them more frequently than low language ability students. Therefore, emphasis should be given to increasing the use of test taking

strategies among low language ability students to improve their performance in ESP tests.

6. One of the main problems of this research was motivating and securing the cooperation of test takers, especially non-engineering students to sit for the test. Their performance in the test depended greatly on their attitudes and motivation. Findings indicate that the test takers in this study had positive attitudes towards the test. The researcher's observation is that elaborating on the development procedures and usefulness of the test, particularly as regards the score report for self-assessment can contribute greatly to positive attitudes among test takers. These procedures could be replicated by other research.

5.5 Recommendations for Future Research

Following are some recommendations for further research.

1. This study could be replicated by
   - 1.1 using more subjects with a wider range of language ability levels such as high, average, and low.
   - 1.2 using passages which are more specifically related to the students' specialized fields of study; another option is to use passages from different disciplines in a test.
   - 1.3 using subjects from a single discipline background; in this way familiarity with content can be better controlled.
2. To improve test authenticity and perhaps reveal some interesting facts, the test could be developed using different types of test formats such as gap filling and short answers.
3. Apart from the engineering field, the test can be developed to assess test takers' background knowledge in other broad discipline areas such as English for Business, or English for Science.
4. The TLU domain and test construct could be defined differently from this study, depending on stakeholders' needs in a particular context.
5. Definitions of language ability developed from criteria other than those developed in this project (i.e. test scores from an English proficiency test

initiated for this research) can be used. Alternative criteria may include the scores from readily available English proficiency tests.

6. Tests for other language skills such as speaking, writing, listening or integrated skills could be developed for future research.
7. An investigation of test taking strategies in other test formats such as gap filling, true/false, matching and short answer could be conducted to provide alternative perspectives on test taking strategies,.
8. Variables other than those included in the present study could be explored. Other test takers' characteristics such as learning strategies and styles, anxiety, personality and risk-taking may be related to test performance.

REFERENCES

Thai

เพ็ญแข ศิริวรรณ. (2546). *สถิติเพื่อการวิจัยโดยใช้คอมพิวเตอร์ (SPSS Version 10.0).* ครั้งที่ 2. กรุงเทพฯ : เท็กซ์แอนด์เจอร์นัลพับลิเคชั่น.

ศิริชัย พงษ์วิชัย. (2549). *การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์.* ครั้งที่ 16. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

English

Adreas, R. (2005). *100 events that shook our world*. New York: Life Books.

Alderson, J. C. (1990a). Testing reading comprehension skills (Part One). *Reading in a Foreign Language* 6(2): 425-438.

Alderson, J. C. (1990b). Testing reading comprehension skills (Part Two). *Reading in a Foreign Language* 7(1): 465-503.

Alderson, C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. C. & Urquhart, A. H. (1983). The effect of student background discipline on comprehension: a pilot study. In A. Hughes, & D. Porter (Eds.), *Current developments in language testing* (pp. 121-130). London: Academic Press.

Alderson, J. C. & Urquhart, A. H. (1985a). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing* 2: 192-204.

Alderson, J. C. & Urquhart, A. H. (1985b). This test is unfair: I'm not an economist. In P. C. Hauptman, R. Le Blanc, & M. B. Wesche (Eds.), *Second language performance testing*. Ottawa: University of Ottawa Press.

Alexander, P.A., Schallert, D.L. & Hare, V.C. (1991). Coming to terms: how researchers in learning and literacy talk about knowledge. *Review of Educational Research* 61(3): 315-343.

Anderson, N.J., Bachman, L., Perkins, K. & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing* 8(1): 41-66.

Arkkelin, D. (2003). *Using SPSS to understand research and data analysis.* Available from: http://wwwstage.valpo.edu/other/dabook/ch13/c13-1.htm [February 20, 2008]

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L.F.(2004). *Statistical analyses for language assessment.* Cambridge: Cambridge University Press.

Bailey, K. M. (1999). *Washback in language testing.* Available from: http://ftp.ets.org/pub/toefl/Toefl-MS-15.pdf [December 1, 2006]

Ballantyne, C. (2004). *Multiple choice test.* Available from: http://www.tlc.murdoch.edu.au/eddev/evaluation/mcq/mctests.html#strengths [February 16, 2007]

Becker, L. (2000). *Effect size (ES).* Available from: http://web.uccs.edu/lbecker/SPSS/glm_effectsize.htm [January 12, 2007]

Beer, F. & Johnston, R. (1977). *Vector mechanics for engineers statics*. London: McGraw-Hill.

Biemans, H.J.A & Simons, P.R., (1996). Contact-2: A computer-assisted instructional strategy for promoting conceptual change. *Instructional Science* 24: 157-176.

Brown, J.D. (1981). An ESP testing project. *English for Specific Purposes* 53: 1-3.

Brown, J.D. (1988). Components of engineering-English reading ability? *System* 16(2): 193-200.

Brown, J.D. (2004a). *Language assessment: Principles and classroom practices.* New York: Longman.

Brown, J.D. (2004b). Test-taker motivations. *JALT Testing and Evaluation SIG Newsletter* 8(2): 16-20. Available from: http://www.jalt.org/test/bro_20.htm [March 26, 2008]

Brown, J.D. (2005). *Testing in language programs.* New York: McGraw-Hill.

Brownstein, S., Weiner, M., & Green, S. (1999). *How to prepare for the GRE: Graduate Record Examination.* New York: Barron’s Educational Series, Inc.

Carrell, P.L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly* 19(4): 727-752.

Chalhoub-Deville, M. (Ed.). (1999). *Issues in computer-adaptive testing of reading proficiency.* Cambridge: Cambridge University Press.

Chatterji, M. (2003). *Designing and using tools for educational assessment.* Boston: Pearson Education.

Child, D. (2006). *Text readability tools.* Available from: http://www.ilovejackdaniels.com/readability.php [September 2, 2006]

Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension.* Cambridge: Cambridge University Press.

Clapham, C. (2000). Assessment for academic purposes: where next? *System* 28: 511-521.

Coe, R. (2000). *What is an effect size?* Available from: http://davidmlane.com/hyperstat/effect_size.html [January 12, 2007]

Cohen, A. (1984). On taking language tests: What the students report. *Language Testing* 1(1): 70-81.

Cohen, A. (1998). Strategies and processes in test taking and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research.* Cambridge: CUP.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^{nd} ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin* 112(1): 155-159.

Dewey, J. C. (1935). The acquisition of facts as a measure of reading comprehension. *The Elementary School Journal* 35: 346-348.

Dochy, F.J. R.C. & Alexander, P.A. (1995). Mapping prior knowledge: a framework for discussion among researchers. *European Journal of Psychology of Education* 10(3): 225-242.

Douglas, D. (2000). *Assessing languages for specific purposes testing.* Cambridge: Cambridge University Press.

Erickson, M., & Molloy, J. (1983). ESP test development for engineering students. In J. Jr. Oller (Ed.), *Issues in language testing research* (pp. 280-288). Massachusetts: Newbury House.

Encyclopedia Britannica (1993). *1993 Britannica book of the year*. Chicago: Encyclopedia Britannica, Inc.

Field, A. (2005). *Discovering statistics using SPSS (2nd ed.)*. London: Sage Publication Ltd.

Floyd, P. & Carrell, P.L. (1987). Effects on ESL reading of teaching cultural content schemata. *Language Learning* 37: 89-108.

Fraenkel, J.R. and Wallen, N.E. (2000). *How to design and evaluate research in education (4th ed.)*. Boston: McGraw-Hill.

Gay, L. & Airasian, P. (2000). *Educational research: competencies for analysis and application.* New Jersey: Prentice-Hall, Inc.

Ghahraki, S. and Sharifian, F. (2005). The relationship between overall reading comprehension and determination of fact/ opinion in L2. *The Reading Matrix* 5(1). Available from: http://www.readingmatrix.com/articles/sharifian/article.pdf. [December 21, 2006]

Grabe, W. (Autumn, 1991). Current developments in second language reading research. *TESOL Quarterly* 25 (3): 375-406.

Halliday, M. & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Hatch, E. & Farhady, H. (1982). *Research design and statistics for applied linguistics*. Massachusetts: Newbury House Publishers, Inc.

Hopkins, W. (2002). *Scale of magnitudes for effect statistics.* Available from: http://davidmlane.com/hyperstat/effect_size.html [January 18, 2007]

Hudson, T. (1982). The effects of induced schemata on the “short circuit” in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning* 32: 1-31.

Hudson, T. (1993). Testing the specificity of ESP reading skills. In D. Douglas & C. Chapelle. (Eds.), *A new decade of language testing research: Selected papers from the 1990 language testing research colloquium* (pp. 58-82). Virginia: Teachers of English to Speakers of Other Languages, Inc.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: CUP

Jensen, C. & Hansen, C. (1995). The effect of prior knowledge on EAP listening test performance. *Language Testing* 12: 99-119.

Johns, M., & Dudley-Evans, T. (1991). English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly* 25(2): 297-314.

Johnson, P. (1981). Effects on reading comprehension of language complexity and cultural background of a text. *TESOL Quarterly* 15: 169-81.

Johnston, P. H. (1983). *Reading comprehension assessment: A cognitive basis.* Newark. Del: International Reading Association

Jordan, R. (1997). *English for academic purposes*. Cambridge: Cambridge University Press.

Kehoe, J. (1995). Writing multiple-choice test items. *Practical Assessment, Research & Evaluation* 4(9). Available from: http://PAREonline.net/getvn.asp?v=4&n=9 [February 18, 2007]

Kish, Leslie. (1965). *Survey Sampling*. New York: John Wiley and Sons, Inc.

Koh, Y. (1985). The role of prior knowledge in reading comprehension. *Reading in a Foreign Language* 3: 375-380.

Kunnan, A.T. (1995*). Studies in language testing: Test taker characteristics and test performance – a structural modeling approach.* Cambridge: Cambridge University Press.

Lynch, B. & Hudson, T. (1991). EST reading. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 216-232). New York: Newbury House.

Mudraya, O. (2006). Engineering English : A lexical frequency instructional model. *English for Specific Purposes* 25: 235-256.

Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

Nieh, A. (2000). *A comparative analysis of the effect of background knowledge upon Taiwanese EFL university students' reading comprehension.* Available from: http://www.melta.org.my/modules/sections/2.doc [January 29, 2007]

Nisa Vongpadungkiat. (2006). *A study of selected predictors of English reading test performance in consumer product marketing of fourth-year university students.* Doctoral dissertation. English as an International Language Program, Graduate School, Chulalongkorn University.

North, B., and Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing* 15(2): 217-262.

Pearson, P.D., Roehler, L.R., Dole, J.A. & Duffy, G.G. (1991). Developing expertise in reading comprehension: what should be taught? how should it be taught*? Technical Report No. 512.* Available from: http://www.madison.k12.wi.us/tnl/langarts/pdf/writersactivate4.PDF [December 1, 2006]

Prapphal, K. (1994). *The relationship between the reading strategies and language background of Thai Science students in performing summary tests.* Available from: http://pioneer.chula.ac.th/~pkanchan/html/sumtest.htm [March 25, 2008]

Professional Testing Inc. (2005). *Conduct the item analysis*. Available from: http://www.proftesting.com/test_topics/steps_9.shtml [February 25, 2008]

Rand, W. (1997). Testing deviant! Are you one also? *Forum* 53(2). Available from: http://exchanges.state.gov/forum/vols/vol53/no2/p48.htm [December 2, 2006]

Robinson, H. M. (1966). *Reading: Seventy-five years of progress.* Chicago: University of Chicago Press.

Saville, N. (2000). *Developing language learning questionnaires*. Available from: http://www.cambridgeesol.org/rs_notes/0001/rs_notes1_4.cfm [November 30, 2006]

Scientific America. (2006, April). *Bug-eyed lens permits panoptic views*. Available from: http://www.sciam.com/article.cfm?chanID=sa002&articleID=0002BFB3-30C8-1451-B0C883414B7F0000 [August 9, 2006]

Scientific America. (2006, August). *Nanotechnology's future*. Available from: http://www.sciam.com/article.cfm?chanID=sa002&articleID=00029E0B-34C6-14C0-AFE483414B7F4945 [August 15, 2006]

Schuele, C. M., & Justice, L. M. (2006, August 15). The importance of effect sizes in the interpretation of research. *The ASHA Leader* 11(10). Available from: http://www.asha.org/about/publications/leaderonline/archives/2006/060815/f060815d.htm [January 12, 2007]

Shoham, M., Perets, S. & Vorhaus, R. (1987). Reading comprehension tests: general or subject specific? *System* 15: 81-88.

Singer, H., & Ruddell, R. B. (1976). *Theoretical models and processes of reading.* Newark, Del.: International Reading Association

Sjoberg, L. (2005). A review of handbook of attitudes. *PsycCRITIQUES* 50(42): 1-8. Available from: http://www.testcentrum.se/pdf/The%20Structure%20and%20Importance%20of%20Attitudes%20as%20published%20oct%202005.pdf [December 2, 2006]

Smith, C. (1997). Vocabulary instruction and reading comprehension. *ERIC DIGEST*, June 1997. Available from: http://www.indiana.edu/~reading/ieo/digests/d126.html [April 21, 2008]

Smith, F. (1970). *Understanding reading.* New York: Halt, Rinehart, and Winston.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the acquisition of literacy. *Reading Research Quarterly* 16: 32-71.

Stevens, K.C. (1980). The effect of background knowledge on the reading comprehension of ninth graders. *Journal of Reading Behavior* 12(2): 151-154.

Strangman, N., & Hall, T. (2004). *Background knowledge.* Available from: http://www.cast.org/publications/ncac/ncac_backknowledge.html [November 3, 2006]

Sudman, Seymour. (1976). *Applied sampling*. New York: Academic Press.

Tabachnick, B. & Fidell, L. (2001). *Computer-assisted research design and analysis.* Massachusetts: Allyn & Bacon.

Tan, S. H. (1990). The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates. In J.H.A.L. de Jong and D. K. Stevensons (eds.) *Individualizing the Assessment of Language Abilities*. Clevedon, PA: Multilingual Matters.

Thalheimer, W. & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology. Available from: http://work-learning.com/effect_sizes.htm [January 16, 2007]

Thorndike, E. L. (1917a). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology* 8: 323-332.

Thorndike, E. L. (1917b). The psychology of thinking and the case of reading. *Psychological Review* 24: 220-234.

Thorndike, E. L. (1917c). The understanding of sentences: A study of errors in reading. *The Elementary School Journal* 18: 98-114.

Timings, R.L. & Twigg, P. (2001). *The pocket illustrated dictionary of engineering terms.* Oxford: Butterworth-Heinemann.

Tulane University. (2006). *Helpful teaching hints.* Available from: http://www.som.tulane.edu/ome/helpful_hints/test_analysis.pdf [February 25, 2008]

University of New England. (2000). *Chapter 7: analysing the data: part IV: analysis of variance.* Available from: http://www.une.edu.au/WebStat/unit_materials/c7_anova/twoway_anova.htm [February 20, 2008]

Whittemore, J. L. (2004). *An overview of attitude structure and formation.* Available from: http://commoncontent.org/catalog/text/reference/4256/ [December 1, 2006]

Wikibook (2008). *Research methods/ Two-way ANOVA.* Available from: http://en.wikibooks.org/wiki/Research_Methods:Two-Way_ANOVA [February 20, 2008]

Wikimedia Foundation Inc.(2006). *Likert scale*. Available from: http://en.wikipedia.org/wiki/Likert_Scales [March 1, 2008]

Wikipedia (2007). *Effect size*. Available from: http://en.wikipedia.org/wiki/Effect_size#Pearson_r_correlation [January 12, 2007]

Wongsothorn, A. (1996). *Approaches to language testing.* Bangkok: Chulalongkorn University Language Institute.

APPENDICES

Appendix A: Interview questions for engineering experts

The Effects of Language Ability and Engineering Background Knowledge on ESP Reading Ability of Thai Graduate Students

ผลกระทบของความสามารถทางภาษา และภูมิหลังความรู้ทางวิศวกรรมศาสตร์ที่มีต่อความสามารถในการอ่านภาษาอังกฤษเฉพาะกิจของนักศึกษาไทย ระดับบัณฑิตศึกษา

Objectives of the study:

ศึกษาผลกระทบของความสามารถทางภาษา และภูมิหลังทางการศึกษา ที่มีต่อความสามารถในการอ่านภาษาอังกฤษสาขาวิศวกรรมศาสตร์ของนักศึกษาไทย ระดับบัณฑิตศึกษา และเพื่อที่จะดูผลกระทบทั้ง 2 ตัวนี้ได้ จำเป็นจะต้องสร้างแบบทดสอบการอ่านภาษาอังกฤษสาขาวิศวกรรมศาสตร์

Purpose of an Interview:

นำข้อมูลจากการสัมภาษณ์ไปสร้าง Specification ของข้อสอบ และสร้างแบบทดสอบให้เหมาะสมกับการอ่านภาษาอังกฤษของสาขาวิศวกรรมศาสตร์ที่ถูกต้อง

Part I: Interviewees' personal information

1. Name: ____
2. Department: ____
3. Faculty/University:
   ____
4. How long have you been teaching? ____________________________
5. What is the highest **degree** you gained and what **field**?
   ____

Part II: Test development

<u>Expectations</u>

1. ในความเห็นของอาจารย์ นักศึกษาวิศวกรรมศาสตร์ควรมีความสามารถในการใช้ทักษะต่าง ๆ ที่เกี่ยวกับภาษาอังกฤษ ในการทำอะไรบ้าง (เน้นระดับปริญญาโท)

(In your opinion, what should engineering students be able to do in terms of)

Listening: ____

Speaking: ____

Reading: ____

Writing: ____

2. ในความเห็นของอาจารย์ 4 ทักษะต่อไปนี้ คือ ฟัง พูด อ่าน เขียน ทักษะใดที่มีความสำคัญและจำเป็นต่อการเรียนของนักศึกษาสาขาวิศวกรรมศาสตร์ ระดับบัณฑิตศึกษามากที่สุด และเพื่อวัตถุประสงค์ใด)
(In your opinion, what is the most necessary language skill used in graduate-level engineering education? and for what purpose?)

Answer:

____

____

3. ถ้ากล่าวถึงวิศวกรรมศาสตร์โดยทั่วไป วิชาใดเป็นวิชาพื้นฐานสำหรับนักศึกษาวิศวกรรมศาสตร์ที่ต้องเรียน
(Regarding general engineering, what are the fundamental courses for testing engineering students?)

1.____________________________________2.________________________________

3.____________________________________4.________________________________

4. . ก่อนที่นักศึกษาวิศวกรรมศาสตร์จะเข้าศึกษาต่อในระดับบัณฑิตศึกษา ความสามารถทางภาษาอังกฤษของนักศึกษาควรอยู่ในระดับใด ถ้าเปรียบเทียบกับ TOEFL คะแนนควรอยู่อย่างต่ำที่ประมาณเท่าไหร่
(Before entering the graduate school in Engineering, what is the expected level of English proficiency? Compare with TOEFL, what is the approximate accepted score?)

Level: (advance, intermediate, or beginner) ________________________________

Compare with TOEFL:

____

5. ความสามารถทางภาษาอังกฤษมีความสำคัญมากน้อยแค่ไหนต่อการเรียนในระดับบัณฑิตศึกษาของนักศึกษาวิศวกรรมศาสตร์ (To what extent is English language ability important for engineering students in studying in graduate level?)
Answer:

____

____

6. ในปัจจุบัน ท่านพอใจกับความสามารถทางภาษาอังกฤษของนักศึกษาวิศวกรรมศาสตร์ ระดับบัณฑิตศึกษาหรือไม่
(At present, are you satisfied with English language ability of your engineering graduate students?)
Answer:

____

____

7. หลังจากที่นักศึกษาเรียนวิชาบังคับภาษาอังกฤษสำหรับบัณฑิตศึกษา ท่านคิดว่านักศึกษาจะมีความรู้ภาษาอังกฤษดีพอหรือไม่ (Do you think students will adequately have a sound knowledge of English by the end of taking compulsory courses?)

Answer:

____

____

8. ท่านคิดว่าควรจะมีการพัฒนาแบบทดสอบภาษาอังกฤษสำหรับนักศึกษาวิศวกรรมศาสตร์หรือไม่ เพราะ เหตุใด

(Should the English test for engineering students be developed? Why?)

Answer:

____

____

Test components

9. ถ้ากล่าวถึงองค์ประกอบของแบบทดสอบการอ่านภาษาอังกฤษสาขาวิศวกรรมศาสตร์ ควรประกอบด้วยอะไรบ้าง (Regarding the components of the Engineering-English reading test (E-ERT), what should be included in the E-ERT?)

Answer:____

____

____

10. แบบทดสอบนี้ควรประกอบด้วยองค์ประกอบทางวิศวกรรมศาสตร์อย่างเดียว หรือ ผสมกันระหว่างองค์ประกอบทางภาษาและองค์ประกอบทางวิศวกรรมศาสตร์ (Should the test consist of only the engineering factor (focus on technical terms) using English as a medium? Or should the test mix between linguistic factor (focus on grammar with general engineering text) and engineering factor (focus on soft technical terms)?

Answer:

____

11. ถ้าผสมกันในสัดส่วนเท่าไหร่ที่เหมาะสม (If mixed, in what proportion?)

Answer: Engineer ______________%______________________________

Linguistic ____________________%______________________________

Content

12. กลุ่มหัวข้อใดที่อาจารย์แนะนำให้ใส่ในแบบทดสอบ

(What topics would you recommend to put in the test? (Physical Science, Technology, general science or general knowledge…etc.)

Answer:____

____

13. กรุณาแนะนำแหล่งที่มาของ text ซึ่งจะเป็นแหล่งเนื้อหาของแบบทดสอบ

(Can you recommend me about the source of text which will provide the content to be put in the test? E.g. textbook, journal, academic report, course book or etc. Give name of them if possible.)

Answer:____

____

____

Test format

14. รูปแบบของแบบทดสอบควรเป็นอย่างไร (What item type should be put in the test?)

(only m/c, short answer, matching etc.. or mixing)

Answer: ____

Test taking time

15. แบบทดสอบควรใช้เวลาทำนานเท่าไหร่ ____

(How long should the test be taken? **1 hour, 2 hours or other** (please specify)

Passage length

16. ถ้ากล่าวถึงความยาวของแบบทดสอบ ต่อ 1 passage ควรยาวเท่าไหร่ ____________________

(Regarding the length of a test passage (only 1), how long should it be?

1 page, 2 pages or other (please specify)

17. ควรมีทั้งหมดกี่ passage ____

(How many passages should be included in the test?)

18. แบบทดสอบควรมีกี่ข้อ ____

(How many test items should be included in the test?)

Comment:____

____

____

____

Appendix B: Summary of interviews for test development

Part I: Interviewees' personal information

10 Engineering experts from different disciplines (i.e. civil engineering, environmental engineering, industrial engineering, food engineering and mechanical engineering) and from different universities (KMUTT, Mahasarakam University and Technology Suranaree University) were interviewed. Questionnaires related to the interview's questions were also distributed to 10 engineering experts at Mahasarakam University. Those experts have been teaching engineering students from approximately 2 to 22 years. They all earned Ph.D. in engineering.

Part II: Test development

Expectations

1. Abilities to deal with each skill:
   Listening: listen to guest lecturers
   Speaking: present in an international seminar
   Reading: read textbooks, research papers/ articles, journals, analyze tables & graphs
   Writing: write abstracts, research articles, assignments, reports, script for presentation
2. The most necessary language skill: 71% of participants answered READING.
3. The fundamental courses for engineering students: Maths, Physics, Statics, Thermodynamic, Mechanics, Materials and ENGLISH (53% of participants answered)
4. Before entering the graduate school, the expected level of English proficiency is at intermediate, approximately 470-500 TOEFL scores.
5. All the informants considered that English language ability was highly important for their engineering students in studying in graduate level.
6. All of them were not satisfied with their students' English language ability.
7. 71% of them were not sure about their students' English language ability if they adequately have a sound knowledge of English by the end of the compulsory courses they are required to take.
8. 71% of them agreed that the English test for engineering students should be developed to self-assess their students' actual English language ability.

Test components

9. The components of the Engineering-English reading test (E-ERT) should include something similar to TOEFL, vocabulary, grammar and reading comprehension but the content based on general engineering text.

10. + 11. The test should consist of both **50% linguistic factor** (focus on grammar (and engineering structure) and general vocabulary with general engineering text) and **50 % engineering factor** (focus on soft technical terms and reading for main idea & details: use background knowledge, not common sense)

Note: additional idea is engineering experiment: experimental design, method, result, discussion and conclusion.

Content

12. Recommended topics: General Science and Technology (advance or hot news), General knowledge about engineering (i.e. Automation, Lane), In-trend technology (i.e. Nano, Biotechnology), Fundamental knowledge for engineering students (i.e. Thermodynamic, Heat transfer and Statics)
13. Source of text:

    Textbook: Static1, Fundamental of Engineering Materials
    Journal: related to engineering, Science and Technology
    magazine: Scientific American (advance terms), Discover (short story & news), Business Week (IT), Popular Science
    internet: www.sciam.com, www.sciencedirect.com, www.bbc.com, www.boa.com
    newspaper: Bangkok Post
    GRE test: Graduate Record Examination

Test format

14. Item type: m/c , matching and short answer

Test taking time

15. The test should take about 1.5 – 2 hrs.

Test length

16. There should be about 4-6 passages in the test.
    1. Linguistic factor (2 passages) + 1 passage (pilot) 10 items / 1 passage
    2. Engineering factor (2 passages) + 1 passage (pilot)

Recommendations:

1. The test passage should be arranged from easy to difficult degree. In other words, the test passages should be included with different degree of difficulty.
2. The test should mix between linguistic and engineering factors. It's not necessary to specially focus on engineering knowledge because students have been tested knowledge in their field already. Then, it is not necessary to focus on any technical terms. It may favor one particular field.
3. Alternatively, the test designer may provide passages with different major disciplines. For example, there are 4 passages in the test focus on: civil engineering, mechanical engineering, computer engineering and electrical engineering. Then, it's fair for all testees to do the test both in their field and out of their field.
4. ESP courses should be offered in order to reinforce the use of English of graduate-level students in engineering or any particular fields with guidance from experts in the fields after they complete the compulsory English courses if possible.

Appendix C: Table of specifications of the Engineering-English Reading Test (E-ERT)

There are two main parts in the test: vocabulary and reading comprehension

Vocabulary Part		
Part 1	**Reading topic**	**Item numbers**
Sub-technical terms	-	10 items
Non-technical words	Relativity Passes Absolute Test	10 items
Reading Comprehension Part		
Part 2	**Reading topic**	**Item numbers**
Reading comprehension	Engineering Projects: Building	8 items
Reading comprehension	Mechanics	8 items
Reading comprehension	Engineering Experiment	9 items
Total		**45 items**

Vocabulary Part

Objectives	Topic	Text length	Item type	Item number	Weight	Points	Time
Part 1 (a): Sub-technical terms in Engineering							
Test takers will be able to recognize the sub-technical terms used in engineering.	-	-	Multiple choice	10	22%	1	15 minutes
Total				**10**	**22%**	**10**	

Part 1 (b): Non-technical words in Engineering							
Test takers will be able to recognize the non-technical words used in engineering.	Relativity Passes Absolute Test	-	Multiple choice	10	22%	1	15 minutes
Total				**10**	**22%**	**10**	

Reading Comprehension Part

Objectives	Topic	Text length	Item type	Item numbers	Weight	Points	Time
Part 2: Passage 1							
Test takers will be able to: 1. skim the text for **main idea**s	Engineering project: Buildings	570 words	m/c	1	2.2%	**1**	30 minutes
2. scan the text for **specific details**			m/c	3	6.6%	**3**	
3. make **inferences** from the reading			m/c	2	4.4%	**2**	
4. identify the **fact that uses background knowledge**			m/c	2	4.4%	**2**	
Total				**8**	**17.6%**	**8**	

Reading Comprehension Part

Objectives	Topic	Text length	Item type	Item numbers	Weight	Points	Time
Part 2: Passage 2							
Test takers will be able to: 1. skim the text for **main ideas**	Mechanics	600 words	m/c	1	2.2%	1	30 minutes
2. scan the text for **specific details**			m/c	4	8.8%	4	
3. make **inferences** from the reading			m/c	1	2.4%	2	
4. identify the **fact that uses background knowledge**			m/c	2	4.4%	2	
Total				**8**	**17.8%**	**9**	

Reading Comprehension Part

Objectives	Topic	Text length	Item type	Item numbers	Weight	Points	Time
Part 2: Passage 3							
Test takers will be able to: 1. skim the text for **main idea**s	Engineering Experiment	650 words	m/c	1	2.2%	1	30 minutes
2. scan the text for **specific details**			m/c	4	8.8	3	
3. make **inferences** from the reading			m/c	3	6.6%	2	
4. identify the **fact that uses background knowledge**			m/c	1	2.2%	2	
Total				**9**	**19.8%**	**8**	

Appendix D: The E-ERT (pilot study version)

Engineering-English Reading Test for Thai Graduate Students

Instructions:

1. The test consists of 2 parts with total 5 sections (excluding the cover).
2. You are allowed 90 minutes to complete the test.
3. Complete all answers in the answer sheet provided.
4. Do not write anything on the test paper.

This test paper consists of the following parts

Part	Item numbers	Scores
Part I: Vocabulary	1-10	10
	11-20	10
Part II: Reading comprehension	21-28	8
	29-36	8
	37-45	9
	45 items	**45 points**

The Test Designer:
Natjiree Jaturapitakkul
EIL Program, Chulalongkorn University
Semester 1/ 2007

Part I: Vocabulary: sub-technical (Questions 1-10)

a) Instructions: Put a cross (X) on the correct answers (a, b, c or d) in the answer sheet provided.

1. A _______________ may be thought of as any influence which tends to change the motion of an object as a lift, a push, or a pull, and has a magnitude and a direction.
   a. force* b. matrix c. stress d. velocity

.

2. The amount of energy transferred by a force is known as _______________.
   a. element b. mass c. shear d. work*

3. The internal resistance of a material when the shape is changed by the application of an external force is generally called _______________.
   a. body b. energy c. entropy d. stress*

4. The _______________ is a quantity that represents the magnitude of force applied to a rotational system at a distance from the axis of rotation.
   a. compression b. moment* c. statics d. tension

5. The deformation of a material substance in which parallel internal surfaces slide past one another is called _______________.
   a. flow b. distortion c. shearing* d. volume

6. The _______________ refers to the state of a body at rest or in uniform motion in which the resultant of all forces on it is zero.
   a. equilibrium* b. magnitude c. origin d. mechanics

7. The force that opposes the movement of one surface relative to another with which it is in contact is known as _______________.
   a. dynamics b. friction* c. resilience d. viscosity

8. The _______________ is speed that has a clearly stated direction.
   a. element b. flow c. momentum d. velocity*

9. An object defined by both magnitude and direction is called _______________.
   a. elasticity b. placement c. vector* d. vortex

10. A/ An _______________ is a fundamental two-dimensional object. Intuitively, it may be visualized as a flat infinite sheet of paper.
    a. area b. circle c. line d. plane*

Part I: Vocabulary: non-technical (Questions 11-20)

b) Instructions: Choose the best answer which is closest in meaning to the word underlined. Put a cross on the correct answers (a, b, c or d) in the answer sheet provided.

Relativity Passes Absolute Test

According to Albert Einstein's general theory of relativity, a massive **11)**body like Earth should bend the space-time fabric of the universe, causing it to curve and flex like a trampoline supporting a bowling ball.

Nearly three years ago, NASA's oft-canceled $750 million Gravity Probe B Relativity Mission finally shot into space with one goal—to quantify Einstein's **12)**predictions from Earth's orbit. Earlier this year, at the meeting of the American Physics Society, principal investigator Francis Everitt **13)**delivered the first results: Gravity Probe B has verified Einstein's theory to within 1 percent.

Four gyroscopes, each the size of a Ping-Pong ball, form the heart of the experiment. The gyroscopes are the most perfectly **14)**spherical man-made objects in existence; if inflated to the size of Earth, they would have mountains no more than eight feet high. (Their near-faultless roundness has landed the spheres in the Guinness World Records.) At the beginning of the experiment, the gyroscopes' **15)**axes pointed to a distant star; as the spacecraft moved around Earth for nearly a year, the researchers carefully **16)**monitored the position of the axes.

Einstein's theory predicts that the axes should shift by a **17)**tiny amount—0.0018 degree—under the influence of Earth's pull on space-time. After 18 months of data analysis, Everitt and his team measured the axial shift to within 1 percent of Einstein's prediction. Everitt, a Stanford physicist who has spent more than 40 years on the project, says the **18)**results are sweet indeed. "It's really extraordinary to look at the output and see Einstein looking back, without any 1**9)**calculations or corrections," he says. "This measurement is unprecedented in any test of general relativity."

In addition to claiming that the universe curves around massive bodies, Einstein said that as these bodies **20)**rotate, they effectively "drag" space behind them, creating a twist in the cosmic fabric. Everitt says his team plans to announce verification of this "frame dragging" effect later this year.

11.	a. content	b. figure	c. group	d. object*
12.	a. analyses	b. calculations	c. evaluations	d. forecasts*
13.	a. declared*	b. handled	c. made	d. sent
14.	a. cylindrical	b. oval	c. round*	d. square
15.	a. arrow lines	b. central lines*	c. dash lines	d. dot lines
16.	a. controlled	b. managed	c. observed*	d. tested
17.	a. decimal	b. huge	c. very low	d very small*
18.	a. numbers	b. outcomes*	c. quantities	d. truths
19.	a. computations*	b. estimations	c. practices	d. predictions
20.	a. move	b. spin*	c. swing	d. twist

Part II: Reading comprehension

Instructions: Read the passages and choose the one best answer for each question by putting a cross (X) in the answer sheet provided.

Passage 1 (Questions 21-28)

The following is an article about engineering projects on buildings.

Hong Kong's Central Plaza building, its 306.5 m-high superstructure making it the world's tallest reinforced-concrete building, was topped out in April 1992. A steel mast would bring the total height to 374 m. The building was an equilateral triangle in shape with each side approximately 50 m in length. The floors extended from closely spaced perimeter columns and around a central core area of elevators and service ducts. The use of concrete for such a tall building was made possible by the employment of high-strength concrete. There were three stories below ground, and under those the building rested on deep caissons up to 7.4 m in diameter.

In an advanced stage of planning during 1992 was the Tour Sans Fins in Paris. This was to be 426 m in height with a circular plan form 43 m in diameter. This caused it to be very slender in terms of building structures and, while the circular plan shape was efficient from the point of view of wind drag, it introduced the problem of wind vortex excitation. This phenomenon arises with rounded shapes and is caused by the periodic shedding of the vortices from alternate sides of the cylinder. At a certain wind speed this shedding can be at the same frequency as that of the building, causing the system to reach a high level of oscillation.

Concrete was being chosen for the structure on account of its efficient stiffness and damping properties (properties that diminish the amplitude of an oscillation) compared with those of steel. Even so, accelerations would likely exceed acceptable limits, and a damper at the top of the building was being considered. This would comprise a pendulum with a mass of only a fraction of that of the building and with a predetermined damping factor introduced between it and the structure. The Tour Sans Fins was to be a peripheral pierced and framed tube with no central core, thereby concentrating the load-carrying members in the area where they would contribute most to the strength and stiffness of the structure.

Another building of significant engineering interest was the Hotel delas Artes Tower in Barcelona, Spain, completed during 1992. The main structure was a 45-story, 135 m-high hotel-apartment tower with an externally expressed structural-steel frame. One of the main design features in a steel-framed building, after strength and stability, is fire resistance. Normally this is achieved by sprayed-on fire protection and/or cladding and encasement. In this case, however, the designers predicted that by having the structural members at least 1.5 m from the window wall, the steel, in the event of fire breaking through the façade, would not overheat. This feature was in addition to careful fire compartmenting within the building

to reduce the risk of fire spread. The structural concept was bold. The tower was 30 m square in plan and each elevation was divided into three bays, the outer ones being cross braced for the full height of the tower; at three levels the central bay was also cross braced, which significantly increased the stiffness of the building and reduced wind movement. Vertical loads were transferred from the floors to the external frame by fire-protected steel beams, four on each elevation. Detailed studies had to be made to predict the likely temperatures that the steel would reach in the event of a fire and to check that the stresses and deflections arising from those temperatures would be acceptable.

21. What is the main idea of the passage?
   a. The comparison between three buildings.
   b. The importance of concrete in building construction.
   c. The buildings with different shapes around the world.
   d. The prominent buildings around the world in the year 1992.*

22. According to a passage, what is a benefit of the circular plan shape of the Tour Sans Fins?
   a. Reduce wind drag *
   b. Increase wind vortex excitation
   c. Increase stiffness of the building
   d. Reduce the amplitude of an oscillation

23. From the passage, when the accelerations of the building exceed the acceptable limit, what do they do to solve this problem?
   a. Installing a pendulum with a mass in the building. *
   b. Installing deep caissons in the building.
   c. Installing steel mast in the building.
   d. Making the building higher.

24. According to the passage, which building seems to be of special interest for engineers in terms of fire resistance?
   a. The Tour Sans Fins
   b. Hong Kong’s Central Plaza
   c. The Hotel delas Artes Tower *
   d. Both a & b

25. It can be implied from the passage that the following fundamental considerations should be included in building construction EXCEPT
   a. steel *
   b. stability
   c. stiffness
   d. strength

26. It can be inferred from the passage that
   a. steel is normally chosen for constructing building structure.
   b. high-strength concrete is not necessarily used for constructing tall buildings.
   c. buildings made from concrete have more fire resistance than buildings made from steel. *
   d. the equilateral triangle plan is more efficient to reduce wind drag as compared to the circular plan.

27. Hong Kong's Central Plaza building is in 1992 the world's tallest reinforced-concrete building. What is a mixture of reinforced concrete?

a. concrete with sand inside
b. concrete with stones inside
c. concrete with steel bar inside *
d. concrete with stainless steel inside

28. Why was Hong Kong's Central Plaza building constructed by high strength concrete?

a. High strength concrete is stronger than steel.
b. The cost of high strength concrete is cheaper than normal concrete.
c. High strength concrete helps reduce building's weight but resist higher load. *
d. High strength concrete is easier to produce and use in constructing the building.

Passage 2 (Questions 29–36)

Here is a part of the text in "Vector mechanics for engineers: Statics".

Mechanics may be defined as that science which describes and predicts the conditions of rest or motion of bodies under the action of forces. It is divided into three parts: mechanics of rigid bodies, mechanics of deformable bodies and mechanics of fluids.

The mechanics of rigid bodies is subdivided into statics and dynamics, the former dealing with bodies at rest, the latter with bodies in motion. In this part of the study of mechanics, bodies are assumed to be *perfectly rigid*. Actual structures and machines, however, are never absolutely rigid and deform under the loads to which they are subjected. But these deformations are usually small and do not appreciably affect the conditions of equilibrium or motion of the structure under consideration. They are important, though, as far as the resistance of the structure to failure is concerned and are studied in mechanics of materials, which is a part of the mechanics of deformable bodies. The third division of mechanics, the mechanics of fluids, is subdivided into the study of incompressible fluids and of compressible fluids. An important subdivision of the study of incompressible fluids is hydraulics, which deals with problems involving liquids.

Mechanics is a physical science, since it deals with the study of physical phenomena. However, some associate mechanics with mathematics, while many consider it as an engineering subject. Both these views are justified in part. Mechanics is the foundation of most engineering sciences and is an indispensable prerequisite to their study. However, it does not have the empiricism found in some engineering sciences, i.e. it does not rely on experience or observation alone; by its rigor and the emphasis it places on deductive reasoning it resembles mathematics. But, again, it is not an abstract or even a pure science; mechanics is an applied science. The purpose of mechanics is to explain and predict physical phenomena and thus to lay the foundations for engineering applications.

Fundamental concepts: Although the study of mechanics goes back to the time of Aristotle (384-322 B.C.) and Archimedes (287-212 B.C.), one has to wait until Newton (1642-1727) to find a satisfactory formulation of its fundamental principles. These principles were later expressed in a modified form by D'Alembert, Lagrange, and Hamilton. Their validity remained unchallenged, however, until Einstein formulated his theory of relativity (1905). While its limitations have now been recognized, newtonian mechanics still remains the basis of today's engineering sciences.

The basic concepts used in mechanics are *space*, *time*, *mass* and *force*. These concepts cannot be truly defined; they should be accepted on the basis of our intuition and experience and used as a mental frame of reference for our study of mechanics

The concept of *space* is associated with the notion of the position of a point P. The position of P may be defined by three lengths measured from a certain reference point, or origin, in three given directions. These lengths are known as the coordinates of P.

To define an event, it is not sufficient to indicate its position in space. The *time* of the event should also be given.

The concept of *mass* is used to characterize and compare bodies on the basis of certain fundamental mechanical experiments. Two bodies of the same mass, for example, will be attracted by the earth in the same manner; they will also offer the same resistance to a change in translational motion.

A force represents the action of one body on another. It may be exerted by actual contact or at a distance, as in the case of gravitational forces and magnetic forces. A force is characterized by its point of application, its magnitude, and its direction; a force is represented by a vector.

29. What is the passage mainly about?
   a. Introduction to mechanics and its types
   b. Introduction to mechanics and its history
   c. Introduction to mechanics and its concepts *
   d. Introduction to mechanics and its principles

30. In paragraph 2, which of the following statements is TRUE?
   a. Statics deals with mean and S.D.
   b. Statics deals with unmoved bodies. *
   c. Dynamics deals with space and mass.
   d. Dynamics deals with immovable bodies.

31. In paragraph 3, which of the following statements about mechanics is FALSE ?
   a. Mechanics is related to mathematics.
   b. It is necessary for engineering students to study mechanics.
   c. Studying mechanics does not require experience or observation, but reasons.
   d. Mechanics focuses on giving specific reasons and then linking to general ideas.*

32. From the passage, the fundamental concepts and principles of mechanics was introduced by….
   a. Newton *
   b. Aristotle
   c. Archimedes
   d. D'Alembert and Lagrange

33. According to the passage, a force that is exerted at a distance is called….
   a. axial force
   b. shear force
   c. body force
   d. gravitational force *

34. It can be inferred that this passage is chosen from which of the following sources?
   a. Journal
   b. Text book *
   c. Science report
   d. Research article

35. What is a condition of equilibrium?
    a. Summation of all forces equal to zero *
    b. Summation of all forces equal to one
    c. All forces in x-direction equal to zero
    d. All forces in y-direction equal to zero

36. The basic concepts used in mechanics are space, time, mass and force. ‘Force’ in this passage probably refers to
    a. moment *
    b rigid bodies
    c. compressible fluids
    d. incompressible fluids

Passage 3 (Questions 37-46)

This is a part of the engineering experiment concerning “Idea feasibility testing”.

Demise of the Pony Express: Idea feasibility testing

In the response to the Gold Rush in the mid-1800s, the Pony Express was developed as an alternative mail service crossing the North American continent. Using mounted riders, the founders of the Pony Express hoped to create a conduit for mail that was faster and more reliable than stagecoaches. The Pony Express only operated for a year and a half—from April 1860 to November 1861—even though it offered excellent wages and employee benefits. The demise of the Pony Express is commonly attributed to the telegraph. The completion of the telegraph network and the termination of the Express occurred within a one-month period. But there is another, lesser-known myth (unknown, actually) that a different technology was the nail in the Express’s coffin. Letters were placed inside baseballs and thrown along a relay of ball carriers. This new system was faster than using horses, and the ball carriers were less tired than the Pony Express riders, who had to endure long days in the saddle. All a ball carrier had to do was to hang out and play catch. Can a ball relay beat a horse? Students applied their estimation skills and rapid idea testing processes to find an answer.

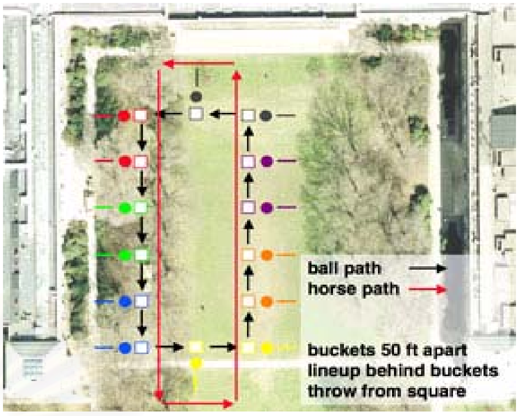
Some numbers: the estimation

Here is how the course instructor approached the problem. Assuming an average ball carrier can throw a 40 mph ball 100 feet and catch-throw in 2 seconds, the ball relay is expected to be a blazing 20 mph, roughly 2 times the Express’s average speed of 10 mph. However, just one drop in 10 might allow the express to win. With strong-armed ball carriers, the relay might even be competitive with a horse running at 30 mph. Not all students agreed. Some thought the ball relay would be much slower (up to 20 times slower), and a few thought it would be much faster (up to 10 times faster). A confirmation sketch model, or validation experiment, was in order.

Resolving the debate: Procedures

The proving ground was MIT’s Killian Court. After a 10-minute warm-up and practice session, the validation experiment was completed under the watchful eye of an impartial cowboy. The GPS-equipped Pony Express rider was Tom, the teaching assistant, on his horse headed mountain bike. 14 ball-throwing stations, one for each lab section consisting of 7-8 students, were set up in a rectangular loop. 50 feet, what we determined to be a manageable throwing distance for the average, non-athletic MIT arm, separated each station.

The Express rider

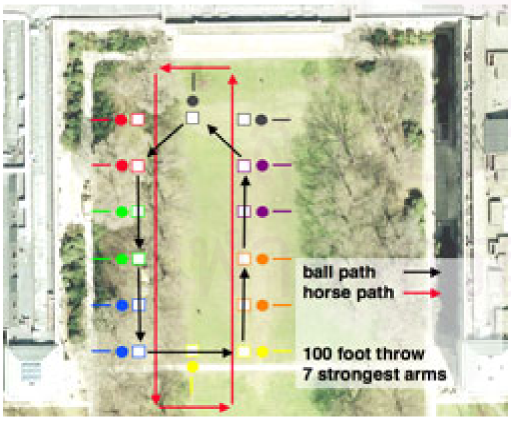


Test setup overview

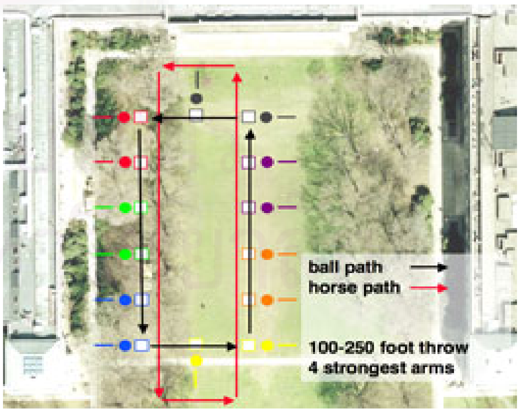
In the first test, all students were ball-carriers regardless of throwing experience, and the baseball completed 8 loops around the courtyard. This allowed each student at all the ball throwing stations to throw at least once. The rider reined his horse at the Express's average speed of 10 mph.

During the second run, two laps were completed with only the strongest-armed ball-carriers throwing from each station, still spaced at 50 feet. The Express rider maintained a speedy 15 mph. Over-throws and missed catches led to a dramatic finish — with the ball just overtaking the Express rider as he crossed the finish line.

The third challenge matched the 100-foot throwing distance assumed in the estimation calculations. Seven strong throwers threw to every second station and the rider rode his horse as fast as possible, competing over 2 loops around the courtyard. Finally, two laps were completed with the 4 strongest arms in the class on the circuit's corners. The winded horse struggled to keep pace.



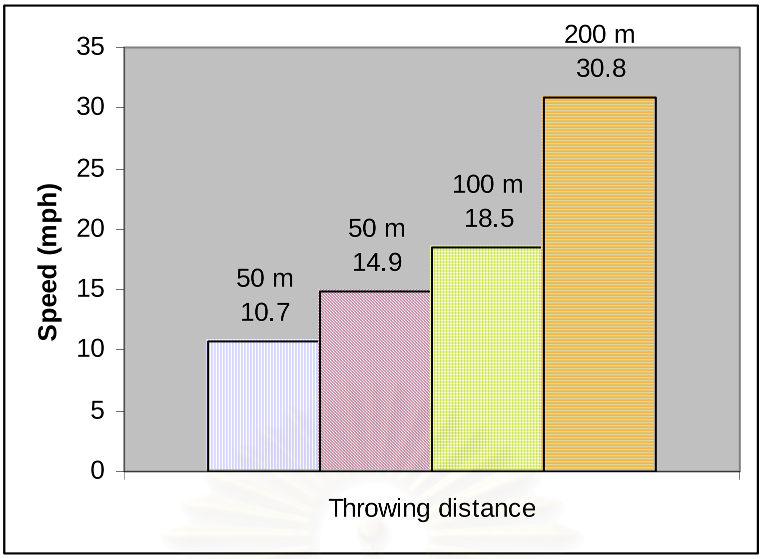
100 foot throw test layout



200 foot throw test layout

Results

Despite suffering from more balls dropped than caught (53% drop rate), the first test still had an average speed of 10.7 mph—just 7% faster than the reported average speed of the Pony Express.



Average speed of the four tests

Under the 100 ft throw conditions assumed in the back-of-the-envelope estimate, the predictions were born-out in reality. We estimated a relay speed of 20 mph and achieved an actual speed of 18.5 mph—an unexpected result given the ball drop rate was 40%. The average drop recovery was under 2 seconds, 10 times less than allowed for in the estimation.

Conclusion: Myth technically feasible! but practically ridiculous.

37. What is the main point of the passage?
    a. A competition between a horse rider and ball carriers.
    b. An experiment to find out whether or not a ball relay can beat a horse. *
    c. A competition on mail service between the Pony Express and the telegraph network.
    d. An experiment to find out about the speed of the ball relay with different throwing distance.
38. According to the passage, which of the following is TRUE about the Pony Express?
    a. It was established because of population growth in the US.
    b. It was a mail service by using a mounted rider as a postman.*
    c. It was faster and more reliable than other kinds of mail service in the North America.
    d. It was operated for a long period because it offered high salary and employee benefits.
39. From the passage, what could be a main cause of the end of the Pony Express?
    a. The telegraph network*
    b. The relay of ball carriers
    c. The exhaustion of the mounted riders
    d. The high wages and employee benefits
40. According to the passage, which of the following factors affected the experimental results?
    a. Gender and Catch-throw time
    b. Gender, catch-throw time and throwing distance
    c. Catch-throw time, throwing experience and throwing distance
    d. Catch-throw time, drop cost, throwing experience and throwing distance*

41. Which of the following statements is NOT inferred as assumptions of the experiment?
    a. The increasing throwing distance can increase the ball speed.
    b. The more balls dropped can allow the horse rider to win.
    c. The more skilled thrower can increase the ball speed.
    d. Force due to wind can decrease the ball speed.*

42. From the experiment, the estimated ball speed is 20 mph. In this case, how long does it take for a ball to travel one mile excluding drop time?
    a. 1 minute
    b. 2 minutes
    c. 3 minutes*
    d. 4 minutes

43. The procedures of the experiment include all of the following controlled parameters EXCEPT
    a. ball speed*
    b. horse speed
    c. throwing distance
    d. throwing experience

44. Which of the following statements about the results is FALSE?
    a. The more throwing distance, the more speed of the throwing relay.
    b. The more throwing experience, the more speed of the throwing relay.
    c. Throwing distance plays more important role than throwing experience when over the 100 foot throw.
    d. Throwing experience has more effects on the relay speed than throwing distance when over the 100 foot throw. *

45. Which is the best conclusion of this passage?
    a. It is proved that a ball relay beats a horse.
    b. It is possible to have a mail service by using a ball relay.
    c. Although the relay speed is satisfied, it is not practical to do.*
    d. Only a horse rider is more practical than a lot of ball carriers.

****This is the end of the test****

Appendix E: Item-Objective Congruence Evaluation Form

Engineering English Reading Test

The objectives of this test are to measure:

A) Vocabulary used in engineering

It involves:

1) Ability to recognize the meanings of sub-technical terms used in engineering.
2) Ability to identify the meaning of non-technical vocabulary in context

B) Comprehension

For comprehension, it involves:

1) Ability to skim the texts for main ideas.
2) Ability to scan the texts for specific information.
3) Ability to draw inferences based on information in the text.
4) Ability to identify the facts that use engineering background knowledge

Guideline for evaluation

Please put a tick (√) in the rating box (High, Medium, Low) the degree to which the item (question) measures the ability indicated in the objectives according to your opinion. Please also specify comments for each item.

H = High degree of congruence with the objective.
M = Medium degree of congruence with the objective.
L = Low degree of congruence with the objective.

Part I: Vocabulary Part

(a) Objective: To assess the ability to recognize the meanings of sub-technical terms used in engineering.

Item	H	M	L	Comments
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Is the content in this part appropriate ?

____________ Yes ______________ No

Comments ____

____

(b) Objective: To assess the ability to identify the meaning of non-technical vocabulary in context.

Item	H	M	L	Comments
11				
12				
13				
14				
15				
16				
17				
18				

19				
20				

Is the content in this part appropriate ?

____________ Yes ______________ No

Comments ____

____

Part II: Engineering Reading Comprehension

Passage1: Engineering projects: Buildings

Objectives: To assess students' reading ability in the following areas:

Item	Ability to	H	M	L	Comments
21	Skim the texts for main ideas.				
22	Scan the texts for specific information.				
23	Scan the texts for specific information.				
24	Scan the texts for specific information.				
25	Draw inference based on information in the text.				
26	Draw inference based on information in the text.				
27	Identify the fact that uses background knowledge.				
28	Identify the fact that uses background knowledge.				

Is the content in this part appropriate ?

____________ Yes ______________ No

Comments

____

____

Passage 2: Mechanics

Objectives: To assess students' reading ability in the following areas:

Item	Ability to	H	M	L	Comments
29	Skim the texts for main ideas.				
30	Scan the texts for specific information.				
31	Scan the texts for specific information.				
32	Scan the texts for specific information.				
33	Scan the texts for specific information.				
34	Draw inference based on information in the text.				
35	Identify the fact that uses background knowledge.				
36	Identify the fact that uses background knowledge.				

Is the content in this part appropriate ?

____________ Yes ______________ No

Comments

____

____

Passage 3: Demise of the Pony Express

Objectives: To assess students' reading ability in the following areas:

Item	Ability to	H	M	L	Comments
37	Skim the texts for main ideas.				
38	Scan the texts for specific information.				
39	Scan the texts for specific information.				
40	Scan the texts for specific information.				

41	Draw inference based on information in the text and graph.					
42	Identify the fact that uses background knowledge.					
43	Scan the texts for specific information.					
44	Draw inference based on information in the text and graph.					
45	Draw inference based on information in the text.					

Is the content in this part appropriate ?

____________ Yes ______________ No

Comments

____

____

<u>Overall evaluation</u>

Please put a tick (√) in front of the answer YES or NO and specify the comments according to your opinion.

1. Is the test appropriate to measure the English reading ability of the graduate-level students ?

   ________ YES ________ NO

   Comments: ____

   ____

2. Is the test appropriate to measure the English reading ability of engineering students?

   ________ YES ________ NO

   Comments: ____

   ____

3. Does the content represent various reading situations and text types found in Engineering academic area ?

_______ YES _______ NO

Comments: ____

____

4. Is the time allotment for the test appropriate?

_______ YES _______ NO

Comments: ____

____

5. Is it appropriate to use visual medias such as pictures and graph in this test?

_______ YES _______ NO

Comments: ____

____

Thank you for your time and kind attention.

Signature ______________________________ Date ________________

Appendix F: A priori validation result for the E-ERT

Note: 5 experts (2 engineering experts + 3 English lecturers)
* items that need to be revised

Part 1: Vocabulary
a: Sub-technical vocabulary

Item	H	M	L
1	**80%**	20%	
2	**80%**	20%	
3	**80%**	20%	
4	**80%**	20%	
5	**80%**	20%	
6	**80%**	20%	
7	**80%**	20%	
8*	**60%**	40%	
9	**80%**	20%	
10*	**60%**	40%	

b: Non-technical vocabulary

Item	H	M	L
11*	**60%**	20%	20%
12	**80%**	20%	
13	**80%**		20%
14	**80%**	20%	
15*	**40%**	60%	
16*	**60%**	20%	20%
17	**100%**		
18	**80%**		20%
19	**80%**	20%	
20	**80%**	20%	

Part 2: Reading comprehension
Passage 1: Engineering projects: Buildings

Item	H	M	L
21	**80%**		20%
22	**100%**		
23	**80%**	20%	
24*	**60%**	40%	
25	**80%**		20%
26	**80%**	20%	
27	**80%**		20%
28	**80%**	20%	

Passage 2: Mechanics

Item	H	M	L
29	**80%**	20%	
30*	**60%**	20%	20%
31	**100%**		
32	**100%**		
33	**80%**		20%
34*	**40%**	40%	20%
35	**80%**		20%
36	**80%**		20%

Passage 3: Engineering experiment

Item	H	M	L
37	**100%**		
38	**100%**		
39	**100%**		
40	**100%**		
41	**80%**	20%	
42	**80%**		20%
43	**80%**	20%	
44	**80%**		20%
45	**80%**	20%	

Appendix G: Item analysis for the pilot study

Items in the pilot study	Items in the main study	Difficulty indices (IF)	Delta	Discrimination Indices (ID)	Biserial Correlation Coefficients
1	1	0.800	8.82	0.200	0.403
2	2	0.750	10.26	0.500	0.567
3	3	0.600	11.95	0.800	0.838
4	4	0.650	11.42	0.400	0.463
5	5	0.425	13.80	0.300	0.348
6	6	0.750	10.26	0.200	0.318
7	7	0.525	12.71	0.700	0.702
8	8	0.475	13.29	0.500	0.548
9	9	0.775	9.94	0.500	0.684
10	10	0.425	13.80	0.400	0.324
11	11	0.750	10.26	0.400	0.471
12	12	0.425	13.80	0.500	0.356
13	13	0.325	14.86	0.300	0.342
14	14	0.450	13.54	0.600	0.503
15	15	0.525	12.71	0.200	0.296
16*		0.850	9.59	0.700	0.889
17	16	0.650	11.42	0.500	0.671
18	17	0.550	12.46	0.900	0.772
19	18	0.650	11.42	0.900	0.788
20	19	0.500	13.00	0.800	0.704
21	20	0.325	14.86	0.300	0.256
22	21	0.600	11.95	0.300	0.200
23	22	0.625	11.69	0.700	0.714
24*		0.900	8.36	0.400	0.791
25	23	0.475	13.29	0.400	0.221
26*		0.375	14.31	-0.200	-0.286
27	24	0.800	9.59	0.500	0.658
28	25	0.400	14.05	0.200	0.280
29	26	0.525	12.71	0.300	0.272
30	27	0.650	11.42	0.700	0.738
31	28	0.350	14.58	0.400	0.428
32	29	0.600	11.95	0.500	0.570
33	30	0.725	10.57	0.600	0.730
34	31	0.650	11.42	0.800	0.705
35	32	0.800	11.42	0.200	0.387
36*		0.400	14.05	0.100	0.161
37	33	0.225	16.06	0.600	0.617
38	34	0.350	14.58	0.600	0.494
39	35	0.450	13.54	0.200	0.161
40	36	0.400	14.05	0.700	0.631

Items in the pilot study	Items in the main study	Difficulty indices (IF)	Delta	Discrimination Indices (ID)	Biserial Correlation Coefficients
41	37	0.275	15.43	0.300	0.364
42	38	0.400	14.05	0.600	0.607
43*		0.325	14.86	-0.300	-0.359
44	39	0.375	14.31	0.200	0.181
45	40	0.400	14.05	0.200	0.331
Mean		0.541	12.508	0.449	0.476

Note: The following is the suggested level of each index by the program (Classical Test Item Analysis) initially developed by Sukamolson.

- Item Difficulty Index (IF) should be between .20-.80.
- Delta should be close to 13.0.
- Item Discrimination index (ID) and Biserial Correlation Coefficient should be higher than 2.0.

The asterisk (*) refers to item that was deleted and excluded from the main study.

Appendix H: The E-ERT (a main study version)

Engineering-English Reading Test for Thai Graduate Students

Instructions:

1. The test consists of 2 parts with total 5 sections (excluding the cover).
2. You are allowed about 90 minutes to complete the test.
3. Complete all answers in the answer sheet provided.
4. Do not write anything into the test paper.

This test paper consists of the following parts

Part	Item numbers	Scores
Part I: Vocabulary	1-10	10
	11-19	9
Part II: Reading comprehension	20-25	6
	26-32	7
	33-45	8
	40 items	**40 points**

The Test Designer:
Natjiree Jaturapitakkul
EIL Program, Chulalongkorn University
Semester 2/ 2007

Part I: Vocabulary: sub-technical (Questions 1-10)

a) Instructions: Put a cross (X) on the correct answers (a, b, c or d) in the answer sheet provided.

1. A _______________ may be thought of as any influence which tends to change the motion of an object as a lift, a push, or a pull, and has a magnitude and a direction.
   a. force* b. shear c. stress d. velocity

2. The amount of energy transferred by a force is known as _______________.
   a. element b. mass c. move d. work*

3. The internal distribution of force per unit area that balances and reacts to external loads applied to a body is generally called _______________.
   a. energy b. entropy c. figure d. stress*

4. The _______________ is a quantity that represents the magnitude of force applied to a rotational system at a distance from the axis of rotation.
   a. compression b. moment* c. statics d. tension

5. The deformation of a material substance in which parallel internal surfaces slide past one another is called _______________.
   a. flow b. distortion c. shearing* d. volume

6. The _______________ refers to the state of a body at rest or in uniform motion in which the resultant of all forces on it is zero.
   a. equilibrium* b. magnitude c. matrix d. origin

7. The force that opposes the relative motion or tendency of such motion of two surfaces in contact is known as _______________.
   a. dynamics b. friction* c. transaction d. viscosity

8. The _______________ is speed that has a clearly stated direction.
   a. act b. flow c. momentum d. velocity*

9. An object defined by both magnitude and direction is called _______________.
   a. elasticity b. placement c. vector* d. vortex

10. A/ An _______________ is a fundamental two-dimensional object. Intuitively, it may be visualized as a flat infinite sheet of paper.
    a. area b. circle c. line d. plane*

Part I: Vocabulary: non-technical (Questions 11-19)

b) Instructions: Choose the best answer which is closest in meaning to the word underlined. Put a cross on the correct answers (a, b, c or d) in the answer sheet provided.

Relativity Passes Absolute Test

According to Albert Einstein's general theory of relativity, a massive **11)**body like Earth should bend the space-time fabric of the universe, causing it to curve and flex like a trampoline supporting a bowling ball.

Nearly three years ago, NASA's oft-canceled $750 million Gravity Probe B Relativity Mission finally shot into space with one goal—to quantify Einstein's **12)**predictions from Earth's orbit. Earlier this year, at the meeting of the American Physics Society, principal investigator Francis Everitt **13)**delivered the first results: Gravity Probe B has verified Einstein's theory to within 1 percent.

Four gyroscopes, each the size of a Ping-Pong ball, form the heart of the experiment. The gyroscopes are the most perfectly **14)**spherical man-made objects in existence; if inflated to the size of Earth, they would have mountains no more than eight feet high. (Their near-faultless roundness has landed the spheres in the Guinness World Records.) At the beginning of the experiment, the gyroscopes' **15)**axes pointed to a distant star; as the spacecraft moved around Earth for nearly a year, the researchers carefully monitored the position of the axes.

Einstein's theory predicts that the axes should shift by a **16)**tiny amount—0.0018 degree—under the influence of Earth's pull on space-time. After 18 months of data analysis, Everitt and his team measured the axial shift to within 1 percent of Einstein's prediction. Everitt, a Stanford physicist who has spent more than 40 years on the project, says the **17)**results are sweet indeed. "It's really extraordinary to look at the output and see Einstein looking back, without any 1**8)**calculations or corrections," he says. "This measurement is unprecedented in any test of general relativity."

In addition to claiming that the universe curves around massive bodies, Einstein said that as these bodies **19)**rotate, they effectively "drag" space behind them, creating a twist in the cosmic fabric. Everitt says his team plans to announce verification of this "frame dragging" effect later this year.

11.	a. content	b. figure	c. group	d. object*
12.	a. analyses	b. calculations	c. evaluations	d. forecasts*
13.	a. declared*	b. handled	c. made	d. showed
14.	a. cylindrical	b. oval	c. round*	d. square
15.	a. arrow lines	b. central lines*	c. dash lines	d. dot lines
16.	a. decimal	b. huge	c. very low	d very small*
17.	a. numbers	b. outcomes*	c. quantities	d. truths
18.	a. computations*	b. estimations	c. practices	d. predictions
19.	a. move	b. spin*	c. swing	d. twist

Part II: Reading comprehension

Instructions: Read the passages and choose the one best answer for each question by putting a cross (X) in the answer sheet provided.

Passage 1 (Questions 20-25)

The following is an article about engineering projects on building.

Hong Kong's Central Plaza building, its 306.5 m-high superstructure making it the world's tallest reinforced-concrete building, was topped out in April 1992. A steel mast would bring the total height to 374 m. The building was an equilateral triangle in shape with each side approximately 50 m in length. The floors extended from closely spaced perimeter columns and around a central core area of elevators and service ducts. The use of concrete for such a tall building was made possible by the employment of high-strength concrete. There were three stories below ground, and under those the building rested on deep caissons up to 7.4 m in diameter.

In an advanced stage of planning during 1992 was the Tour Sans Fins in Paris. This was to be 426 m in height with a circular plan form 43 m in diameter. This caused it to be very slender in terms of building structures and, while the circular plan shape was efficient from the point of view of wind drag, it introduced the problem of wind vortex excitation. This phenomenon arises with rounded shapes and is caused by the periodic shedding of the vortices from alternate sides of the cylinder. At a certain wind speed this shedding can be at the same frequency as that of the building, causing the system to reach a high level of oscillation.

Concrete was being chosen for the structure on account of its efficient stiffness and damping properties (properties that diminish the amplitude of an oscillation) compared with those of steel. Even so, accelerations would likely exceed acceptable limits, and a damper at the top of the building was being considered. This would comprise a pendulum with a mass of only a fraction of that of the building and with a predetermined damping factor introduced between it and the structure. The Tour Sans Fins was to be a peripheral pierced and framed tube with no central core, thereby concentrating the load-carrying members in the area where they would contribute most to the strength and stiffness of the structure.

Another building of significant engineering interest was the Hotel delas Artes Tower in Barcelona, Spain, completed during 1992. The main structure was a 45-story, 135 m-high hotel-apartment tower with an externally expressed structural-steel frame. One of the main design features in a steel-framed building, after strength and stability, is fire resistance. Normally this is achieved by sprayed-on fire protection and/or cladding and encasement. In this case, however, the designers predicted that by having the structural members at least 1.5 m from the window wall, the steel, in the event of fire breaking through the façade, would not overheat. This feature was in addition to careful fire compartmenting within the building to reduce the risk of fire spread. The structural concept was bold. The tower was 30 m square in plan and each elevation was divided into three bays, the outer ones being cross braced for the full height of the tower; at three levels the central bay was also cross braced, which significantly increased the stiffness of the building and reduced wind movement. Vertical loads were transferred from the floors to the external frame by fire-protected steel beams, four on each elevation. Detailed studies had to be made to predict the likely temperatures that the steel would reach in the event of a fire and to check that the stresses and deflections arising from those temperatures would be acceptable.

20. What is the main idea of the passage?
    a. The comparison between three buildings.
    b. The importance of concrete in building construction.
    c. The buildings with different shapes around the world.
    d. The prominent buildings around the world in the year 1992.*

21. According to a passage, what is a benefit of a circular plan shape of the Tour Sans Fins?
    a. Reduce wind drag*
    b. Increase wind vortex excitation
    c. Increase stiffness of the building
    d. Reduce the amplitude of an oscillation

22. From the passage, when the accelerations of the building exceed the acceptable limit, what do they do to solve this problem??
    a. Installing a pendulum with a mass in the building.*
    b. Installing deep caissons in the building.
    c. Installing steel mast in the building.
    d. Making the building higher.

23. It can be implied from the passage that the following fundamental considerations should be included in building construction EXCEPT
    a. steel*
    b. stability
    c. stiffness
    d. strength

24. Hong Kong’s Central Plaza building is in 1992 the world’s tallest reinforced-concrete building. What is a mixture of reinforced concrete?
    a. concrete with sand inside
    b. concrete with stones inside
    c. concrete with steel bar inside*
    d. concrete with stainless steel inside

25. Hong Kong’s Central Plaza building was constructed by using high strength concrete. What might be a reason for using high strength concrete?
    a. High strength concrete is stronger than steel.
    b. High strength concrete uses a special kind of cement for tall buildings.
    c. High strength concrete helps reduce building’s weight but resist higher load.*
    d. High strength concrete is easier to produce and to use to construct the building.

Passage 2 (Questions 26–32)

Here is a part of the text "Vector mechanics for engineers statics".

Mechanics may be defined as that science which describes and predicts the conditions of rest or motion of bodies under the action of forces. It is divided into three parts: mechanics of rigid bodies, mechanics of deformable bodies and mechanics of fluids.

The mechanics of rigid bodies is subdivided into statics and dynamics, the former dealing with bodies at rest, the latter with bodies in motion. In this part of the study of mechanics, bodies are assumed to be perfectly rigid. Actual structures and machines, however, are never absolutely rigid and deform under the loads to which they are subjected. But these deformations are usually small and do not appreciably affect the conditions of equilibrium or motion of the structure under consideration. They are important, though, as far as the resistance of the structure to failure is concerned and are studied in mechanics of materials, which is a part of the mechanics of deformable bodies. The third division of mechanics, the mechanics of fluids, is subdivided into the study if incompressible fluids and of compressible fluids. An important subdivision of the study of incompressible fluids is hydraulics, which deals with problems involving liquids.

Mechanics is a physical science, since it deals with the study of physical phenomena. However, some associate mechanics with mathematics, while many consider it as an engineering subject. Both these views are justified in part. Mechanics is the foundation of most engineering sciences and is an indispensable prerequisite to their study. However, it does not have the empiricism found in some engineering sciences, i.e. it does not rely on experience or observation alone; by its rigor and the emphasis it places on deductive/ reasoning it resembles mathematics. But, again, it is not an abstract or even a pure science; mechanics is an applied science. The purpose of mechanic is to explain and predict physical phenomena and thus to lay the foundations for engineering applications.

Fundamental concepts: Although the study of mechanics goes back to the time of Aristotle (384-322 B.C.) and Archimedes (287-212 B.C.), one has to wait until Newton (1642-1727) to find a satisfactory formulation of its fundamental principles. These principles were later expressed in a modified form by D'Alembert, Lagrange, and Hamilton. Their validity remained unchallenged, however, until Einstein formulated his theory of relativity (1905). While its limitations have now been recognized, newtonian mechanics still remains the basis of today's engineering sciences.

The basic concepts used in mechanics are *space, time, mass* and *force*. These concepts cannot be truly defined; they should be accepted on the basis of our intuition and experience and used as a mental frame of reference for our study of mechanics

The concept of *space* is associated with the notion of the position of a point P. The position of P may be defined by three lengths measured from a certain reference point, or origin, in three given directions. These lengths are known as the coordinates of P.

To define an event, it is not sufficient to indicate its position in space. The *time* of the event should also be given.

The concept of *mass* is used to characterize and compare bodies on the basis of certain fundamental mechanical experiments. Two bodies of the same mass, for example, will be attracted by the earth in the same manner; they will also offer the same resistance to a change in translational motion.

A force represents the action of one body on another. It may be exerted by actual contact or at a distance, as in the case of gravitational forces and magnetic forces. A force is characterized by its point of application, its magnitude, and its direction; a force is represented by a vector.

26. What is the passage primarily about?
    a. Introduction to mechanics and its types
    b. Introduction to mechanics and its history
    c. Introduction to mechanics and its concepts*
    d. Introduction to mechanics and its principles

27. In paragraph 2, which of the following statements is TRUE?
    a. Statics deals with mean and S.D.
    b. Statics deals with unmoved bodies.*
    c. Dynamics deals with space and mass.
    d. Dynamics deals with immovable bodies.

28. In paragraph 3, which of the following statements about mechanics is FALSE?
    a. Mechanics is related to mathematics.
    b. It is necessary for engineering students to study mechanics.
    c. Studying mechanics does not require experience or observation, but reasons.
    d. Mechanics focuses on giving specific reasons and then linking to general ideas.*

29. From the passage, fundamental concepts and principles of mechanics was laid by….
    a. Newton*
    b. Einstein
    c. Aristotle
    d. D'Alembert, and Lagrange

30. According to the passage, a force that is exerted at a distance is called….
    a. axial force
    b. shear force
    c. body force
    d. gravitational force *

31. It can be inferred that this passage is chosen from which of the following source?
    a. Journal
    b. Text book *
    c. Science report
    d. Research article

32. What is a condition of equilibrium?
    a. Summation of all forces equal to zero*
    b. Summation of all forces equal to one
    c. All forces in x-direction equal to zero
    d. All forces in x-direction equal to one

Passage 3 (Questions 33-40)

This is a part of the engineering experiment concerning "Idea feasibility testing".

Demise of the Pony Express: Idea feasibility testing

In the response to the Gold Rush in the mid-1800s, the Pony Express was developed as an alternative mail service crossing the North American continent. Using mounted riders, the founders of the Pony Express hoped to create a conduit for mail that was faster and more reliable than stagecoaches. The Pony Express only operated for a year and a half—from April 1860 to November 1861—even though it offered excellent wages and employee benefits. The demise of the Pony Express is commonly attributed to the telegraph. The completion of the telegraph network and the termination of the Express occurred within a one-month period. But there is another, lesser-known myth (unknown, actually) that a different technology was the mail in the

Express's coffin. Letters were placed inside baseballs and thrown along a relay of ball carriers. This new system was faster than using horses, and the ball carriers were less tired than the Pony Express riders, who had to endure long days in the saddle. All a ball carrier had to do was to hang out and play catch. Can a ball relay beat a horse? Students applied their estimation skills and rapid idea testing processes to find an answer.

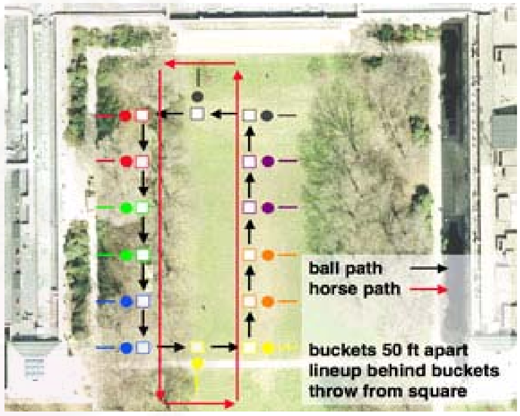
Some numbers: the estimation

Here is how the course instructor approached the problem. Assuming an average ball carrier can throw a 40 mph ball at 100-foot throwing distance and catch-throw in 2 seconds. The ball relay is expected to be a blazing 20 mph, roughly 2 times the Express's average speed of 10 mph. However, just one drop in 10 might allow the express to win. With strong-armed ball carriers, the relay might even be competitive with a horse running at 30 mph. A validation experiment, was in order.

Resolving the debate: Procedures

The proving ground was MIT's Killian Court. After a 10-minute warm-up and practice session, the validation experiment was completed under the watchful eye of an impartial cowboy. The GPS-equipped Pony Express rider was Tom, the teaching assistant, on his horse headed mountain bike. 14 ball-throwing stations, one for each lab section consisting of 7-8 students, were set up in a rectangular loop. 50 feet was determined to be a manageable throwing distance for the average, non-athletic MIT arm, separated each station.

The Express rider

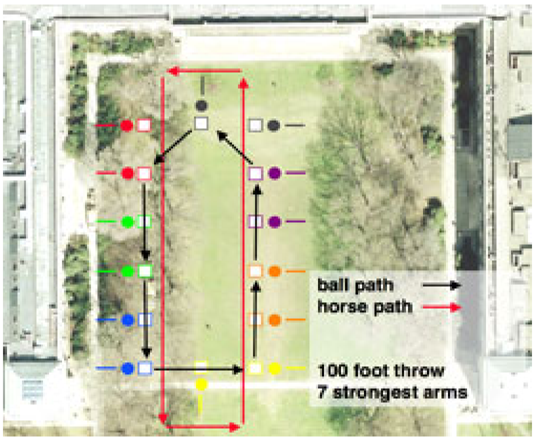


Test setup overview

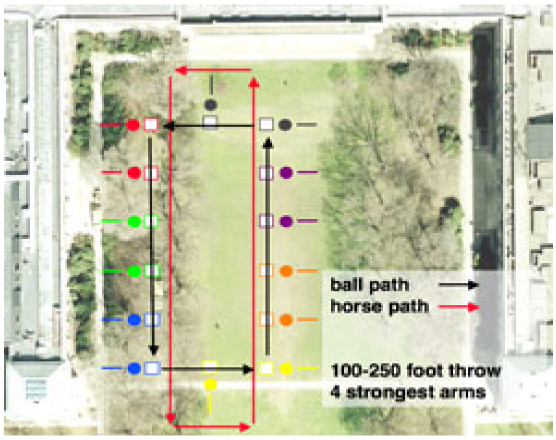
In the first test, all students were ball-carriers regardless of throwing experience, and the baseball completed 8 loops around the courtyard. This allowed each student at all the ball-throwing stations to throw at least once. The rider reined his horse at the Express's average speed of 10 mph.

During the second run, two laps were completed with only the strongest-armed ball-carriers throwing from each station, still spaced at 50 feet. The Express rider maintained a speedy 15 mph. Over-throws and missed catches led to a dramatic finish — with the ball just overtaking the Express rider as he crossed the finish line.

The third challenge matched the 100-foot throwing distance assumed in the estimation calculations. Seven strong throwers threw to every second station and the rider rode his horse as fast as possible, competing over 2 loops around the courtyard. Finally, two laps were completed with the 4 strongest arms in the class on the circuit's corners. The winded horse struggled to keep pace.



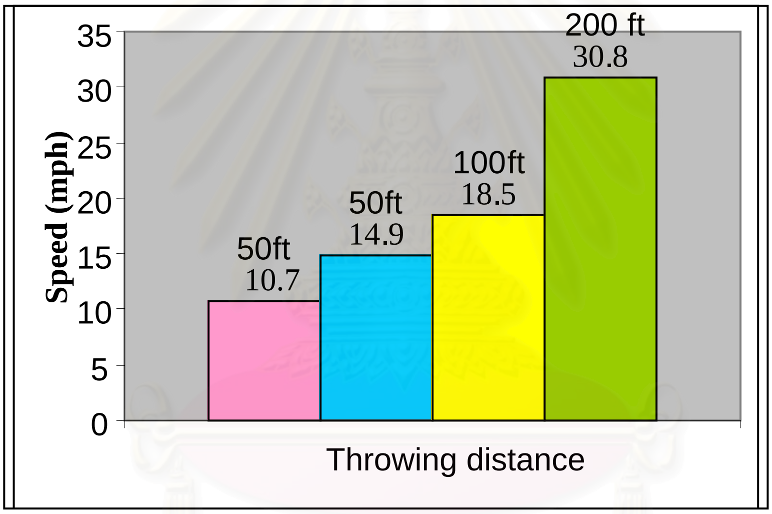
100 foot throw test layout



200 foot throw test layout

Results

Despite suffering from more balls dropped than caught (53% drop rate), the first test still had an average speed of 10.7 mph—just 7% faster than the reported average speed of the Pony Express.



Average speed of the four tests

Under the 100 ft throw conditions, the predictions were born-out in reality. A relay speed of 20 mph was estimated and it was achieved an actual speed of 18.5 mph—an unexpected result given the ball drop rate was 40%. The average drop recovery was under 2 seconds, 10 times less than allowed for in the estimation.

Conclusion: Myth technically feasible! but practically ridiculous.

33. What is the main point of the passage?

a. A competition between a horse rider and ball carriers.
b. An experiment to find out whether or not a ball relay can beat a horse.*
c. A competition on mail service between the Pony Express and the telegraph network.
d. An experiment to find out about the speed of the ball relay with different throwing distance.

34. According to the passage, which of the following is TRUE about the Pony Express?
    a. It was established because of population growth in the US.
    b. It was a mail service by using a mounted rider as a postman.*
    c. It was faster and more reliable than other kinds of mail service in the North America.
    d. It was operated for a long period because it offered high salary and employee benefits.

35. From the passage, what could be a main cause of the end of the Pony Express?
    a. The telegraph network *
    b. The relay of ball carriers
    c. The exhaustion of the mounted riders
    d. The high wages and employee benefits

36. According to the passage, the experimental results were affected by which of the following factors?
    a. Gender and Catch-throw time
    b. Gender, catch-throw time and throwing distance
    c. Catch-throw time, throwing experience and throwing distance
    d. Catch-throw time, drop cost, throwing experience and throwing distance *

37. Which of the following statements is NOT inferred as assumptions of the experiment?
    a. The increasing throwing distance can increase the ball speed.
    b. The more balls dropped can allow the horse rider to win.
    c. The more skilled thrower can increase the ball speed.
    d. Force due to wind can decrease the ball speed.*

38. From the experiment, the estimated ball speed is 20 mph. In this case, how long does it take for a ball to travel one mile excluding drop time?
    a. 1 minute
    b. 2 minutes
    c. 3 minutes*
    d. 4 minutes

39. Which of the following statements about the results is FALSE?
    a. The more throwing distance, the more speed of the throwing relay.
    b. The more throwing experience, the more speed of the throwing relay.
    c. Throwing distance plays more important role than throwing experience when over the 100 foot throw.
    d. Throwing experience seems to have more effects on the relay speed than throwing distance when over the 100 foot throw.*

40. Which is the best conclusion of this passage?
    a. It is proved that a ball relay beats a horse.
    b. It is possible to have a mail service by using a ball relay.
    c. Although the relay speed is satisfied, it is not practical to do.*
    d. Only a horse rider is more practical than a lot of ball carriers.

****This is the end of the test****

Appendix I: Interview questions for test taking strategies of the test takers

In the interview session, the researcher asked quite the same pattern of questions to interviewees. The interviewees were given an opportunity to express themselves first as well as to illustrate the strategies they employed in some items. Then, some probing questions which were adapted from test taking strategies in the literature review (Cohen, 1998) were delivered to them in order to get more details. The following are a list of interview questions asked in the main study.

By focusing on each part of the test, what kind of test taking strategies do you use in answering each type of test items? Please elaborate type by type.

Part 1: A > Sub-technical terms

1. What strategies do you employ in answering sub-technical term items?
2. Can you tell me how you could arrive at an answer for questions 1 and 2?
3. Do you make use of key words?
4. Do you guess from context clues?
5. Do you translate a whole sentence?
6. Do you take advantage of clues appearing in other items?
7. What else of test taking strategies do you employ?

Part 1: B > Non-technical vocabulary

1. What strategies do you employ in answering non-technical vocabulary items?
2. Can you tell me how you could arrive at an answer for questions 11 and 14?
3. Do you make use of key words?
4. Do you guess from context clues?
5. Do you translate a whole sentence?
6. Do you use background knowledge in recognizing the meaning of the words?
7. What else of test taking strategies do you employ?

Part 2: Passages 1-3

Main idea

1. What strategies do you employ in answering main idea items?
2. Can you tell me how you could arrive at an answer for questions 20 and 26?
3. Do you skim for general idea of the text?
4. Do you read every single word?
5. Do you work on main idea item before doing other items?
6. What else of test taking strategies do you employ?

Specific detail

1. What strategies do you employ in answering specific detail items?
2. Can you tell me how you could arrive at an answer for questions 21 and 29?
3. Do you use scanning technique?
4. Do you focus on key words in a question before scanning for details?
5. Do you use an eliminating option strategy? How do you eliminate those choices?
6. What else of test taking strategies do you employ?

Inference

1. What strategies do you employ in answering inference items?
2. Can you tell me how you could arrive at an answer for questions 23 and 31?
3. Do you read for details and then analyze the information?
4. Do you employ an eliminating option technique? If so, how?
5. Do you return to the text passage to look for the answer?
6. What else of test taking strategies do you employ?

Fact

1. What strategies do you employ in answering inference items?
2. Can you tell me how you could arrive at an answer for questions 24 and 32?
3. Do you use engineering background knowledge?
4. Do you make of the world knowledge?
5. Can you produce your own answer before looking at the options?
6. What else of test taking strategies do you employ?

Appendix J: Questionnaire for Test Takers (Thai version)

แบบสอบถามความคิดเห็นเกี่ยวกับ

Engineering-English Reading Test (E-ERT) for Thai Graduate Students

ชื่อ (นาย / นางสาว) ....เลขประจำตัว..

คณะ.. สาขา ....ปีที่.......................

สาขาวิชา/คณะที่เรียนในระดับปริญญาตรี ....

เวลาในการสอบ เริ่ม .... เสร็จ ....

ส่วนที่ 1 : ข้อมูลด้านภูมิหลังความรู้ของท่าน

ในการเรียนของท่านก่อนหน้านี้ ขณะนี้ หรือในยามว่าง ท่านเคยอ่านหนังสือ นิตยสาร บทความวิชาการหรือบทความในหนังสือพิมพ์ในหัวข้อดังต่อไปนี้บ้างหรือไม่ กรุณาใส่เครื่องหมาย √ ในช่องที่ท่านเลือกตามเกณฑ์ดังต่อไปนี้

3 = บ่อยครั้ง, **2** = บางครั้ง, **1** = ไม่เคย

หัวข้อ	เกณฑ์ 3	เกณฑ์ 2	เกณฑ์ 1
Engineering Mechanics (กลศาสตร์วิศวกรรม)			
Engineering Materials (วัสดุวิศวกรรม)			
Mechanics of Materials (กลศาสตร์วัสดุ)			
Mechanics of Fluids (กลศาสตร์ของเหลว)			
Engineers Statics (วิศวกรรมสถิตศาสตร์)			
Statistics for Engineers (สถิติสำหรับวิศวกร)			
Mathematics (คณิตศาสตร์)			
Science and Technology (วิทยาศาสตร์และเทคโนโลยี)			
Computer Science (วิทยาการคอมพิวเตอร์)			
Education (การศึกษา)			
Business (ธุรกิจ)			
Literature (วรรณคดี)			
Others (อื่นๆ โปรดระบุ) _________________			

ส่วนที่ 2 : ความคิดเห็นของผู้ตอบแบบสอบถามที่มีต่อข้อสอบ E-ERT

กรุณาใส่เครื่องหมาย √ ในช่องที่ท่านเลือกตอบตามเกณฑ์ระดับความคิดเห็นต่อไปนี้

4 = มากที่สุด, **3** = มาก , **2** = น้อย , **1** = น้อยที่สุด

		ระดับความคิดเห็น			
ลำดับ	คำถาม	4	3	2	1
1	คำสั่งที่ใช้ในข้อสอบชัดเจน				
2	ลักษณะและขนาดของตัวอักษรที่ใช้ในข้อสอบเหมาะสม				
3	กราฟและตารางที่ใช้ในข้อสอบชัดเจน				
4	กราฟและตารางที่ใช้ในข้อสอบเหมาะสมกับข้อสอบ				
5	จำนวนข้อในข้อสอบมีความเหมาะสม				
6	ระยะเวลาที่ใช้ในการสอบมีความเหมาะสม				
7	ระดับความยากง่ายของข้อสอบ E-ERT มีความเหมาะสม				
8	ท่านพอใจกับรูปแบบโดยรวมของข้อสอบ E-ERT				
9	การสอบ E-ERT มีประโยชน์ต่อการเรียนและการพัฒนาภาษาอังกฤษของท่าน				
10	ท่านคิดว่า เนื้อหาของข้อสอบมีความคล้ายคลึงกับการอ่านภาษาอังกฤษในสาขาวิศวกรรมศาสตร์ในสถานการณ์จริง				
11	รูปแบบของข้อสอบมีความคล้ายคลึงกับการอ่านภาษาอังกฤษในสาขาวิศวกรรมศาสตร์ในสถานการณ์จริง				

12. โปรดแสดงความคิดเห็นเกี่ยวกับข้อดี และ ข้อควรปรับปรุงของข้อสอบ E-ERT

ข้อดี

....

....

....

ข้อควรปรับปรุง

....

....

....

ขอขอบคุณในความร่วมมือของท่าน

Appendix K: Questionnaire for Test Takers (English version)

Engineering-English Reading Test (E-ERT) for Thai Graduate Students

Name (Mr./ Ms.)... ID.No. ....

Faculty...Major...Year......................................

Field of study in Bachelor's degree……………………………………………………………………….....

Test time Begin at ………………………………………… Finish at ……………………………...………

Part 1: Background Knowledge

1. Think about the reading you do for your previous/ present study or during your spare time. Do you read books, magazines, academic papers or newspaper articles on any of the following topics? Please put √ in the space that corresponds to your background knowledge by using the following scale:

3 = Often, **2** = Sometimes, **1** = Never

	Scale		
Topics	**3**	**2**	**1**
Engineering Mechanics			
Engineering Materials			
Mechanics of Materials			
Mechanics of Fluids			
Engineers Statics			
Statistics for Engineers			
Mathematics			
Science and Technology			
Computer Science			
Education			
Business			
Literature			
Others (please specify) ________________			

Part 2: Engineering-English Reading Test (E-ERT)

Please put √ in the space that corresponds to your opinion by using the following scale:

4 = Strongly agree, **3** = Agree , **2** = Disagree , **1** = Strongly disagree

		Scale			
Item	Statement	4	3	2	1
1	Instructions of the test are clear.				
2	Typeface and size of letters in the test are appropriate.				
3	Graph, layouts and pictures in the test are clear.				
4	Graphs, layout and pictures in the test are appropriate.				
5	The number of questions in the test is appropriate.				
6	Time allotment for the test is appropriate.				
7	Level of difficulty of the test is appropriate for graduate students.				
8	I am satisfied with the test in general.				
9	The test is useful for my English language learning and development.				
10	The content of the test is similar to the English reading for Engineering in real situations.				
11	The format of the test is appropriate for English reading test for Engineering in real situations.				

12. Please comment on the strong points and areas for improvement of the E-ERT
Strong points

....

....

....

Areas for improvement

....

....

....

Thank you for your kind co-operation

Appendix L: Statistical procedures used in this study

In order to select an appropriate statistical test for this research study, the following is a list of the criteria that is taken into consideration:

1. The objectives of research
2. Research questions/ hypotheses of a study
3. Research design
4. Basic assumptions of the selected statistical tests

Typically, in conducting a research, there must be a plan. In planning a good research, many steps are included in a research plan. Some steps of the research plan such as stating objectives of a research, setting research questions or hypothesis and designing research procedures can help determine the statistical analysis. Briefly explained, objectives of a research usually go hand in hand with research questions and hypotheses of the study. When research questions and hypotheses of a (quantitative) study are set, they can determine the nature of research design, which in turn determines the statistical analysis or statistical tests appropriate for the study. In case of a quantitative research, some key words in the hypotheses can help determine the statistical tests. For example, "There is a significant interaction effect between language ability, background knowledge and ESP reading scores. From this directional hypothesis, some words like *interaction effect* can help indicate to use a two-way ANOVA since it concerns an interaction between two independent variables.

Apart from this, which statistical test should be selected also depends on a number of factors as suggested by Gay & Airasian (2000) as follows:

- how the groups will be formed (for example, by random assignment, by matching, or by existing groups)
- how many different treatment groups will be involved
- how many variables will be involved
- and the kind of data to be collected (e.g. counts of the numbers of times, test scores, or place students into categories)

In addition, the basic assumptions of each statistical test should be taken into consideration. Generally speaking, the statistical procedures used to analyze data have assumptions that underlie the statistics. For example, most parametric statistical tests assume an underlying normal distribution and that each participant's scores are independent of any other participants' scores. If these assumptions are not met, bias enters into the statistics used, then weakening research generalization.

2.9.1 A logical sequence in selecting the statistical tests

Regarding this research study, two kinds of statistical tests are used: a 2-way ANOVA and partial Eta squared. The following shows a logical sequence in selecting these tests.

Step 1: Identify the objectives of the study in terms of observable behavior.

Step 2: Set research questions and hypotheses corresponding to the objectives of the study.

Step 3: Design research procedures including subjects, sampling technique, instrument used, data collection and data analysis.

Step 4: Determine the statistical tests based on research design and suggestions by Gay & Airasian (2000).

Step 5: For an interaction effect question in this study, "Is there any significant interaction effect between language ability and engineering background knowledge on ESP reading ability? If there is, what is its effect size? " A two-way ANOVA is selected to be used in this study since it concerns an interaction effect between variables and there are two independent variables with two levels in each variable. Before using ANOVA, the data must be checked if it meets the basic assumptions of the test. In addition, partial Eta squared is chosen in order to measure the effect sizes of the treatment.

Step 6: When a significant interaction is presented in a two-way ANOVA, comparisons of the main effects are inappropriate (Hatch & Farhady, 1982: 158). A comparison of single main effects should be made. Then, comparing two means of each main effect (background knowledge: engineering and non-engineering and language ability: high and low) was conducted

In this research study, it is necessary to use a two-way ANOVA. This is because this research aims at investigating the interaction effect between language ability and background knowledge on ESP reading scores. Then, the two-way ANOVA could help to exhibit the interactions between these two independent variables with each has two levels. It also allows making reasonable conclusions about the performances of groups of students on the reading test. The next part presents the detail of two-way ANOVA.

2.9.2 Two-way Analysis of Variance (ANOVA)

Introduction to Two-Way ANOVA

A more sophisticated ANOVA model that allows the researcher to test the effectiveness of two independent variables is called Two-Way ANOVA (sometimes also called Factorial ANOVA). That is, two-way ANOVA improves on one-way ANOVA in that the researcher can simultaneously assess the effects of two independent variables on a single dependent variable within the same analysis. Thus, two-way ANOVA yields the same information that one-way ANOVA would, but it does so in one analysis.

In addition, two-way ANOVA allows the investigator to determine the possible combined effects of the independent variables (Arkkelin, 2003). That is, it also assesses the ways in which these variables interact with one another to influence scores on the dependent variable. Although understanding such interaction effects can be a complex and difficult task, it is essential since in the real world many variables interact with one another to determine behavior.

Two-way ANOVA generates three F-values: one to test the **main effects** of each variable, and a third to test the **interaction effect** (when two IVs are considered simultaneously**)** which may or may not be significant. To elaborate on each type of effects, the following are details.

A main effect refers to the effect that one independent variable has on the dependent variable holding the effects of the other variables constantly. Specifically, a main effect represents a special form of the between-groups variance of a single-independent variable. In a two-way ANOVA, there are two main effects, one for each factor. When the data was examined by using an ANOVA, each main effect can be either statistically significant or not statistically significant.

An interaction effect indicates that the effect of one variable is not consistent across all levels of the other variables. That is, the relationship between one independent variable is different at different levels of the other variable.

When interpreting interaction and main effects in ANOVA, many texts including Ray (p. 198) cited in University of New England (2000) stipulates that the interaction should be interpreted first. If the interaction is not significant, the main effects can then be examined without needing to qualify the main effects because of the interaction. If the interaction is significant, the main effects cannot be examined because the main effects do not tell the complete story. Most statistics texts follow this line.

Advantages of Two-way ANOVA

Before employing any statistical tests, benefits of them should be recognized. The following are advantages of Two-way ANOVA proposed by Wikibook (2008).

In research using a two-variable design offers many advantages over using a one-variable design. The first advantage is *increased efficiency*. This is because the two-variable design contains all of the elements investigated in one experiment. From this, using two variable design in one experiment is more cost-effective than researching two variables in two experiments (one variable in one experiment).

Another advantage is that *the interaction of the two variables in the design can be analyzed*. This helps us understand how combinations of variables influence behavior. In particular, it allows us to understand and analyze the interactive effects between the two independent variables on the dependent variable. In this, interaction means that the effect of one independent variable is influenced by another independent variable; or, interaction means that the relationship between an independent variable is different at various levels (types) of another independent variable.

The last advantage of using a two-way ANOVA is *an increase in statistical power*. By recalling on statistical "power", it is the ability to confidently reject a false NULL hypothesis. This type of research design increases statistical power because the within groups variance tends to be smaller than the within-group variance of a comparable one-variable study. The smaller the variance, the less fluctuation in measure. Therefore, the smaller the F-ratio, the smaller the confidence interval which means that we are more likely to have chosen a smaller range of possible values. Consequently, greater statistical power in correctly rejecting a false NULL hypothesis is encouraged.

Basic assumptions for ANOVA

As mentioned earlier, all statistical tests have assumptions that underlie the statistics. If these assumptions are not met, bias enters into the statistics used, then weakening research generalization. In this study, two-way ANOVA is selected as one of statistical tests used for data analysis. According to Tabachnick & Fidell (2001: 83-86), the following is the most well-known basic assumptions of two-way ANOVA and methods to test each assumption.

1. Normality of sampling distributions

One assumption of two-way ANOVA is that the sampling distribution of means for each level (or combination of levels) of the IV(s) is normal. The assumption is for the sampling distribution, not the raw scores. Normality of sampling distributions is usually assured by having sufficiently large and relatively equal sample sizes among levels (or combinations of levels) of the IV. The rationale behind hypothesis testing relies on having normally distributed data and so if this assumption is not met then the logic behind hypothesis testing is flawed.

To check normality, a normal Q-Q plot (Field, 2005) could be employed. To deal with normality, tests for skewness and kurtosis can be applied to the raw scores within each group. If sample sizes are unequal or too small and there is excessive skewness and kurtosis, or if outliers are present, data transformation may be necessary to achieve normality of distributions of raw scores within each group.

2. Homogeneity of variance

This assumption means that the variances should be the same throughout the data. In designs like two-way ANOVA in which several groups of participants are tested, this assumption means that each of these samples comes from populations with the same variance.

The Levene’s test could be used to test homogeneity of variance. However, ANOVA is known to be robust to violation of this assumption as long as there are no outliers, sample sizes are large and fairly equal, the sample variances within levels are relatively equal, and a two-tailed hypothesis is tested.

3. Absence of outliers

An outlier is a score that is usually far from the mean of its own group and apparently disconnected from the rest of the scores in the group. Outliers are deviant cases that may have undue impact on the results of analysis. They can raise means or lower means and, by doing so, create artificial significance or cover up real significance. They almost always increase measures of dispersion which makes findings significance less likely. Their conclusions in a data set, in short, makes the outcome of analysis unpredictable and not generalize unless to a population that happens to include the same sort of outliers.

A boxplot could be employed to detect outliers. To deal with outliers, the researcher may remove them from the analysis and endeavour to explain them on a separate basis.

Apart from using two-way ANOVA for this study, partial Eta squared is also conducted as it reveals the meaningful degree of the effect sizes of the treatment. The detail of effect sizes is presented in the next part.

2.9.3 Effect sizes

In reporting the result in an experimental research, researchers should report study findings in a manner that is not only significant but meaningful to their readers. However, statistical significance does not automatically equate to a *meaningful* or *practical* effect (Schuele & Justice: 2006). Some statistically significant effects are meaningful, yet others are not. Because statistical significance and practical significance are often conflated when one interprets research findings (i.e., statistical significance is assumed to establish practical significance), researchers now are asked to explicitly interpret the practical import of statistical results by providing estimates of effect sizes. In this part, the meaning of effect size is firstly explained as introductory information. Then, the reasons to use effect sizes, kinds of them, the meaning of their indexes or how to interpret the magnitude of effect sizes are elaborated respectively. Finally, what kind of effect sizes should be used in this research study is discussed with reasons to substantiate answers.

Definition of effect sizes

According to Becker (2000) and Coe (2000), "effect size" is a name given to a family of *indices that measure the magnitude of a treatment effect* or the effectiveness of the treatment. Effect size is simply a way of *quantifying the difference between two groups or means*. In addition, effect size is a *measure of the strength of the relationship between two variables*: independent and dependent (APA: 2001 cited in Schuele & Justice, 2006**;** Wikipedia: 2007). In scientific experiments, it is often useful to know not only whether an experiment has a statistically significant effect, but also the size of any observed effects. In practical situations, effect sizes are helpful for making decisions. Effect size measures are the common currency of meta-analysis studies that summarise the findings from a specific area of research.

Why using effect sizes

Whereas statistical tests of significance tell us only the likelihood that experimental results differ from chance expectations, effect size measurements tell us the relative magnitude of the experimental treatment. They tell us the size of the experimental effect (Thalheimer and Cook: 2002). In addition, effect sizes are especially important because they allow us to compare the magnitude of experimental treatments from one experiment to another. Although percent improvements can be used to compare experimental treatments to control treatments, such calculations are often difficult to interpret and are almost always impossible to use in fair comparisons across experimental paradigms. Furthermore, unlike significance tests, effect sizes are independent of sample size and alpha. Moreover, effect sizes are considered as a function of statistical power.

Kinds of effect sizes

There are many different types of ES measures, each suites to different research situations. Each ES type may also have multiple methods of computation. In general, there are three major types of effect sizes that measure two independent groups as follows:

1. Standardized mean difference
2. Correlation coefficient
3. Odd-ratio

This part presents details of the three basic ES measures. A scale of magnitudes for differences in means (using the standardized mean difference), linear trends (using the correlation coefficient), and relative frequencies (using odds ratios) are explained and interpreted, respectively.

1. Standardized mean difference

Standardized mean difference is a kind of effect size measuring a scale of magnitudes for differences or changes in means. Effect Size uses the idea of 'standard deviation' to contextualise the difference between the two groups. What we end up with is a number that represents "how many standard deviations" the two groups differ. It represents a standardized group contrast on an inherently continuous measure. Cohen (1988) used the letter *d* to represent the standardized difference or difference between two means and it is often known as *Cohen's d*. The standardized mean difference probably has more methods of calculation than any other effect size type. It provides a method for calculating both t-test and some F-test significance. Basically, it uses the pooled standard deviation (some situations use control group standard deviation) as a method of calculating. For more formulas, see Appendix M.

$$\overline{ES} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}}$$

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

In order to interpret this kind of effect size, there are three ways to employ: as the magnitude of the effect of a treatment on a DV, as the average percentile standing of the average treated (experimental) participant relative to the average untreated (control) participant, and as the percent of nonoverlap of the treated group's scores with those of the untreated group. The following table is usually referred when interpreting the effect size (Becker, 2000).

Table L1: Interpreting the Effect Size Standardized Mean Difference

Cohen's Standard	Effect Size (ES)	Percentile Standing	Percent of Nonoverlap
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0

For the first way of interpretation, Cohen (1988, 1992) defined this kind of effect sizes as *a scale of magnitude for three degrees*: small, moderate, and large are at 0.20, 0.50 and 0.80 respectively. Anything less than 0.2 standard deviations is not worth worrying about.

For the second way of interpretation, effect sizes can also be thought of as the *average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant*. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group.

For the last way, effect sizes can be interpreted in terms of *the percent of nonoverlap of the treated group's scores with those of the untreated group*. An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group, there is 0% of nonoverlap. An ES of 0.8 indicates a nonoverlap of 47.4% in the two distributions. An ES of 1.7 indicates a nonoverlap of 75.4% in the two distributions.

2. Correlation coefficient

The effect size correlation represents the strength of relationship or association between two *inherently continuous* measures. There are many ways to calculate effect size correlation which depend on each research situation. For more formulas, see Appendix N. The results are generally reported directly as “r” (the Pearson product moment coefficient).

$$\overline{ES} = r$$

Pearson's *r* correlation is one of the most widely used effect sizes. It can be used when the data are continuous or binary, thus the Pearson *r* is arguably the most versatile effect size. This was the first important effect size to be developed in statistics, and it was introduced by Karl Pearson. Pearson's *r* can vary in magnitude from -1 to 1, with -1 indicating a perfect negative relationship, 1 indicating a perfect positive relationship, and 0 indicating no relationship between two variables.

For interpreting the effect size correlation, Jacob Cohen has written the most on this topic. In his well-known book he suggested that a correlation of 0.5 is large, 0.3 is moderate, and 0.1 is small (Cohen, 1988). The usual interpretation of this statement is that anything greater than 0.5 is large, 0.5-0.3 is moderate, 0.3-0.1 is small, and anything smaller than 0.1 is insubstantial, trivial, or otherwise not worth worrying about. Any data for which we can calculate a standardized mean difference effect size, we can also calculate a correlation type of effect size. *d* can be converted to *r* and when r is squared, it can be converted to r-squared (r^2) as presented in the following table.

Table L2: Interpreting the Effect Size Correlation

Cohen's Standard	*d*	r	r²
	2.0	.707	.500
	1.9	.689	.474
	1.8	.669	.448
	1.7	.648	.419
	1.6	.625	.390
	1.5	.600	.360
	1.4	.573	.329
	1.3	.545	.297
	1.2	.514	.265
	1.1	.482	.232
	1.0	.447	.200
	0.9	.410	.168
LARGE	0.8	.371	.138
	0.7	.330	.109
	0.6	.287	.083
MEDIUM	0.5	.243	.059
	0.4	.196	.038
	0.3	.148	.022
SMALL	0.2	.100	.010
	0.1	.050	.002
	0.0	.000	.000

For example, the d value of 0.8 corresponds to an r value of .371. An r^2 of 0.138 would suggest that 13.8% of the variance is shared by the two variables.

Apart from the aforementioned effect size correlation (r), there is another kind of correlation measured particularly in Analysis of Variance. Measures of effect size in ANOVA are measures of the degree of association between an effect (e.g., a main effect, an interaction, a linear contrast) and the dependent variable. They can be thought of as the correlation between an effect and the dependent variable. If the value of the measure of association is squared, it can be interpreted as the proportion of variance in the dependent variable that is attributable to each effect. Four of the commonly used measures of effect size in AVOVA are: Eta squared (η^2), partial Eta squared (η_p^2), omega squared (ω^2), and the Intraclass correlation (ρ_I) (Becker: 1998-1999).

1. Eta squared (η^2), is an estimate of the degree of association for the sample. It is the proportion of the total variance that is attributed to an effect. It is calculated as the ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}) --

$$\eta^2 = SS_{effect} / SS_{total}$$

However, one of the problems with η^2 is that the values for an effect are dependent upon the number of other effects and the magnitude of those other effects. For this reason, many people prefer an alternative computational procedure called the partial Eta squared. SPSS reports the partial Eta squared rather than Eta squared. Some authors (e.g., Tabachnick & Fidell, 2001) call partial Eta squared an "alternative" computation of Eta squared.

2. Partial Eta squared (η_p^2), is an estimate of the degree of association for the sample. It is the proportion of the effect + error variance that is attributable to the effect. The formula differs from the Eta squared formula in that the denominator includes the SS_{effect} plus the SS_{error} rather than the SS_{total} --

$$\eta_p^2 = SS_{effect} / (SS_{effect} + SS_{error})$$

Luckily, researchers now can find “partial Eta squared”, from SPSS Program when using the “General Linear Model” for ANOVA testing and also ANCOVA, MANOVA and MANCOVA testing. When interpreting the result, the same idea of magnitude of r can be applied.

One of the problems with η_p^2 is the sums of the partial Eta squared values are not additive. They do not sum to the amount of dependent variable variance accounted for by the independent variables. It is possible for the sums of the partial Eta squared values to be greater than 1.00.

3. Omega squared (ω^2), is an estimate of the dependent variance accounted for by the independent variable in the population for a fixed effects model. The between-subjects, fixed effects, form of the ω^2 formula is --

$$\omega^2 = (SS_{effect} - (df_{effect})(MS_{error})) / MS_{error} + SS_{total}$$

Because η^2 and η_p^2 are sample estimates and ω^2 is a population estimate, ω^2 is always going to be smaller than either η^2 or η_p^2.

4. The intraclass correlation is an estimate of the degree of association between the independent variable and the dependent variable in the population for a random effects model. Because it is for a random effects model, it is not commonly used in psychology experiments. The formula for r_I is --

$$r_I = (MS_{effect} - MS_{error}) / (MS_{effect} + (df_{effect})(MS_{error}))$$

In conclusion, Eta squared and partial Eta squared are estimates of the degree of association for the sample. Omega squared and the intraclass correlation are estimates of the degree of association in the population.

3. Odd-ratio

The Odds-Ratio is a kind of effect size which compares difference in frequencies. It is based on a 2 by 2 contingency table, such as the one below.

Frequencies

	Success	Failure
Treatment Group	*a*	*b*
Control Group	*c*	*d*

$$\overline{ES} = \frac{ad}{bc}$$

The Odds-Ratio is the odds of success in the treatment group relative to the odds of success in the control group. Results are typically reported in one of three forms: frequency of successes in each group, proportion of successes in each group or 2 by 2 contingency table.

The odds ratio is another useful effect size. It is appropriate when both variables are binary. For example (taken from Wikipedia, 2007), consider a study on spelling. In a control group, two students pass the class for every one who fails, so the odds of passing are two to one (or more briefly 2/1 = 2). In the treatment group, six students pass for every one who fails, so the odds of passing are six to one (or 6/1 = 6). The effect size can be computed by noting that the odds of passing in the treatment group are three times higher than in the control group (because 6 divided by 2 is 3). Therefore, the odds ratio is 3. However, odds ratio statistics are on a different scale to Cohen's d. So, this '3' is not comparable to a Cohen's d of '3'.

In summary, there are three major types of effect sizes: standardized mean difference, correlation coefficient and odd ratio. Each kind uses different magnitudes when interpreting the effect size as small, moderate or large. As a good summary, Hopkins (2002) provided a table of scale summary for each kind of effect sizes as follows:

Table L3: Scale Summary for Each Kind of Effect Sizes

	trivial	small	moderate	large	very large	nearly perfect	perfect
Standardized diff.	0.0	0.2	0.6	1.2	2.0	4.0	infinite
Correlation	0.0	0.1	0.3	0.5	0.7	0.9	1
Odds ratio	1.0	1.5	3.5	9.0	32	360	infinite

To focus on this research study, one main kind of effect size that should be used is correlation particularly partial Eta squared (η_p^2). This type of effect size is selected based on statistical tests used in this research study. That is, this study includes a two-way ANOVA. Now reasons on why to select this type of effect size are explained.

For correlation particularly partial Eta squared, it is a special kind of effect size measure in Analysis of Variance which is included in this study. In addition, it is appropriate for factorial design as suggested by Tabachnick and Fidell (2001). It helps estimate the effect size in a sample. Furthermore, it is convenient to compute partial Eta squared by the SPSS program. Actually, Eta squared can also be used as it is a simple way to calculate effect sizes. However, one of problems of the Eta squared is that the values for an effect are dependent upon the number of other effects and the magnitude of those other effects. For this reason, the partial Eta squared is the best choice. However, as the simplest and quickest way, r^2 can be an option.

Appendix M: Different formulas for "standardized mean difference"

The different formulas represent degrees of approximation to the ES value that would be obtained based on the means and standard deviations

Great	Good	Poor
• direct calculation based on means and standard deviations • algebraically equivalent formulas (t-test) • approximations based on continuous data (correlation coefficient)	• estimates of the mean difference (adjusted means, regression B weight, gain score means) • estimates of the pooled standard deviation (gain score standard deviation, one-way ANOVA with 3 or more groups, ANCOVA)	• approximations based on dichotomous data

1. Direct calculation based on means and standard deviations

$$ES = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}} = \frac{\overline{X}_1 - \overline{X}_2}{s_{pooled}}$$

2. Algebraically equivalent formulas (t-test)

2.1 For independent t-test

$$ES = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

2.2 For two-group one-way ANOVA

$$ES = \sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}}$$

Exact *p-values* from a *t*-test or *F*-ratio can be converted into *t*-value and the above formula applied.

3. Approximations based on continuous data (correlation coefficient)

Close Approximation Based on Continuous Data --Point-Biserial Correlation. For example, the correlation between treatment/no treatment and outcome measured
on a continuous scale.

$$ES = \frac{2r}{\sqrt{1-r^2}}$$

4. Estimates of the mean difference (adjusted means, regression B weight, gain score means)

5. Estimates of the pooled standard deviation (gain score standard deviation, one-way ANOVA with 3 or more groups, ANCOVA)

5.1 For standard error of the mean

$$s_{pooled} = se\sqrt{n-1}$$

5.2 For one-way ANOVA> 2 groups

$$s_{pooled} = \sqrt{\frac{MS_{between}}{F}}$$

$$MS_{between} = \frac{\sum \bar{X}_j^2 n_j - \frac{(\sum \bar{X}_j n_j)^2}{\sum n_j}}{k-1}$$

5.3 For standard deviation of gainscores, where *r* is the correlation between pretest and posttest scores

$$s_{pooled} = \frac{s_{gain}}{\sqrt{2(1-r)}}$$

5.4 For ANCOVA, where *r* is the correlation between the covariate and the DV

$$s_{pooled} = \sqrt{\frac{MS_{error}}{1-r^2} \cdot \frac{df_{error}-1}{df_{error}-2}}$$

5.5 For a two-way factorial ANOVA, where B is the irrelevant factor and AB is the interaction between the irrelevant factor and group membership (factor A)

$$s_{pooled} = \sqrt{\frac{SS_B + SS_{AB} + SS_W}{df_B + df_{AB} + df_W}}$$

6. Approximations based on dichotomous data

6.1 For the difference between the probits transformation of the proportion

successful in each group, converts proportion into a *z*-value.

$$ES = probit(p_{group_1}) - probit(p_{group_2})$$

6.2 For chi-square, it must be based on a 2 by 2 contingency table (i.e., have only 1 df)

$$ES = 2\sqrt{\frac{\chi^2}{N - \chi^2}}$$

6.3 For phi coefficient

$$ES = \frac{2r}{\sqrt{1 - r^2}}$$

Appendix N: Different formulas for "effect size correlation"

1. The point-biserial correlation: it is correlation between the dichotomous independent variable and the continuous dependent variable.

$$r_{Y\lambda} = r_{dv,iv}$$

2. Phi correlation: it can be computed from a single degree of freedom Chi Square value by taking the square root of the Chi Square value divided by the number of cases, N.

$$r_{Y\lambda} = \Phi = \sqrt{(X^2(1) / N)}$$

3. The ES correlation can be computed from the *t*-test value.

$$r_{Y\lambda} = \sqrt{[t^2 / (t^2 + df)]}$$

4. The ES correlation can be computed from a single degree of freedom *F* test value (e.g., a oneway analysis of variance with two groups).

$$r_{Y\lambda} = \sqrt{[F(1,\_) / (F(1,\_) + df\ error)]}$$

2. The ES correlation can be computed from Cohen's *d*.

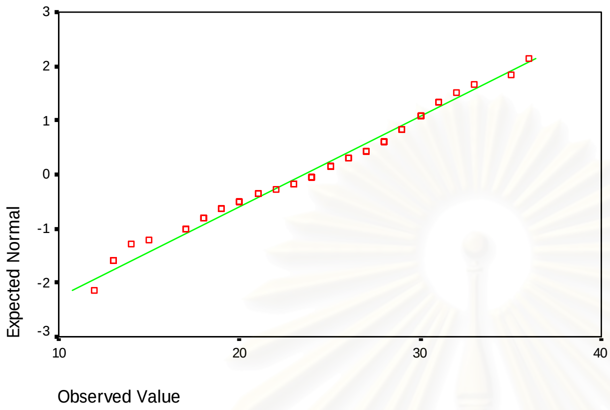
$$r_{Y\lambda} = d / \sqrt{(d^2 + 4)}$$

Appendix O: The Check on Basic Assumptions of Two-way ANOVA

1. Normality of sampling distributions

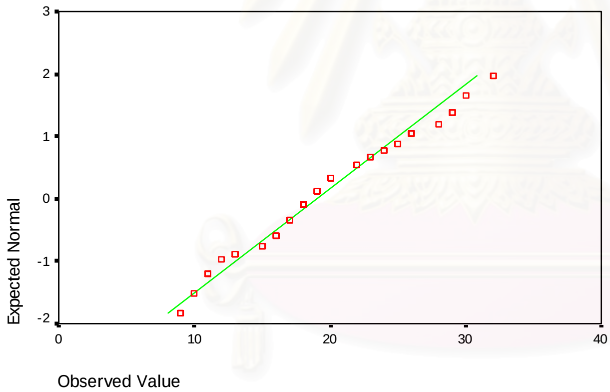
Normal Q-Q Plot

For MAJOR= Engineering



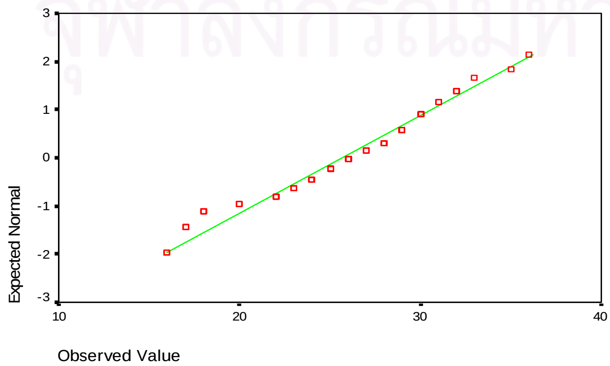
Normal Q-Q Plot

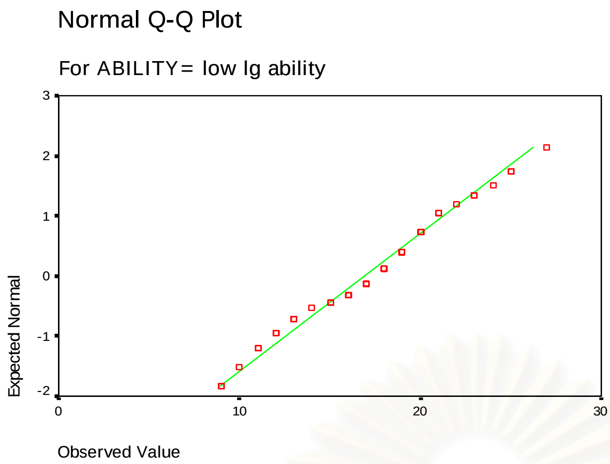
For MAJOR= non-Engineering



Normal Q-Q Plot

For ABILITY= high lg ability





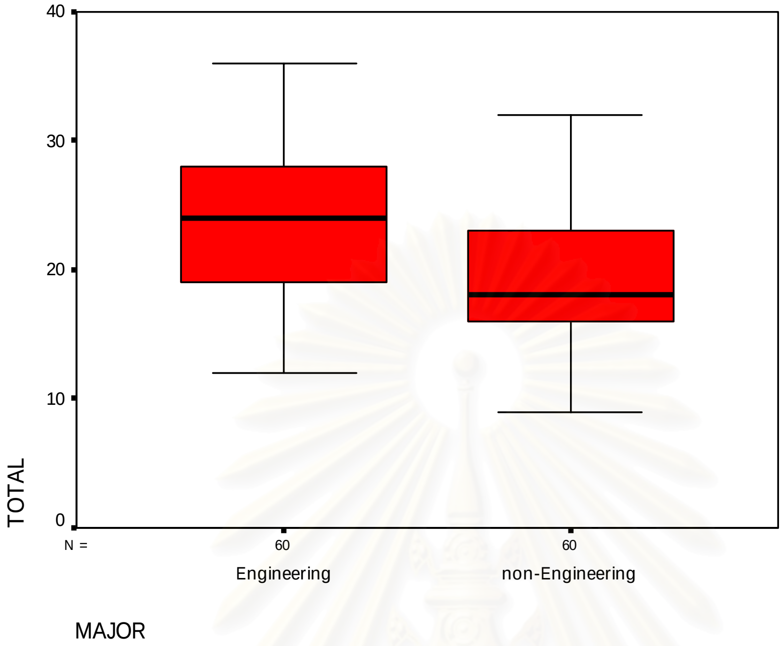
2. Homogeneity of variance

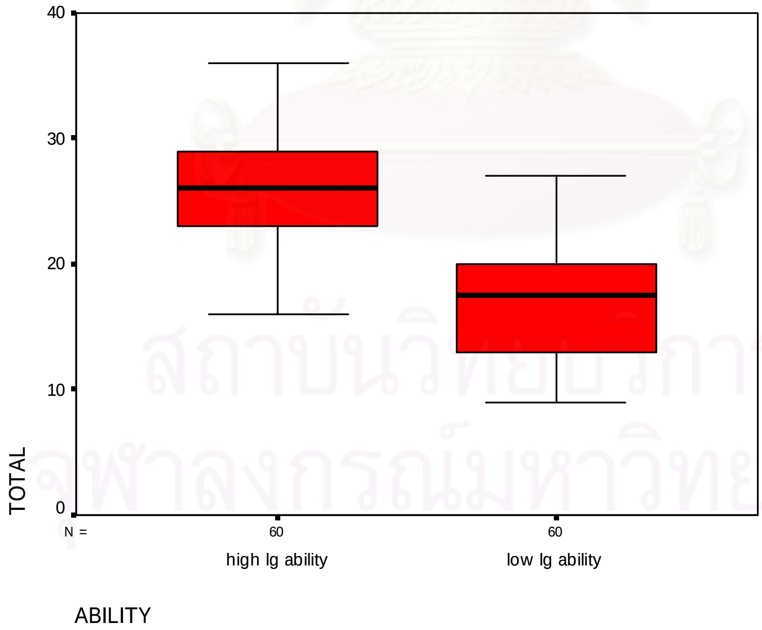
Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
E-ERT scores	Based on Mean	**.231**	1	118	.631
	Based on Median	.207	1	118	.650
	Based on Median and with adjusted df	.207	1	115.900	.650
	Based on trimmed mean	.249	1	118	.619

Focusing on “Based on Mean”, the observed value (.231) was higher than the set value (.05) which means variances are equal. So this assumption is met the requirement.

3. Absence of outliers





BIOGRAPHY

Natjiree Jaturapitakkul was born on September 11, 1977. She graduated with a Bachelor's degree of Education in English (Second class honors) from Silpakorn University. In 1999, she got a scholarship from King Mongkut's University of Technology Thonburi (KMUTT) to continue a study at KMUTT and received her M.A. in Applied Linguistics (English for Science and Technology) in 2001. Since then she has worked as an English instructor at KMUTT. In 2004, she got a financial support from the Commission on Higher Education, Ministry of Education, under Cooperative Research Network (CRN) Program to further her doctoral study at Chulalongkorn University. Her research interests include language assessment and evaluation, test taking strategies and ESP testing, which are the main area of her dissertation.