การศึกษาองค์ประกอบทางความสามารถด้านการพูดจากแบบสอบการพูดด้วยคอมพิวเตอร์สำหรับนักศึกษาไทยระดับอุดมศึกษา

นางสาวสุทธิรักษ์ ทรัพย์สิรินทร์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต

สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2549

A STUDY OF TRAIT FACTORS OF ORAL LANGUAGE ABILITY IN A COMPUTER-BASED SPEAKING TEST FOR THAI UNIVERSITY STUDENTS

Ms. Sutthirak Sapsirin

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

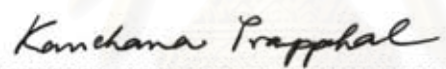Chulalongkorn University

Academic Year 2006

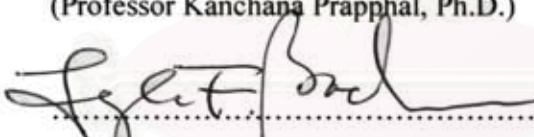|	
Thesis title	A study of trait factors of oral language ability in a computer-based speaking test for Thai university students
By	Miss Sutthirak Sapsirin
Field of Study	English as an International Language
Thesis Advisor	Professor Kanchana Prapphal, Ph.D.
Thesis Co-advisor	Professor Lyle F. Bachman, 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.)
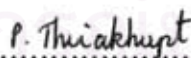
THESIS COMMITTEE
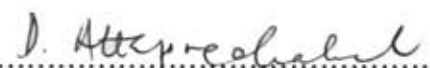
.... Chairperson

(Associate Professor Suphat Sukamolson, Ph.D.)

.... Thesis Advisor

(Professor Kanchana Prapphal, Ph.D.)

.... Thesis Co-advisor

(Professor Lyle F. Bachman, Ph.D.)

.... Member

(Assistant Professor Chansongklod Gajaseni, Ph.D.)

.... Member

(Assistant Professor Pavinee Thirakhupt, Ph.D.)

.... Member

(Assistant Professor Damrong Attaprechakul, Ph.D.)

สุทธิรักษ์ ทรัพย์สิรินทร์: การศึกษาองค์ประกอบทางความสามารถด้านการพูดจากแบบสอบการพูด ด้วยคอมพิวเตอร์สำหรับนักศึกษาไทยระดับอุดมศึกษา (A STUDY OF TRAIT FACTORS OF ORAL LANGUAGE ABILITY IN A COMPUTER-BASED SPEAKING TEST FOR THAI UNIVERSITY STUDENTS) อ.ที่ปรึกษา: ศ. ดร. กาญจนา ปราบพาล, อ.ที่ปรึกษาร่วม: Professor Lyle F. Bachman, Ph.D.

งานวิจัยนี้มีจุดมุ่งหมายเพื่อศึกษาโครงสร้างทางความสามารถของแบบทดสอบการพูดภาษาอังกฤษ ด้วยคอมพิวเตอร์ กระบวนการและกลยุทธ์ในการทำข้อสอบ และผลของข้อสอบต่อคะแนน คำตอบ และเจตคติของผู้เข้าสอบ กลุ่มตัวอย่างเป็นนิสิตมหาวิทยาลัยชั้นปีที่ 1 เครื่องมือที่ใช้ได้แก่ แบบสอบการพูด ด้วยคอมพิวเตอร์ มาตราส่วนประมาณค่า และแบบสอบถาม ความสามารถที่ศึกษา ได้แก่ ความรู้ด้านการ ออกเสียง ด้านวากยสัมพันธ์ คำศัพท์ การเชื่อมโยงข้อความ และหน้าที่ของภาษา แบบสอบกำหนดให้เล่า เรื่อง แสดงความคิดเห็น บรรยายเหตุการณ์ในจินตนาการ และโน้มน้าวใจผู้ฟัง การวิเคราะห์ข้อมูลใช้วิธี วิเคราะห์เชิงปริมาณ ได้แก่ ทฤษฎีการอ้างอิงสรุปของผลการวัด การวิเคราะห์องค์ประกอบเชิงยืนยัน การ วิเคราะห์ความแปรปรวนแบบพหุ และวิธีวิเคราะห์เชิงคุณลักษณะ ได้แก่ การวิเคราะห์เนื้อหาของการราย งานปากเปล่า เจตคติต่อแบบทดสอบ และการวิเคราะห์คำตอบของแบบสอบ

ผลการศึกษาพบหลักฐานที่สนับสนุนและโต้แย้งความตรงเชิงทฤษฎีของแบบสอบ หลักฐานที่ สนับสนุนได้แก่ ค่า factor loading ของความสามารถมีค่าปานกลางถึงสูงอย่างมีนัยสำคัญทางสถิติ (p < .05) ค่าความเที่ยงของความสามารถมีค่าสูงกว่า 0.70 ซึ่งบ่งบอกว่าแบบสอบสามารถวัดความรู้ที่ออกแบบให้วัด ได้ แบบสอบนี้จึงมีความตรงในการวัดความสามารถด้านการพูดที่ศึกษา นอกจากนี้ ผู้เข้าสอบใช้กระบวน การและกลยุทธ์ที่เกี่ยวข้องกับความสามารถที่ต้องการวัด ซึ่งบ่งบอกว่าคะแนนสอบอาจเป็นผลมาจากการ ใช้ความสามารถทางภาษาซึ่งต้องการวัดโดยแบบสอบ คำตอบของผู้สอบเป็นประเภท หน้าที่ของภาษา และ ลักษณะทางไวยากรณ์ ที่สอดคล้องกับสิ่งที่แบบสอบต้องการวัด และผู้สอบมีเจตคติเชิงบวกต่อแบบสอบ อย่างไรก็ตาม พบหลักฐานที่แย้งความตรงเชิงทฤษฎีของแบบสอบด้วย ซึ่งเกิดจากค่าความเที่ยงที่ต่ำของ ความสามารถด้านหน้าที่ ซึ่งอาจเกิดจากข้อสอบที่ให้ผู้สอบบรรยายโลกในจินตนาการ พบว่าข้อสอบข้อนี้ ยากที่สุดและกำกวม ทำให้คะแนนข้อนี้ต่ำที่สุดในเกือบทุกด้าน ผู้สอบแสดงหน้าที่ของภาษาที่ไม่เกี่ยวข้อง กับข้อสอบ และมีเจตคติทางลบต่อข้อสอบข้อนี้ นอกจากนี้ การไม่มีผู้สนทนาด้วยทำให้ผู้เข้าสอบมองว่า แบบสอบไม่เหมือนสภาพจริง โดยสรุปแล้วหากแบบสอบนี้ได้รับการปรับปรุง ก็อาจมีประโยชน์ในการ ทดสอบความสามารถด้านการพูดภาษาอังกฤษได้

สาขาวิชา ภาษาอังกฤษเป็นภาษานานาชาติ   ลายมือชื่อนิสิต...

(สหสาขาวิชา)   ลายมือชื่ออาจารย์ที่ปรึกษา..........................

ปีการศึกษา 2549   ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....................

4589687520: MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
KEY WORDS: TRAIT FACTORS / ORAL LANGUAGE ABILITY / COMPUTER-BASED SPEAKING TEST
SUTTHIRAK SAPSIRIN: A STUDY OF TRAIT FACTORS OF ORAL LANGUAGE ABILITY IN A COMPUTER-BASED SPEAKING TEST FOR THAI UNIVERSITY STUDENTS. THESIS ADVISOR: PROF. KANCHANA PRAPPHAL, PH.D., THESIS COADVISOR: PROF. LYLE F. BACHMAN, PH.D.

The purposes of the study were to investigate the trait structure of oral language ability in a computer-based speaking test (CBST), the test-taking processes and strategies of the examinees taking the test, the effects of test tasks on test scores and speech produced, and the attitudes of the examinees to the test. The participants were Thai first-year university students. The research instruments included the CBST, rating scales, and questionnaire. The traits being investigated were knowledge of pronunciation, syntax, vocabulary, cohesion and function. The test tasks comprised the narrative, opinion, imaginary and persuasive tasks which generated planned monologic responses. The data were analyzed through both quantitative and qualitative approaches. The quantitative approaches were multivariate generalizability theory, confirmatory factor analysis, and MANOVAs. The qualitative approaches consisted of content analysis of verbal protocols, and attitudes to the test, and discourse analysis of the test responses: analyses of genre, speech functions and grammatical features. The limitations of the study concerned the homogeneous nature of the participants in terms of their native language and educational level as well as the relatively small sample size.

The findings seemed to provide evidence which supported and challenged the validity of the CBST score interpretations. The supportive evidence may be, first, most of the trait factor loadings were significant ($p < .05$) and ranged from moderate to high. Also, the reliability of most constructs was above 0.70. This suggests that most measures were good indicators of the construct they were designed to measure, indicating the CBST was a valid measure of most of the speaking ability as defined. Second, the examinees engaged in test-taking processes and strategies relevant to the constructs of interest. This implies the test scores may be a product of using the language ability aimed to be measured by the test. Moreover, the genre, speech functions and grammatical features found in the test responses corresponded with the task requirements. Finally, the examinees generally had positive attitudes to the test. However, evidence that may threaten the validity of the score interpretation came from the low reliability of the functional knowledge construct. This may be the result of the imaginary task as the difficulty and ambiguity of the task prompt has led to the lowest mean score in almost all areas, unintended speech functions and negative attitude to the test. Also, the lack of interlocutors caused the examinees to view the test as unauthentic. Despite these weak points, with some revisions and careful interpretations made from the test scores, the CBST can serve as a potentially useful instrument in oral language assessment.

Field of study English as an International Language   Student's signature...................

(Interdisciplinary Program)   Advisor's signature....................

Academic Year 2006   Co-advisor's signature................

ACKNOWLEDGEMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background of the study

Current technological advances have made it possible to deliver a speaking test via computers with various sources of input such as text, visual and audio (Flewelling & Snider 2001; Kenyon & Malabonga, 2001). The characteristics of these computer-based speaking tests (CBSTs) differ to a great extent from other types of oral proficiency tests such as language interview or group discussion. The major differences lie in the absence of interlocutors whom the candidate interacts with. The candidate usually responds to preset questions presented with visual and/or audio input and record their answers which will be rated afterwards.

However, the implementation of CBSTs has raised concerns about the validity of such tests (Chapelle, 2001; Norris, 2001). This is because little is known about the extent to which the test methods of such tests may alter the candidate's performance, which in turn may affect the assessment of their oral language ability. Test method facets have long been acknowledged as one of the factors that influence language test performance (Bachman, 1990; Chalhoub-Deville, 1996; Skehan, 1998). Several research studies have reported the effects of test methods on test scores, language produced by the examinees, and attitudes towards several kinds of speaking tests (Brown, 1993; Hill, 1998; Hoekje & Linnell, 1994; Kenyon & Malabonga, 2001; Norris, 2001; O' Loughlin, 1995; Shohamy, 1982, 1994; Stansfield, Kenyon, Paiva, Doyle, Ulsh, & Cowles, 1990; Stansfield & Kenyon, 1992; Zeidner & Bensoussan, 1988). Unfortunately, few studies have investigated how the characteristics of CBSTs may have an impact on traits, test performance and attitudes of test takers taking the test. Thus, there is a need for a deeper investigation into the effects of the test methods so that we would better understand the constructs of speaking. This in turn would help determine the degree to which we can justify the interpretations that we can make of test scores from CBSTs.

1.2 Research questions

The present study addresses the following research questions:

1. What is the trait structure of oral language ability in a Computer-based Speaking Test (CBST) for Thai first-year university students?
2. What processes do the examinees employ in taking the CBST?
3. To what extent do different types of CBST tasks affect the test scores and characteristics of speech produced?
4. To what extent do examinees with different levels of speaking ability differ in terms of their attitudes towards the CBST?

1.3 Objectives of the study

The objectives of the present research are:

1. to examine the trait factors of oral language ability in a CBST.
2. to investigate the test taking processes of examinees taking the test.
3. to study the effect of different types of test tasks on test performance in terms of test scores and speech produced.
4. to examine test takers' attitudes towards the test.

1.4 Statements of hypotheses

The hypotheses in the study are as follows:

1. The trait factors of oral language ability for Thai first-year university students are multidimensional consisting of the factors of knowledge of pronunciation, syntax, vocabulary, cohesion and function.
2. Different types of test tasks in the CBST have a significant effect on the test scores.
3. There are no significant differences in terms of attitudes towards the CBST among examinees with different levels of speaking ability.

1.5 Scope of the study

The scope of the study is as follows:

1. The study used descriptive, correlational and "ex post facto" research designs (Isaac & Michael, 1995). Specifically, it used a multitrait-multimethod (MTMM) design (Campbell & Fiske, 1959).

2. The population was non-native English speaking university students in a country where English is taught and used as a foreign language. The sample was Thai first-year university students.

3. The data were collected by the following instruments and data collection methods: a computer-based speaking test, a computer-based speaking test questionnaire, and a retrospective verbal protocol technique.

4. The CBST is an English oral proficiency test which consists of such functions as narrating, expressing opinions, talking about an imaginary situation and persuading. The kinds of genre, speech acts, and monologic test responses represented in the test may not be generalized to other kinds of oral interaction.

5. In order to make the study feasible, the trait factors to be investigated include only some of those proposed in Bachman and Palmer's model (1996), namely, knowledge of pronunciation, knowledge of syntax, knowledge of vocabulary, knowledge of cohesion and knowledge of functions.

6. The data analysis methods includes quantitative methods: MTMM approach, confirmatory factor analytical technique (Jöreskog, 1993), Multivariate Analysis of Variance (MANOVA), and generalizability theory as well as qualitative methods: discourse analysis, content analysis and verbal protocol analysis.

1.6 Assumptions of the study

The assumption of the study is:

1. The participants are willing to do their best in taking the tests, producing verbal protocol and answering the questionnaire.

1.7 Limitations of the study

The limitations of the present study are as follows:

1. The participants were 167 Thai first-year university students. They were selected by a purposive method. Therefore, the results of the study may not generalize beyond this group.

2. The number of the participants with complete data for the CFA was relatively small. There should be about 10-20 cases per variable, that is, 200 cases for the present study (Schumacker & Lomax, 1996). However, some participants dropped out of the study and some data were missing. Thus, there were only 158 complete cases for the CFA.

3. Due to the limitations of the computer labs, the CBST could not be delivered to all participants in one session. However, the test administration followed the same procedure across sessions.

1.8 Definition of terms

Trait, which is a synonym for 'knowledge' and 'ability', refers to "entities which we may hypothesize to be in the minds of language users" (Bachman, 1990, p. 108). It is a latent variable which may be measured indirectly from a test or measurement.

Knowledge of pronunciation is knowledge of English pronunciation rules, for example, individual sounds, stress, rhythm, and intonation. In the present study, knowledge of pronunciation is measured by examinees' ability to produce comprehensible English utterances through appropriate use of sounds, stress, rhythm and intonation. It is represented by the average scores from the two raters applying the pronunciation rating scales.

Knowledge of syntax is knowledge of English syntactic structures. In the present study, knowledge of syntax is measured by examinees' ability to use syntactic structures appropriate to the task requirement. It is represented by the average scores from the two raters applying the syntax rating scales.

Knowledge of vocabulary is knowledge of general purpose vocabulary items, including cultural references. In the present study, knowledge of vocabulary is assessed by examinees' ability to select and use words appropriate to the task requirement. It is represented by the average scores from the two raters applying the vocabulary rating scales.

Knowledge of cohesion is knowledge of features for marking cohesive textual relationships to make relationships between words, phrases and clauses clear. In the present study, knowledge of cohesion is measured by examinees' ability to select appropriate cohesive markers to link the utterances. It is represented by the average scores from the two raters applying the cohesion rating scales.

Knowledge of functions is knowledge of how to create and interpret language functions of utterances. Knowledge of functions comprises that of ideational, manipulative, heuristic, and imaginary functions (Bachman & Palmer, 1996). In the present study, knowledge of function is measured by examinees' ability to use speech functions appropriate to the task requirement. It is represented by the average scores from the two raters applying the function rating scales.

Test tasks in the computer-based speaking test are activities which involve test takers in using language to achieve a particular goal in a particular situation. The test tasks in the present study include a narrative, opinion, imaginary and persuasive task. The narrative test prompt aims to elicit an enjoyable experience the examinees had on campus. The opinion prompt requires the examinees to express their opinion about studying for a Bachelor's degree abroad. The imaginary task intends to bring out a description of an imaginary ideal world. The purpose of the persuasive task is for the examinees to persuade foreigners to visit Thailand.

Test-taking processes and strategies Test-taking processes refer to general ongoing stages of preparing for responses and performing the test. Test-taking strategies refer to communication strategies an examinee uses when facing with problems preparing for responses and performing the test. In this study, test-taking processes are the components of strategic competence as proposed by Bachman and Palmer (1996), for example, goal setting and planning. On the other hand, test-taking strategies are communication strategies as defined by Færch and Kasper (1983). They operate when the speaker has communication problems. Examples of communication strategies are paraphrase and avoidance.

Attitudes refer to examinees' opinions towards the CBST regarding its usefulness, interface design, difficulty and time provided for preparing for and responding to the test tasks. In this study, the attitudes were measured by the CBST questionnaire which consisted of Likert-scaled items and open-ended questions.

1.9 Significance of the study

The findings of this study are expected to be significant in several ways. First, in terms of theoretical significance, the findings can contribute to a better understanding of trait factors of oral language ability in a CBST. They also can yield data on the validity of such test score interpretations. The information on what constitutes speaking abilities can assist language researchers and educators in designing and developing computer-based speaking tests. Finally, the developed CBST has a potential to be a useful instrument for several purposes such as a placement or proficiency test in assessing speaking abilities of Chulalongkorn University students since currently there is a lack of such speaking test for these students. The use of computers in test administration can be practical in giving the test to thousands of students.

1.10 An overview of the study

This chapter presents the background of the study and the research questions and objectives to respond to the current problem. It also describes the scope, assumptions, definitions of terms, and significance of the study.

Chapter 2 reviews the literature in the following areas: the nature of oral language proficiency of second/foreign language speakers, computer-based language assessment, the assessment of oral proficiency, and research approaches and analyses used in the study.

Chapter 3 focuses on research methodology. It presents the research design, population and sample, research instruments, data collection and analyses.

Chapter 4 presents the findings of the four research questions.

Chapter 5 summarizes the study, discusses the findings, describes the theoretical and practical implications and offers recommendations for future research and test users.

CHAPTER II
LITERATURE REVIEW

The review of relevant literature will cover the following topics. First of all, it includes the nature of oral language proficiency of second/foreign language speakers. Next, it will discuss studies on computer-based language assessment, the assessment of oral proficiency, and test taker reactions and attitudes to oral tests. It will also present the review of related research approaches in language assessment, that is, generalizability theory, confirmatory factory analysis, verbal protocol analysis and discourse analysis for oral language assessment.

2.1 The nature of oral language proficiency of second/foreign language speakers

The question of what speaking ability in a second/foreign language (SL/FL) is has been investigated from different perspectives (Bygate, 1987; Brown, 2001; Bachman & Palmer, 1996; He & Young, 1998). On one hand, spoken skill is viewed as consisting of sub-skills. Bygate (1987) described such sub-skills as production, accuracy and management skills. Brown (2001) suggests a variety of micro skills of oral communication:

a. "Produce chunks of language of different lengths.

b. Orally produce differences among the English phonemes and allophonic variants.

c. Produce English stress patterns, words in stressed and unstressed positions, rhythmic structure, and intonational contours.

d. Produce reduced forms of words and phrases.

e. Use an adequate number of lexical units (words) in order to accomplish pragmatic purposes.

f. Produce fluent speech at different rates of delivery.

g. Monitor your own oral production and use various strategic devices—pauses, fillers, self corrections, backtracking—to enhance the clarity of the message.

h. Use grammatical word classes (nouns, verbs, etc.), systems (e.g., tense, agreement, pluralization), word order, patterns, rules, and elliptical forms.

i. Produce speech in natural constituents—in appropriate phrases, pause groups, breath groups, and sentences.

j.   Express a particular meaning in different grammatical forms.

k.   Use cohesive devices in spoken discourse.

l.   Accomplish appropriately communicative functions according to situations, participants, and goals.

m.   Use appropriate registers, implicature, pragmatic conventions, and other sociolinguistic features in face-to-face conversations.

n.   Convey links and connections between events and communicate such relations as main idea, supporting idea, new information, given information, generalization, and exemplification.

o.   Use facial features, kinesics, body language, and other nonverbal cues along with verbal language to convey meanings.

p.   Develop and use a battery of speaking strategies, such as emphasizing key words, rephrasing, providing a context for interpreting the meaning of words, appealing for help, and accurately assessing how well your interlocutor is understanding you." (p. 272)

On the other hand, speaking ability can be defined by using a model of language ability. One of the most influential models is the theoretical framework proposed by Bachman (1990) and Bachman and Palmer (1996). In these models, language ability consists of two components: language knowledge and strategic competence. Language knowledge is defined as "a domain of information in memory that is available for use by the metacognitive strategies in creating and interpreting discourse in language use". Language competence is made up of various kinds of knowledge that are used in communication via language. This competence can be categorized into two types: organizational competence and pragmatic competence. Each of these comprises of several competencies. In communicative language use, all of these competencies interact with each other and with the language use context.

Organizational competence relates to two components: grammatical competence and textual competence. Grammatical competence is the knowledge of vocabulary, morphology, syntax, and phonology/graphology. It enables language learners to choose words to express specific significations, their forms, their order in utterances, and their physical realization as sounds or as written symbols. Textual competence refers to the knowledge of how to construct spoken and written discourse. This includes cohesion and rhetorical organization. Cohesion is conventions of marking semantic relationships such as reference, conjunction, and lexical cohesion,

and conventions that are involved in the ordering of old and new information in discourse. Rhetorical organization is related to text structure and the effect of the text on the language user. Conventions of rhetorical organization include common methods of development a written text such as narration, description, and comparison as well as those used in conversational discourse such as attention greeting, topic development, and conversation maintenance.

Pragmatic competence, the second component of language competence, is involved in functional competence and sociolinguistic competence. Functional competence is the knowledge of using language to express various functions and of interpreting the illocutionary force of utterances or the speaker or writer's intentions. This includes the ability to express or exchange ideas and emotions (ideational function); to get things done (manipulative function); to use language to extend our knowledge in such acts as teaching, learning, problem solving, and conscious memorizing (heuristic function); and to be creative for enjoyment (imaginative functions). The second component of pragmatic competence, sociolinguistic competence, enables us to perform language functions as illustrated above appropriately to the context in which they are used. It consists of sensitivity to differences in dialect or variety (the ability to use language that is appropriate to a certain regional and social group); sensitivity to differences in register (the ability to perform language functions in ways that are appropriate to register); sensitivity to naturalness (the ability to formulate and interpret an utterance in a nativelike way); and ability to interpret cultural references and figures of speech.

Strategic competence is "a set of metacognitive components, or strategies, which can be thought of as higher order executive processes that provide a cognitive management function in language use, as well as in other cognitive activities" (p. 70). In carrying out a language as well as non-language activity, these metacognitive strategies act as a mediator among topical knowledge, language knowledge, personal characteristics and affect as well as between these components and the features of language use and setting. They enable us to employ the following strategies: goal setting, assessment and planning.

First, goal setting concerns making a decision about what one is going to do by identifying the tasks, choosing a task if choices are provided and deciding whether to

carry out the task(s) or not. Next, assessment involves assessing the characteristics of language use setting, determining our topical knowledge, affective schemata, language knowledge which will be required in the task, and evaluating our response to the task in terms of correctness and appropriateness. Finally, planning enables us to select the components of language knowledge, topical knowledge, and affective schemata to be used in a plan, formulating and selecting a plan to achieve the task successfully.

Another theoretical framework of language competence, the interactional competence, is proposed by He and Young (1998). This framework is different from the communicative competence in a number of ways. First of all, interactional competence not only consists of discourse, pragmatic, and strategic competence but also includes competence in (at least) five interactional features. These features are a knowledge of rhetorical scripts, a knowledge of certain lexis and syntactic structures, a knowledge of managing turns, a knowledge of topical organization and a knowledge of signaling boundaries of an interactive practice.

Interactional competence also differs from communicative competence in that it "is not an attribute of an individual participant" (p. 7). Rather, it is

"co-constructed by all participants in an interactive practice and is specific to that practice. Participants' knowledge and interactive skills are *local*: they apply to a given interactive practice and either do not apply or apply in a different configuration to different practices". (p. 7)

According to this view, individuals acquire interactional competence that is dependent of a specific practice. Thus, interactional competence is jointly constructed by all participants who bring in their knowledge or resources they have acquired to a given practice.

The following sections survey empirical research studies that explore components of oral proficiency using a variety of research approaches.

2.1.1 Studies of organizational competence

Research on organizational competence reveals such problems SL/FL learners have as vocabulary, syntax, prosodic and discourse features which lead to

miscommunication or comprehension difficulties. For example, Tyler (1992) used a qualitative discourse-analytic framework to compare the spoken discourse of an international teaching assistant (ITA) and native speaker of English. It was found that the ITA did not use or he misused a number of discourse structuring devices in his presentations, which led to unsuccessful communication. The discourse structures that were examined were lexical discourse markers, lexical specificity, and syntactic incorporation. For instance, the additive markers were unclear. There was no explicit link between the words or phrases used. Finally, the pattern of interclausal connection was weak throughout the discourse. This misuse of discourse structuring devices had a negative effect on the listening comprehension of native listeners.

In a study of the effect of foreign accent and speaking rate of native and nonnative speakers on native speaker comprehension, Anderson-Hsieh and Koehler (1988) found that passages read by native speakers of English generated significantly higher comprehension scores than those read by nonnative speakers. Also, the passages read by all speakers at the regular rate resulted in significantly higher scores than those read at the fast rate. Speaking rate is more critical for the comprehension of the most heavily accented speaker than other speakers. Finally, prosodic features may have greater effect on comprehension than segmental features. Thus, it was concluded that speaking rate may be one of the factors that can affect comprehension of the listener.

He (1998) examined answer-related interactive features which contributed to failure of a candidate in a language proficiency interview of ITAs. Discourse competence as shown by the data in the study includes understanding of questions, repairs, and turn-constructional units; ability to elaborate responses; ability to strategically position pauses in between speech; and ability to specify the part one has problems with during communication. The first problematic aspect of the candidate's discourse was his use of "yeah". In English conversation "yeah" is used to acknowledge understanding and show cooperation. However, the candidate's use of the word often suggested that he did not understand the examiner. Another problem in the candidate's answers was in his turn which immediately followed the examiner's initial question. The candidate often gave one-word responses and left long pauses after the examiner's question. When he did not understand the question, his responses

did not indicate where the problem was. Also, he sometimes gave wrong or contradictory answers. All of these characteristics of his turn led the examiners to ask a follow-up question or repeat or recast the question throughout the interview. The researcher concluded that these problematic discourse features affected how the candidate's oral proficiency was evaluated, which led to his failing the interview. The data also suggested that grammatical competence and discourse competence are inseparable. The short responses, for example, may be due to the candidate's limited linguistic knowledge. Pauses after being asked a question and the misuse of "yeah" may result from the candidate not understanding grammatical and interactional structures.

Young and Halleck (1998) compared the degree of topical organization among Japanese and Mexican learners of English in language proficiency interviews. It was found that two factors influence topical organization in the interviews: transfer of first language conversational style and proficiency level. The effects of these factors seem indistinguishable. The Japanese appeared to elaborate their answers less than the Mexicans and lower proficiency speakers did the same when compared to higher proficiency speakers. The transfer of Japanese conversational style was judged negatively by American interviewers who prefer topical elaboration. However, the researchers mentioned stylistic differences between individuals from the same culture and personality differences may also explain topical organization in the interviews.

Williams (1992) examined the planned and unplanned production of 24 nonnative speaking teaching assistants. The data indicated that the degree of discourse marking differ significantly between the two conditions while the degree of grammatical accuracy did not. In the planned production, the ITAs used more explicit discourse marking, which contributed to more comprehensibility of the listeners. Thus, the researcher suggested that ITAs should be taught how to use discourse markers. This method viewed as a part of strategic competence can help overcome their comprehensibility difficulties.

2.1.2 Studies of pragmatic competence

Pragmatic competence is another type of ability necessary for second/foreign language learners to communicate effectively. Ross (1998) has demonstrated that insufficient pragmatic competence may affect how candidates responded to interviewer's questions in a context of interview interaction. In this study, Japanese examinees seemed to transfer the pragmatics rules in their first language into second language performance. They may not realize that the norms for talk about certain topics in the interview are different from Japanese norms. Thus, they were reluctant to elaborate on their answers, which may lead the interviewer to view them as uncooperative or to misjudge their language proficiency. Another aspect of pragmatic competence shown in this study is the ability to interpret and create meaning in a particular context. The candidates who did not understand the frame of the interview question may mistakenly interpret the questions as real questions, not as ways to elicit speech to be evaluated. As a result, the examinees may not be willing to answer some questions. The researcher concluded that the underelaboration of answers is frequent even in speech of high-level second language learners. This would suggest that pragmatic competence may not be acquired at the same level as any other second language skills.

In a study of sociolinguistic competence of SL speakers of English, Ranney (1992) examined cultural knowledge of the choice and sequencing of language functions in medical consultation of native speakers of Hmong studying at a university in the United States. The results showed that the learners acquired considerable knowledge of the sociolinguistic and discourse norms for the medical consultation in the US although some aspects of this knowledge were different from those of Americans due to cultural differences. For example, the Hmong used on-record strategies for requesting medication or refusing surgery, which Americans found to be inappropriate and showing disrespect for the doctor. To conclude, the researcher marked that although these Hmong were highly educated and fluent in English, they may have some difficulties in medical consultation due to sociolinguistic knowledge different from that of native speakers.

Kim and Suh (1998) analyzed confirmation sequences in a Korean Language Proficiency Interview. This feature of the Korean language reflects a social hierarchy

between the interviewer and interviewee. When an examinee ignored the interviewer's confirmation request, the interviewer would view this as challenging the hierarchy, and that led to a negative evaluation of his/her proficiency. The researcher noted that Korean language learners should attend to this aspect of the language since it is a part of sociolinguistic competence in Korean.

2.1.3 Studies of trait structure of oral language ability

In addition to identifying components of speaking ability, language assessment research has investigated the trait structure of SL/FL ability that underlies language performance in test and non-test situations. Since oral language ability is subsumed under SL/FL ability, the research in this area can also shed light on the trait structure of oral language ability.

A number of studies seem to show that SL/FL ability comprises a general language ability factor and several specific factors (for example, Bachman & Palmer, 1981, 1982; Carroll, 1983; Fouly, Bachman & Cziko, 1990; Sawaki, 1993). The intercorrelations between the specific factors may be accounted by the second-order general factor. It should be noted that in the case where there are two or three first-order factors, a model of the correlated traits and that of a second-order trait accounting for the uncorrelated first-order traits cannot be discriminated.

For instance, Bachman and Palmer (1981) conducted a study to investigate the construct validity of the Foreign Service Institute (FSI) oral interview. The study used six measures involving two traits, speaking and reading ability, each of which was measured by three methods, interview, translation and self-ratings. The results showed the most parsimonious and fit model which could explain the data contained two distinct but correlated traits, speaking and reading ability and three method factors. The findings supported a model of partly divisible language ability.

In a study of construct validation of communicative performance, a model with similar structure was found (Bachman & Palmer, 1982). It consisted of a higher order general factor and two uncorrelated trait factors as well as method factors. The first trait can be interpreted as the grammatical and pragmatic competence, while the

second trait the sociolinguistic competence. The higher order general factor may involve information processing for handling extended spoken or written discourse.

The multidimensionality of SL/FL proficiency was supported by subsequent studies. Fouly et al (1990) discovered evidence of a general higher-order trait factor accountable for the correlations among the first-order trait factors. The higher order factor was hypothesized a general language proficiency, and the first-order traits an oral-aural ability to understand and speak the second language; a structure-reading comprehension ability; and an ability to interpret cohesion and organization of extended utterances.

Sawaki (1993) found similar results: a general SL proficiency factor and three uncorrelated specific first-order factors. These first-order factors were associated with composition ability, an ability to understand short context, and an ability to understand long context.

2.1.4 Studies of communication strategies

Communication strategies (CSs) can be defined from different viewpoints. Some researchers see them as intra-individual (Tarone, 1977; Færch & Kasper, 1983; Poulisse et al., 1984), while others view them as inter-individual (Tarone, 1983). The former restrict the notion of CSs to problem solving activity. For instance, CSs are (cited in Kasper & Kellerman, 1997)

… used by an individual to overcome the crisis which occurs when language structures are inadequate to convey the individual's thought. *(Tarone,1977. p. 195)*

… potentially conscious plans for solving what to an individual presents itself as a problem in reaching a particular communicative goal *(Færch & Kasper, 1983, p. 36)*

… strategies which a language user employs in order to achieve his intended meaning on becoming aware of problems arising during the

planning phase of an utterance due to his own linguistic shortcomings." *(Poulisse et al.,1984)*

From these concepts, CSs are used when communication is problematic and they occur internally, without the interlocutor's help. On the other hand, the second view of CSs is concerned with the inter-individual characteristic of CSs. Thus, CSs requires the collaboration between the speaker and the listener when there is a gap in communication. As Tarone (1983) put it, CSs are

…a mutual attempt of two interlocutors to agree on a meaning in situations where requisite meaning structures do not seem to be shared... attempts to bridge the gap between the linguistic knowledge of the second-language learner, and the linguistic knowledge of the target language interlocutor in real communication situations" *(Tarone,1983, p. 65)*

According to Cohen and Dornyei (2002), CSs can be classified into four major groups: avoidance or reduction strategies, achievement or compensatory strategies, stalling or time-gaining strategies, and interactional strategies.

1. Avoidance or reduction strategies

These strategies consist of

- Message abandonment: the speaker does not finish a message due to some language difficulty
- Topic avoidance: the speaker avoids some topic areas which s/he has difficulty in talking about
- Message replacement: the speaker substitutes the original message with a new one when feeling incapable of delivering it

2. Achievement or compensatory strategies

- Circumlocution: the speaker describes words they can't remember
- Approximation: the speaker uses an alternative form which is as close in meaning as the target word
- Use of all-purpose word: the speaker uses a general 'empty' word when s/he can't think of the specific word

- Word-coinage: the speaker creates a new L2 word based on a supposed rule

- Use of non-linguistic means such as mime, gestures and facial expression

- Literal translation: the speaker translates literally from L1 to L2

- Foreignizing: the speaker uses an L1 word but with L2 phonology and/or morphology

- Code switching: the speaker includes L1 words in L1 pronunciation when producing L2 speech

   3. Stalling or time-gaining strategies

- Use of fillers or other hesitation devices: the speaker uses fillers to gain time to think

- Repetition: the speaker repeats a word or expression right after they were said

   4. Interactional strategies

- Appeal for help: the speaker asks for help from his/her conversation partner directly or indirectly

- Asking for repetition: the speaker asks the conversation partner to repeat the utterances s/he does not understand

- Asking for clarification: the speaker requests explanation of something s/he does not understand

- Asking for confirmation: the speaker requests confirmation that s/he understood something correctly

- Expressing non-understanding: the speaker expresses that s/he does not understand something

- Interpretive summary: the speaker paraphrases what the conversation partner said to check that s/he understood it correctly

These taxonomies of strategies have been criticized from the researchers of the Nijmegen Group (Bongaerts & Poulisse, 1989; Kellerman, 1991). They have argued that these categorizations deal with the linguistic realization but do not reflect the psychological processes underlying the strategies. The researchers, thus, have proposed a new, process-based taxonomy: conceptual strategies, which learners use to manipulate the intended concept, and code or linguistic strategies, which they use to manipulate encoding media. Conceptual strategies can be categorized into holistic and

analytic approaches. The holistic approach involves naming a referent which is similar to the target referent, for example, the use of 'bird' for 'sparrow' and 'chair' for 'stool'. The analytic approach is the selection and listing of properties of the intended referent, for instance, 'haircutter' used for 'hairdresser' and 'hot beast' for 'scorpion'. The second process-based taxonomy, linguistic strategies, entails resorting to another language such as the speaker's first language or to the knowledge of the rule systems of the target language. Examples of these strategies are morphological creativity (e.g. 'representator' used for 'representative'), literal translation, and borrowing. Another subtype of linguistic strategies is non-verbal means such as onomatopoeic devices and pictorial representation.

Research studies on CSs have revealed how SL/FL speakers of English use CSs in communication. For example, Paribakht (1985) and Yoshida-Morise (1998) compared CSs used by students at different levels of English proficiency. Paribakht (1985) found that there is a directional relationship between English proficiency level of Persian speakers and CSs. That is, advanced learners of English used more approximation strategies than lower-level learners. Native speakers (NSs) of English and advanced learners used linguistic approach (semantic contiguity, circumlocution, metalinguistic clues) more often than intermediate learners. On the other hand, intermediate learners employed more conceptual approach (demonstration, exemplification, metonymy) than NSs and advanced learners. Finally, intermediate and advanced learners used mime more often than NSs.

Yoshida-Morise (1998) analyzed the types of CSs 12 Japanese interviewees used in the Oral Proficiency Interview (OPI). The findings showed the greatest difference between speakers at different levels of English proficiency: levels 1+, 2, 2+ and 3. Low-proficiency participants (level 1+) used more strategies than other participants to compensate for their insufficient English knowledge. However, the groups of higher proficiency levels did not differ much in the mean frequency of CSs they used. Further, the data showed six out of 11 strategies had significant differences according to proficiency levels. First, the groups at the highest proficiency level used more repair strategies than the other groups. Next, as the level of proficiency increased, the use of telegraphic, paraphrase and interlingual strategies decreased. Finally, the intermediate level students used reduction strategies and fillers more than

other groups. The researcher also noted that in addition to proficiency levels, individual differences influenced several strategy categories.

Katona (1998) focused on how participants who were familiar and not familiar negotiated meaning when there was communication breakdown in an oral proficiency examination. The researcher looked at the CSs which resulted in negotiation and grouped the negotiation exchanges containing those CSs into the helping sequence, the eliciting sequence and the clarifying sequence. The helping sequence consists of an (optional) initiation, a response which includes one or more CSs, a follow-up move containing the discourse act of Help, an (optional) second response move, and an (also optional) second follow-up move. The eliciting sequence consists of an (optional) initiation, a response including one or more CSs, a follow-up move which contains some type of Elicit, and a second response (which may have some CSs). The last sequence, the clarifying sequence, consists of an (optional) initiation, a response which includes one or more CSs, a follow-up move with some type of Clarify, a second response, and an (optional) follow-up move. It was found that there seemed to be a relatively strong relationship between familiarity of the candidates with the interlocutor and the sequence types used by the interlocutors. That is, the unknown interviewers used fewer negotiation sequences than the known interviewers. When different types of meaning negotiation strategies were compared, the familiar examiners used mainly "help-based" types, while the unfamiliar examiners used a combination of the three.

Some research has investigated CSs used by native and nonnative speakers. Tarone and Yule (1989) studied the differences between NSs and non-native speakers (NNSs) of English in using CSs when they described objects pictured on a card to a listener. The results showed that NSs used circumlocution and approximation more often than NNSs. They also differed in the level of detail provided in transmitting information. In addition, Bongaerts and Poulisse (1989) compared the similarities and differences in L1 and L2 referential communication. In this study, native speakers of Dutch described a set of unconventional abstract shapes, first in Dutch and then in English. It was found that the participants in both versions dealt with the tasks in much the same way. They preferred a strategy in which they described the pictures from holistic perspectives, that is, the referent is conceived as a whole.

In the research by Shohamy (1994), CSs from a direct and semi-direct speaking test were compared. The data showed a significant difference for one strategy, paraphrasing. The examinees taking the semi-direct test paraphrased more often than those taking the direct one. There were two strategies that were nearly significant, self-correction and switch to L1. The former was used more on the semi-direct test while the latter on the direct test.

2.2 Computer-based language assessment

Regarding the use of computers in language assessment, it can be seen that recent advances in computers have been incorporated into many areas of language assessment: the development, administration, scoring, and evaluation of language tests (Alderson, 2000; Brown, 1997; Chalhoub-Deville & Deville, 1999; Chapelle, 2001). Computer-based tests (CBTs) for language assessment have largely focused on receptive skills; the assessment of speaking skills via computer has been limited, mainly because of technology constraints (Alderson, 2000; Chalhoub-Deville & Deville, 1999).

There are few CBT projects in the language assessment area. To begin with, a computer-adaptive reading test was discussed by Madsen (1991). The test aimed to measure English as a Second Language (ESL) learners' reading ability. The test results were used for student placement into an intensive ESL program. The test tasks consisted of multiple-choice reading items. The adaptive nature of the test did not allow examinees to return to previously completed items. The total time for completing the test and the number of the items varied from one test taker to the next so there was not set time limit.

Another example of a computer-based reading test is from a research project by Henning, Anbar, Helm, and D'Arcy (1993). The test in the project was used for the purpose of comparing different item formats and scoring techniques. The formats included multiple-choice and corresponding binary open-ended items (as correct/incorrect) to check examinees' reading passage comprehension. The items were scored by computer and then checked manually. The open-ended items were scored by degree of correctness manually first. After that, the scores were incorporated into the computerized scoring system.

The next example is a computer-based Test of English as a Foreign language (TOEFL). Taylor, Kirsch, Eignor and Jamieson (1999) have described the features of the CBT TOEFL which is different from the paper-and-pencil test in various ways. The CBT listening and reading comprehension items, for instance, display one question at a time on the screen. The examinees respond to the questions in the order they are presented before moving to the next questions. Further, the test makes use of visual, non-linguistic as well as aural input. Finally, new test formats are included in the test. In addition to multiple-choice questions, the CB listening test requires the examinees to select two answers and match or order information presented in the listening stimuli. The CB reading test contains two new item types: clicking on a word, phrase or sentence within a passage, and inserting a sentence in a passage.

Two projects of CB listening tests have been reported (Dunkel, 1991 and 1999). The first test is a prototype of a computer-adaptive listening test (Dunkel, 1991). This ESL prototype test was divided into three sections: understanding/making appropriate responses to questions; recognizing synonymous statements; and comprehending monologues and conversations. The test items were two, three, or four multiple-choice questions. The test contained some of the aspects of computer adaptive testing. That is, the examinee was branched up or down levels as a result of their correct or incorrect responses to items. In another test, Dunkel (1999 cited in Chapelle, 2001) described a computer adaptive listening proficiency test of Hausa. The test required the examinees to perform four listener functions: identification/recognition, orientation, main idea comprehension, and detailed comprehension.

A computer program has also been implemented as a proofing cloze test to measure English language proficiency (Coniam, 1997). The test presented a reading text which contained errors that examinees had to identify and correct. The examinees could identify the incorrect word by clicking on it. If they were correct, a window would appear with the word and multiple choices for them to choose. During this process, the original reading text would still appear so it can be read. To prevent examinees from clicking words randomly, a scoring system of rewards and penalties was adopted.

A CBT has also been implemented to measure writing ability of native speakers of English. In a pilot project by Powers, Fowles, Farnum, and Ramsey (1994

cited in Chapelle, 2001), a computer-assisted writing test gives the examinees 50 minutes to write an essay on a single topic. The essays were scored by human raters using a holistic scale.

Currently, there has been progress in speech recognition technology (e.g. PhonePass, Ordinate, 1998) and the development of software for evaluating oral responses. For instance, CARLA (Flewelling & Snider, 2001) is a software developed by the University of Windsor to evaluate students' performance orally. The software has graphic and video capability, allowing different types of prompts such as sound material, still picture and moving picture. The students' responses to questions are stored on either the computer hard drive or a school server. Thus, teachers can evaluate the responses from any computer, and they can record their feedback and comments. Another CBT test for oral assessment is Computerized Oral Proficiency Instrument-Center of Applied Linguistics (COPI) (Kenyon & Malabonga, 2001). COPI is a multi-media computer adaptive test which is an adaptation of the tape-mediated Simulated Oral Proficiency Interview (SOPI). The COPI allows test takers to choose the amount of preparation and response time, speaking function, topics, level of difficulty and language of the directions for each performance task. The scoring program makes it possible for raters to hear the responses for the tasks in any order. While raters assess each task, elements of the task appear on the screen. In addition, the raters can write notes to test takers to give them overall comments and task-specific feedback.

These technological advances have offered several advantages of computer-based assessment in terms of test delivery, test construction, test compilation, response capture, test scoring, result calculation and delivery, test analysis, storing of tests and details of candidates (Alderson, 1996). However, the computer delivery methods have raised a concern on the validity of CBTs (Alderson, 2000; Chapelle, 2001; Norris, 2001). Alderson (2000) has called for more research on the validity of the tests, test taking processes and strategies, constructs being measured, and the impact of the use of the technology on learning, learners and the curriculums.

Chapelle (2001) addressed an issue of the undesirable and unpredictable effects the computer may have on examinees' performance. The question that follows is whether a computer-based and a paper-and-pencil version of the 'same' test measure the same or different traits. Similarly, Norris (2001) pointed out several key

issues that need to be explored for COPI, for instance, whether the adaptive feature of the test will lead the examinees to select tasks which adequately reflect their actual abilities and whether there are sufficient numbers of tasks of varying difficulty levels and topics for the examinees to choose. Moreover, Norris (2001) stressed the importance of research exploring whether the computer delivery method can offer us evidence about learners' L2 speaking ability, rather than asking what computers are capable of doing.

2.3 The assessment of oral proficiency

2.3.1 Characteristics of oral proficiency tests

The assessment of oral proficiency is different from other kinds of language tests (Underhill, 1987). He defined a genuine oral test as a test in which "real people meet face to face, and talk to each other" (p. 3). Their performance is then assessed based on the speech. The characteristics of speech normally include a two-way communication in which the speaker and listener interact and switch their roles. This feature makes oral tests distinct from listening, reading and writing tests. Thus, people taking part in the test and their interaction are primary concern in oral language test design and development.

Oral proficiency tests come in different forms (Clark, 1985; Underhill, 1987; He & Young, 1998; Flewelling & Snider, 2001; Kenyon & Malabonga, 2001). Underhill (1987) described general types of oral tests: self-assessment, teacher assessment, direct interview type, pre-arranged information gap, prepared oral tests, mechanical/entirely predictable tests, recorded tests and telephone tests. To select the most appropriate type of oral tests depends on the testing purposes, available resources, and the needs and expectations of learners.

Clark (1985) differentiates two approaches to oral proficiency interview: a direct and a semi-direct technique. They differ in that the latter is '… any procedure which elicits active speech by the examinee in response to tape recorded stimuli, printed test booklets or other "non-human" means' (p. 133). The direct speaking test can also be described as what He and Young (1998) define a language proficiency interview (LPI). A LPI is a face-to-face spoken interaction usually between two participants: an interviewer, usually a native or near-native speaker of the language in

which the interview is conducted, and an interviewee, a non-native speaker (NNS) or second/foreign language learner. The purpose of the LPI is for the interviewer to assess the NNS's ability to speak the language. In the case of scripted interviews, the interviewer will use an agenda containing the topics for conversation and the activities to take place during the LPI. In addition to the agenda, the interviewer (but usually not the NNS) has access to one or more scales for rating the NNS's ability in the language of the interview.

A LPI that is widely used to measure second and foreign language speaking ability in the U.S. is the ACTFL Oral Proficiency Interview (OPI) (Johnson & Tyler, 1998). It has been developed by the American Council on the Teaching of Foreign Languages (ACTFL), which has conducted a number of projects which aimed to develop, refine and disseminate procedures for describing and measuring foreign/second language proficiency since 1981 (Dandonoli & Henning, 1990). The ACTFL Proficiency Guidelines and the oral proficiency interview (OPI) are one of the major contributions of the ACTFL. In the OPI, the test taker converses face-to-face for 10-30 minutes with one or two trained testers who try to elicit from him / her a rich sample of speech by using questions and role-play situations. The candidate's speech sample will be evaluated on a scale ranging from 0 (no functional ability) to 5 (proficiency equivalent to that of a well-educated native speaker).

Another type of oral language interviews, a semi-direct test, can be delivered through different kinds of technology. For instance, the SOPI is a tape-mediated speaking test designed to be used in situation where the OPI is not possible to be conducted (Stansfield & Kenyon, 1992). A satellite-based video teletraining (VTT) has been explored by the Defense Language Institute Foreign Language Center to measure oral language proficiency (Clark & Hooshmand, 1992). Finally, the CARLA and COPI as mentioned earlier use the computer delivery method to assess examinees' speaking ability (Flewelling & Snider, 2001; Kenyon & Malabonga, 2001).

2.3.2 Factors which influence speaking performance

There are several factors which can influence the candidate's performance in an oral test (Bachman, 1990; Bachman & Palmer, 1996; Chalhoub-Deville, 1996; Skehan, 1998). Bachman (1990) and Bachman and Palmer (1996) stated that two sets

of characteristics can affect language use and test performance: characteristics of language users or test takers and characteristics of tasks. The individual characteristics include language ability, personal characteristics, topical knowledge and affective schemata. The characteristics of tasks can be explained by using the framework consisting of features describing five aspects of tasks: setting, test rubric, input, expected response and relationship between input and response.

Chalhoub-Deville (1996) mentioned that test methods and raters have potential influence on the use and interpretation of oral test scores. In addition to rating procedures, a model of oral test performance proposed by Skehan (1998) presented other factors as well. These are task characteristics and conditions, the interactive conditions under which performance was elicited, and the ability for use of the candidate.

A number of empirical studies have reported on factors, in addition to test takers' language ability, that have an effect on test performance and scores derived, for example, examinees' behaviors, interlocutor effects, test tasks, and raters.

First of all, candidates' nonverbal behaviors in an oral proficiency interview may influence the way raters judge their language ability (Jenkins & Parra, 2003). They demonstrated that the nonnative speaker interviewees who used nonverbal behavior considered appropriate by North American raters and who negotiated and reduced the interview asymmetry were able to compensate for their limited linguistic proficiency. Active nonverbal behavior and appropriate paralinguistic features exhibited during their talk ensured high scores for linguistically proficient candidates and low passing scores for linguistically weaker ones. The researchers suggest that both test takers and evaluators should be aware of the powerful role these nonverbal behaviors play in a proficiency test.

The impact of test taker characteristics on performance in an oral language test has been investigated through cross-cultural communication studies. Ross (1998) demonstrated that Japanese examinees used their own cultural norms of talk in responding to interview questions, which led to underelaboration of responses. Young and Halleck's study (1998) also supported the findings of Ross (1998) in that Japanese interviewees transferred their topical organization in L1 into a language test. This feature contradicted the norms of American interviewers; thus, the Japanese may be misjudged in terms of their English proficiency. Davies (1998) found that the

characteristics of not only the interviewee but also the interviewer may affect how they interact in an oral exam. In this study, American examinees and a Korean bilingual examiner sometimes code switched into English during the exam. Davies explained that 'the complex cross cultural context of the face-to-face interaction creates competing role expectations for the participants' (p. 271).

Interlocutor behaviors can have an impact on the assessment of speaking ability. McNamara and Lumley (1997) studied the effect of perceived competence of interlocutors to carry out their role on rating in the Occupational English Test (OET). In the case in which interlocutors were viewed as relatively incompetent, raters tended to give higher scores for the candidates interacting with the interlocutor. The researchers recommend more research on variables that affect the candidates' scores to understand the nature of performance assessment and to ensure fairness to candidates.

Lazaraton (1996) identified eight types of linguistic and interactional support that examiners provided to nonnative speaker candidates in the Cambridge Assessment of Spoken English (CASE). This includes priming topics; supplying vocabulary or engaging in collaborative completions; giving evaluative responses; echoing and/or correcting responses; repeating questions with slowed speech, more pausing and overarticulation; stating question prompts as statements that merely require confirmation; drawing conclusions for candidates; and rephrasing questions. The researcher noted that these speech behaviors need to be further investigated since they are not consistent and may have impact on candidate language use as well as rating of the performance.

The relationship between the candidate and interviewer may also influence the way they interact in a speaking test. Katona (1998) found differences in negotiations of candidates with unknown and known interlocutors. It was pointed out that the interaction between familiar participants seemed to have the characteristics of natural NS-NNS oral interaction, while that with unfamiliar participants often led to misunderstanding and was more like a formal talk.

Moreover, task types can create differences in the interactional structure between candidates and interviewers. In an analysis of non-scripted interviews and guided role-play activities in oral proficiency exams, Kormos (1999) found that the role-plays have more features of real-life conversations than the interviews. In the

role-plays, the conversational interaction was more symmetrical. The candidates had the opportunity to introduce and ratify approximately the same number of topics as the interviewers; to interrupt and hold the floor more effectively; and to open and close a conversation. On the other hand, in the interviews in which there was unequal distribution of power, candidates rarely had the chance to initiate a new topic and had no right to open or close the conversations. Some of these characteristics of speech in interviews are supported by the study of Lazaraton (1992). She pointed out most of the sequences in the interview interaction were initiated by the interviewers and controlled by recommended agenda forms. However, some sequences such as the greetings and closing were initiated by the candidates. Further, in general the interview seemed to share some structural organization of conversation although the turn taking is a pre-specified system which is a distinctive characteristic of interviews, not conversations.

There has been an investigation of rater variables and speaking assessment. A study explored the effects of rater's occupational and linguistic background on rating of the Japanese Language Test for Tour Guides (Brown, 1995). The assessors in this study were native and near-native speakers of Japanese with background in teaching or tour guiding. The data showed no significant differences in rating between groups of raters. However, there were significant differences in rating for some individual criteria such as pronunciation and task fulfillment assessments. The researcher, then, raised an issue of who is more appropriate in devising assessment scheme in occupational proficiency.

Task performance conditions are another factor that has been examined. Wigglesworth (1997) in a study of planning time and oral test discourse demonstrated that planning time had some effect on discourse measures of complexity and fluency although it did not affect subjects at all proficiency levels.

In conclusion, it can be seen that several variables play a role in speaking test performance, interaction and scoring. Thus, it is important that these variables should be further studied so that we can understand the impact they have on oral language assessment.

2.3.3 Studies on semi-direct speaking tests

As mentioned earlier, semi-direct tests of speaking differ from other kinds of tests of spoken proficiency in terms of test methods. The examinees do not have an interlocutor to interact with. They respond to test questions and situations via technology such as audiotapes, teleconferencing and computers. Also, the rating of the test performance usually takes place after the test and information on the nonverbal behaviors of the candidates may not be available. The introduction of semi-direct speaking tests, then, has raised issues on their validity, authenticity and comparability with direct tests of speaking (Hoekje & Linnell, 1994; Norris, 2001; O' Loughlin, 1995; Shohamy, 1994; Stansfield & Kenyon, 1992).

Regarding the comparability of the SOPI and OPI, Stansfield and Kenyon (1992) reviewed research on the two test formats. In terms of reliability, there was very high inter-rater agreement on the SOPIs across languages and forms and on the OPI as well, with the exception in the test of the Chinese language. Using data from Clark (1987), the researchers calculated test-retest reliabilities and showed that the reliability of the SOPI compared very favorably with that of the OPI. Finally, in the comparability of scores, the SOPI showed a strong enough correlation with the OPI so it was concluded that the SOPI can serve as a variant of the OPI.

The results of the study by Clark and Hooshmand (1992) supported the previous studies. In this study, Clark and Hooshmand (1992) examined a speaking test which used satellite-based video teletraining technology and a face-to-face oral language test of Arabic and Russian. The quantitative analyses showed high agreement between the two test formats in the test scores.

Studies which compare the speech elicited by a direct and semi-direct test provide insights into the qualities of the tests. Shohamy (1994) examined the validity of the OPI and SOPI through both quantitative and qualitative methods. When the language samples from the two tests were compared, produced oral discourse varied from one test format to another. The SOPI tasks could elicit language that was mostly description, reporting and some narration. They could not elicit complaining and thanking as intended in the 'situation' part of the test. In contrast, the OPI tasks allowed the examinees to produce a variety of speech functions, for example, elaboration, expansion, negotiations for meaning and interviewing the examiner. In addition, when elicitation tasks were compared between the two tests, it was found

that the SOPI had an advantage in sampling for low-proficiency examinees, and was similar to the OPI for middle-proficiency levels. However, the OPI had an advantage for the high-level examinees. The researcher concluded that the test format can influence or even dictate the type of language examinees produce.

Koike (1998) described the output of examinees taking an OPI and SOPI Spanish test. The results were different from what Shohamy (1994) found in that the significant differences between the discourse from the two tests were not numerous. Also, what seemed to influence the language of the candidates was the task type and specific topic rather than the test formats themselves. However, the test modalities did affect the use of particular strategic and discourse factors. The OPI tended to produce interactive discourse, while the SOPI elicited more formal language and more fillers and self correction. The speech from the SOPI was also better organized, more focused and contained more ideas.

In a study comparing direct and semi-direct versions of **access:** test (the Australian Assessment of Communicative English Skills), O' Loughlin (1995) revealed the differences in lexical density of test takers' discourse. The findings showed that the effect of test format, task and their interaction were all statistically significant although the differences were not large overall. Further, the candidates' speech in the semi-direct format was more literate, lexically dense. However, he concluded that 'the degree of interactiveness, rather than test format, emerges as perhaps the single most important determinant of candidate output in the study' (p. 236).

Hoekje and Linnell (1994) evaluated three tests: the SPEAK test, the OPI and a performance test, in terms of their authenticity as language tests for international teaching assistants. The results indicated that the tests differed substantially with regard to the language produced. The SPEAK, a semi-direct audiotaped test of spoken language proficiency developed by Educational Testing Service (ETS), elicited relatively short responses; thus, several aspects of discourse such as the organization and chunking of material and transitions between topics, and complex referencing cannot be evaluated on the test. On the other hand, the OPI can elicit much more extended and complex language use but the interaction on the OPI was different from that in the classroom. The performance test allowed examinees to organize their language of the classroom and demonstrate grammatical, discourse, sociolinguistic,

and strategic competence. In conclusion, the performance test was better than the other two tests in terms of authenticity of task.

2.3.4 Test taker reactions and attitudes to oral tests

Several studies have reported examinee reactions to varying types of oral tests with mixed results (Brown, 1993; Hill, 1998; Kenyon & Malabonga, 2001; Shohamy, 1982; Stansfield et al., 1990; Zeidner & Bensoussan, 1988). College students in an advanced EFL reading comprehension course were found to prefer written over oral tests of English as a foreign language (Zeidner & Bensoussan, 1988). Written tests were viewed as more pleasant, valuable, fair, less anxiety evoking and more reflective of text comprehension while oral tests were perceived as more interesting to take. In contrast, Shohamy (1982) showed that students of Hebrew as a foreign language perceived oral interviews as low anxiety producing, pleasant and reflecting their knowledge of the language assessed.

Another research study has presented results regarding candidate reactions to semi-direct oral tests (Brown, 1993). In a study of test takers' reactions to a tape-mediated test of proficiency in spoken Japanese, Brown (1993), using a post-test questionnaire, reported that, in general, examinees reacted positively to the format of the test. There was strong support from the test takers who considered the test to be valid although the majority of them found the test difficult because of inadequate response time.

In addition, some studies have compared a direct and a semi-direct oral language test in terms of test taker reactions and attitudes to the tests (Clark & Hooshmand, 1992; Hill, 1998; Stansfield et al., 1990). Clark and Hooshmand (1992) asked the candidates and examiners participating in a teleconferencing and face-to-face speaking test. According to their feedback on the testing process, it indicated that the semi-direct test was acceptable although there was a general preference for the direct test.

Stansfield et al (1990) investigated the extent to which a face-to-face oral proficiency interview (OPI) and a Simulated Oral Proficiency Interview (SOPI) were valid and acceptable to test takers. They discovered that while SOPI seemed to be valid and reliable, examinees did not react quite positively to it. A large majority of them preferred the live test, were more nervous in the tape version, and considered the

tape test more difficult. Stansfield (1991) suggested that the cause of such result may come from the less natural nature of the tape test which led to students feeling more stressful. In another comparison of candidate reactions to a live and tape-based oral test, **access**: test (the Australian Assessment of Communicative Skills in English), Hill (1998) demonstrated that examinees had a clear preference for the interview format although both test methods appeared to have face validity for the examinees. Moreover, females found the live test more difficult than the tape version and the Asian candidates felt more nervous during the interview.

There have been few studies on candidate reactions to a CBST. Kenyon and Malabonga, (2001) reported examinee attitudinal reactions to different speaking test formats of OPI, SOPI and COPI in Spanish, Arabic, and Chinese. All 55 students took both the SOPI and COPI. In addition, the 24 students in the Spanish study took OPI. The students completed a Likert questionnaire on six aspects of their attitudes towards and perceptions of the test after finishing each test. Then, after all the tests and questionnaires were administered, the students were asked to compare the SOPI and COPI directly on those six aspects. The results showed that examinees, particularly at the lower proficiency levels felt that the COPI was less difficult than the SOPI. In most other aspects, both tests were rated similarly. That is, none of the comparisons were statistically significant in terms of the test fairness, the opportunity for the candidates to adequately demonstrate their speaking ability, the feeling of nervousness while taking the tests, the clearness of the test directions, and the reflections of the accurate picture of their speaking ability. For the comparison of OPI, SOPI and COPI, all tests were rated similarly although the OPI appeared to be a better measurement of real-life speaking skills.

2.4 Research approaches and analyses

2.4.1 Generalizability theory

Generalizability theory (G-theory) is a measurement theory concerning with the dependability of scores, a measurement of behavior. That is, it focuses on how accurate one can generalize from one observation of test scores or other measures to a universe of admissible observations (Shavelson & Webb, 1991). G-theory provides us with a generalizability coefficient which reflects the level of dependability of scores.

This coefficient is analogous to reliability coefficient in classical test theory (CTT). However, unlike CTT, G-theory can estimate multiple sources of variation or measurement error in a single analysis such as inconsistencies among raters and inconsistencies of students' behavior across occasions. These sources or variation are called a facet, thus, in this example, raters and occasions are facets of the study.

Several designs are allowed to be investigated depending on the characteristics of the facet of interest (Shavelson & Webb, 1991). A facet may be crossed or nested, and random or fixed. In a crossed design, all conditions of one facet appear with all conditions of another facet. For example, in a design in which all students are rated by the same two raters, persons or students which are the object of measurement in this case are crossed with raters. In a nested design, one facet has two or more conditions which occur with one and only one condition of another facet. For instance, each student is rated by different pairs of raters. A facet can be treated as random or fixed. A random facet is when the sample of conditions of the facet is exchangeable for any other same-size set of conditions from the universe. For example, a rater facet is random if the two rates in a test can be exchanged for any other two raters. However, a facet is considered fixed if the conditions of the facet equal the number of conditions in the universe of generalization. For instance, if a test user does not want to generalize over other test prompts, the test prompts used in the study are a fixed facet.

There are two stages of investigation: a generalizability study (G-study) and a decision study (D-study) (Shavelson & Webb, 1991). The purpose of a G-study is to identify and estimate the magnitude of as many potential sources of variation or measurement error as possible. For example, in a one-facet G-study which aims to investigate how accurate the generalization is from a set of two raters rating students' English essays to all admissible raters, there are four sources of variability. They include systematic differences in the students' knowledge in English; differences in the stringency of raters; the interaction between students (persons) and raters; and finally random or unidentified source of variability. The third and fourth sources of variability cannot be distinguished. They are combined and called residual. The G-study gives estimated of the variance components of the persons, the rater facet and the residual.

The results from the G-study can be used in a D-study to determine appropriate measurement design which minimizes error and at the same time suitable for a particular purpose (Shavelson & Webb, 1991). Two kinds of the use of score dependability are offered in a D-study: relative and absolute decision, yielding the generalizability coefficient and the index of dependability, respectively. The relative decision is made in a situation where the test user aims to rank order individuals, while the absolute decision is used when they aim to index an individual's absolute level of performance regardless of how their peers perform. With a particular purpose in mind, one can conduct and compare several D-studies, each of which varies in the number of conditions of each facet. This allows us to select an optimal measurement design which suits our purpose.

Generalizability theory also makes possible an analysis of dependability of a profile of scores. The estimates include both the variance components associated with the object of measurement and the facets, and the covariance of these sources of variability among the subsets of scores. This would benefit the study using analytical ratings in written and spoken data (Shavelson & Webb, 1991).

2.4.2 Factory analysis

Factor analysis can be performed in an exploratory or confirmatory mode (Hair, et al., 1995). Exploratory Factor Analysis (EFA) is used to discover the underlying structure in a set of variables (Child, 1990). It does not impose any predetermined structure in the data. EFAs can be used to achieve the following purposes (Hair, et al. 1995). First, it can examine the structure of relationships in variables or respondents. Factors that underlie the data are identified and factor loadings are estimated. Second, EFAs can produce representative variables from a group of variables. The representative variables of a factor are those that have substantively higher factor loading than the others; they can be further used in other multivariate analyses. Finally, EFAs can create a new smaller set of variables to replace the original data. In this case, factor scores which are composite scores for each factor which represent each case are computed. Factor scores can be used in, for example, correlation, discriminant analysis and cluster analysis.

The assumptions in EFAs include conceptual and statistical ones (Hair, et al, 1995). The former is more critical than the latter. The conceptual assumption is there

is some underlying dimension in the set of variables of interest. The statistical assumptions include normality, homoscedasticity, and linearity. However, Hair, et al. (1995) explain that the departures from the statistical assumptions are not critical since they apply only to the degree they lessen the correlations. Also, normality is important only when one uses a statistical test to the significance of the factors, but the use of these tests are not common. In addition, some degree of multicollinearity is preferable since the purpose of factor analysis is to examine the intercorrelations among variables. Several statistical tests are available to assess the assumptions. The Kaiser-Meyer-Olkin (KMO) statistics can be used to check for this sampling adequacy (Hinton, 2004). It should be 0.5 or higher indicating the data are suitable for factor analysis. Another assumption which can help determine the appropriateness of using EFAs is there should be sufficient correlations of many variables. This can be tested by the Bartlett test of sphericity (Hair et al., 1995). The significant test shows that there are relationships among the variables. Finally, there must be adequate sample size. Many criteria have been proposed to determine the sample size. For example, there should be 10-20 cases per variables (Schumacker & Lomax, 1996). However, another criterion is the number of subject should be at least 5 times the number of variables, or 100 but when communalities are low and/or there are few variables loading on each factor, more cases are required (Hatcher, 1994 cited in http://www.2.chass.ncsu.edu/garson/pa765/factor/htm). Similarly, MacCallum, et al (2001) suggest that using the rule of thumb only may not be sufficient. When the mean communality is high, using four cases per variable may be adequate, but when it is low, much larger cases per variable are needed.

There are several options in conducting EFAs depending on the research objective (Hair, et al., 1995). For example, when one aims to identify underlying factors among the variables, the common factor analysis or principal factor analysis (PFA) extraction method should be selected since it reflects only the common variance of variables, not including the unique variance. Factor rotation methods should be considered also. They include an orthogonal and oblique one. The former is chosen when the factors are assumed to be uncorrelated, while the latter correlated. Hair et al (1995) suggest that the orthogonal is suitable when the purpose is to reduce a larger set of variables to a smaller one, but the oblique rotation is suitable to find theoretically meaningful factors since in reality very few variables are uncorrelated.

On the other hand, confirmatory factor analysis (CFA) is a statistical procedure for testing the extent to which sample data are consistent with or confirm a given set of hypothesis (Long, 1983). In conducting CFA, the researcher, first, posits a model generated from a theory. This model specifies the number of unobserved factors underlying the measured or observed variables, the exact relationships between the factors and variables and among the factors themselves. Then, the confirmatory factor model is identified, estimated, and assessed in terms of its overall fit. The identification of the model involves determining if there is a unique solution for the parameters. After the model has been identified, it can be estimated. Estimation involves using sample data which are collected, for example, from a language test to make estimates of population parameters. Finally, an assessment of model fit will be made to determine the model that best fits the data.

When CFA are performed, it is recommended that EFA should be done as a preliminary data analysis to explore the structure of the data (Kunnan, 2002). Unlike EFA, CFA can take into account the nonnormality of the data in estimating robust statistics (Byrne, 1994). These robust tests include Satorra-Bentler scaled chi-square (Satorra & Bentler, 1988) and robust standard errors (Bentler & Dijkstra, 1985).

CFA, which is a type of *ex post facto* correlational design, has been used recently in the language assessment research especially in construct validation studies because the advantages it has over experimental and quasi-experimental approaches (Bachman & Eignor, 1997). CFA has made it possible to investigate the relationships among a relatively large number of variables in a single study. Examples of research employing CFA are studies on construct validation of components of communicative proficiency (Bachman & Palmer, 1982; Fouly et al., 1990); and construct validation of the FSI oral interview (Bachman & Palmer, 1981) and self-ratings (Bachman & Palmer, 1989).

Bachman and Palmer (1982) using a multitrait-multimethod design and CFA examined the construct validity of some tests of communicative competence and of a hypothesized model. It was found that the model which best fit the data was the model consisting of a general and two specific trait factors, grammatical/pragmatic competence and sociolinguistic competence. In another study of construct validation, the researchers discovered that a model with two correlated trait factors and three method factors fit the data for the FSI interview (Bachman & Palmer, 1981).

In an investigation of the nature of second language proficiency, the findings showed that two models were good representation of the data: the model with separate traits or factors underlying performance on language tests, and correlating with each other; and the model with separate traits underlying performance on language tests, and influenced by a single higher-order factor (Fouly et al., 1990).

2.4.3 Verbal protocol analysis

Verbal protocol analysis is 'a methodology which is based on the assertion that an individual's verbalisations may be seen to be an accurate record of information that is (or has been) attended to as a particular task is (or has been) carried out' (Green, 1998 p.1-2). The informant usually either talks aloud or thinks aloud what s/he does to complete a task. This technique has increasingly been used in cognitive psychology, educational psychology, psychology of assessment, cognitive science, and social psychology (Green, 1998). In the area of language testing, verbal protocol analysis is currently used to supplement data obtained from quantitative techniques to investigate the validity of assessment methods and reliability of judgments (Banerjee & Luoma, 1997; Cohen, 2000; Green, 1998). For example, verbal report data has been used to explore the processes and strategies examinees use to complete a reading comprehension task (Anderson, Bachman, Perkins & Cohen, 1991; Cohen, 1984). The technique has also been employed to analyze the sequence of rating, the interpretations the raters made while using an analytic rating scale, and the difficulties raters had in rating written language performance (Lumley, 2002). Finally, insights into strategies in oral production tasks have been provided through verbal reports in studies by Cohen and Olshtain (1993), Cohen, (1998) and Cohen, Weaver and Li, (1998) (cited in Cohen, 2000).

Verbal protocol analysis can yield data on cognitive processes of the informants; however, to ensure that the technique and the data it provides are valid and reliable, researchers should be aware of the following aspects (Green, 1998). For example, the informant should be trained to give protocol report beforehand. The appropriate instructions should be used to guide the production of the verbal report. Moreover, the informant should be encouraged to express their thoughts but not to rationalize them. During the report, the researcher should not intervene the process but should only encourage the informant to keep speaking. Finally, the report should

be conducted during the informant taking the test because a delay may introduce an error in the data.

Little work has been done in using verbal protocols in a speaking test. However, Green (1998) suggests that a retrospective reporting technique should be applied in this condition. This is because an examinee has to converse with or without another interlocutor when performing a speaking test. To avoid disrupting this process, a retrospective report is recommended. Also, it should be conducted very soon after the test.

2.4.4 Discourse analysis for oral language assessment

Discourse analysis can be used in a validation study. For monologic test responses, Lazaraton (2002) recommends such approaches as genre, functional and structural analyses.

Bhatia (2002) has described genres as

"recognisable communicative events, characterised by a set of communicative purpose(s) identified and mutually understood by members of the professional or academic community in which they regularly occur" (p. 23)

Also, genres are highly structured although their lexio-grammatical and discoursal forms are dynamic (Bhatia, 2002). Genres can be analyzed in two levels (Bunton, 2002). In a micro level, genre analysis shows how certain grammatical features are used in various genres, and within the same genre. In a macro level, genre analysis explores the overall structures of a variety of genres. This structure presents the moves one makes to achieve one's communicative purpose in that genre (Bhatia, 1993). Examples of genres in the literature are narrative, description and exposition.

Storytelling genres can be further subcategorized as: narrative, anecdote, exemplum and recount (Plum, 1988 cited in Eggins & Slade, 1997). All of these share the same generic structures: abstract (a summary of the story to come), orientation (description of time, place, situation, characters) and coda (a part that evaluates the whole story and returns to the present), but the four subtypes differ in the middle phase. That is, narratives contain crisis and a resolution to that crisis. Anecdotes also focus on a crisis but an explicit resolution is not presented. Rather, such expressions as amazement, frustration and embarrassment are provided. Exemplums contain an

explicit moral message. Finally, recounts comprise retelling of events and sharing the speaker's emotional views to the story. They are not necessarily about a problem or crisis.

The structure of descriptions is a description of an object, place, etc (Martin, 1989). Both description and expositions do not have a set template (Hatch 1992). An example of a variant of expositions is provided by Martin (1989). An exposition may consist of thesis and argument. A thesis presents the speaker's judgment toward an issue, and an argument provides support to the judgment. Expositions can be subcategorized into an analytical exposition and a hortatory exposition. The former aims to persuade the audience that what is being argued for is well formulated. On the other hand, the latter's aim is to persuade the audience to do as recommended.

Functional analysis is a micro-structure analysis of text (Eggins & Slade, 1997). It aims to identify the speech function or purpose of a unit of utterance called move. Each move in a text has a specific speech function, for example, narrating, describing, giving opinion and justifying. Both genre and functional analyses look at the communicative purposes of the utterances but they differ in that genre analysis aims to identify the template or text structure which differentiates one genre from another (Hatch, 1992). Genre analysis is a macro-structure analysis of text. On the other hand, speech function analysis aims to identify the speaker's intent in a move (Eggins & Slade, 1997). The units of functional analysis are based on their grammatical and prosodic features. Thus, it is a micro-structure analysis of text.

Finally, structural analysis of text involves describing such features as grammar, and vocabulary associated with a particular text. For example, a story may contain past simple, copula sentences and relative clauses (Hatch, 1992).

2.5 Literature review and its relatedness to the study's hypotheses

From the review of the literature above, the hypotheses in this study were formed. The first hypothesis stated: the trait factors of oral language ability for Thai first-year university students are multidimensional. This was based on the research on trait factors of SL/FL ability (Bachman & Palmer, 1981, 1982; Carroll, 1983; Fouly et al., 1990; Sawaki, 1993). These studies found evidence of a general language ability factor which explains the intercorrelations among several specific factors. This may

suggest that more than one ability underlies the SL/FL performance; that is, the SL/FL ability is multidimensional. Thus, the hypothesis was stated accordingly.

The second hypothesis was: different types of test tasks in the CBST have a significant effect on the test scores. The literature has shown that many variables can have an impact on the examinees' performance in a speaking test (e.g. Bachman & Palmer, 1996; Chalhoub-Deville, 1996; Lazaraton, 1996; Skehan, 1998). One of these factors is test task characteristics, for example, setting, test rubric, input, expected responses and relationship between input and responses (Bachman & Palmer, 1996). Other studies have supported the impact of test methods or characteristics on oral test performance (e.g. Chalhoub-Deville, 1996; Kormos, 1999; Skehan, 1998). Therefore, in this study it was hypothesized that the CBST test tasks may have a significant effect on the test scores as well.

The final hypothesis was: there are no significant differences in terms of attitudes towards the CBST among examinees with different levels of speaking ability. Previous research has looked into attitude and reactions to several types of oral tests, for example, oral interview (Shohamy, 1982); semi-direct oral tests (Brown, 1993), and a comparison among OPI, SOPI, and COPI (Kenyon & Malabonga, 2001). However, there has been no study which compare that attitudes to a computer-based oral test among examinees at different oral proficiency levels. When there is no evidence which suggests which direction of the results may be anticipated, a nondirectional hypothesis should be used (Wiersma & Jurs, 2005). Thus, the third hypothesis was set accordingly.

2.6 Summary of the literature review

The literature review discussed above has shown a recent body of research on the features of oral language proficiency, computer delivered language assessment, oral language assessment, factors influencing speaking performance, and research methods used in language test validation studies. It can be seen that oral language proficiency is multi-componential and performance in a speaking test is multifaceted as well. Test takers' performance is influenced not only by their language ability but also by other personal characteristics as well as the nature of test tasks. One of the test task features is test method facets which cover such aspects as test rubric, input, testing conditions and raters. The assessment of oral proficiency which employs

emerging technology such as the computer, then, needs to be investigated in terms of the effects of new test methods on test takers' performance. In order to study these influences, more and more researchers have recommended using both quantitative and qualitative approaches such as confirmatory factor analysis and verbal protocol analysis. One of the advantages of combining these methods is to triangulate the data collected. This will enable us to gain deeper insights into how accurately we can interpret test scores from computer-delivered tests as meaningful and appropriate indicators of oral language ability. The next chapter presents the research methodology of the present study.

CHAPTER III
RESEARCH METHODOLOGY

This chapter presents the methodology used in this current study.

3.1 Research design and approach

The research design and approach was set to explore the four research questions as specified earlier. The questions included:

1. What is the trait structure of oral language ability in a Computer-based Speaking Test (CBST) for Thai first-year university students?
2. What processes do the examinees employ in taking the CBST?
3. To what extent do different types of CBST tasks affect the test scores and characteristics of speech produced?
4. To what extent do examinees with different levels of speaking ability differ in terms of their attitudes towards the CBST?

The research design was descriptive, correlational and "ex post facto" designs (Isaac & Michael, 1995) to examine these questions. Specifically, it used a multitrait-multimethod (MTMM) design (Campbell & Fiske, 1959). Both quantitative and qualitative data were collected. First, the relationships among the test scores were investigated through Confirmatory Factor Analysis (CFA) to answer the first research question. In addition, a collection of verbal reports on test-taking processes and strategies was employed to answer the second question. The CFA, Repeated-measures MANOVA and discourse analysis were carried out to answer the third question. Finally, independent MANOVA and content analysis of opinion to the CBST were used to explore the fourth research question.

3.2 Population and sample

The population was non-native English speaking university students in a country where English is taught and used as a foreign language. The sample was selected from Thai first-year Chulalongkorn University (CU) students in the academic year 2005 by a purposive sampling technique. The total number of students in this group was 4,969. The students were studying in diverse academic fields: biological

science (Faculty of Allied Health Sciences, Faculty of Dentistry, Faculty of Medicine, Faculty of Pharmaceutical Sciences, Faculty of Veterinary Science, and School of Sports Science), social science (Faculty of Commerce and Accountancy, Faculty of Communication Arts, Faculty of Economics, Faculty of Education, Faculty of Law, and Faculty of Political Science), physical science and technology (Faculty of Architecture, Faculty of Engineering, and Faculty of Science) and humanities (Faculty of Arts and Faculty of Fine and Applied Arts).

The sample of the current study was 167 students: the data from 158 students were used for the analyses of test scores, discourse of the responses and attitude; and the data from additional 9 students for the verbal protocol analysis. The selection criteria were their general English language proficiency measured by the Chulalongkorn University Test of English Proficiency (CU-TEP) and their academic fields. First, the first-year CU students in Year 2005 were categorized into three general English proficiency levels according to their CU-TEP scores. The mean score was 473.34 and the standard deviation (S) was 49.79. Those whose scores were between -1 S to -2 S were grouped as the lower intermediate group, those between -.5 S to .5 S were the intermediate group, and those between 1 S to 2 S were the advanced group. The students who scored lower than -2 sd, and higher than 2 sd were not included because they may be extreme cases toward both ends. After the scores were set, an attempt was made to select the sample from all faculties. Since the students' class schedules varied and were in conflict, only those from the following faculties were contacted: the Faculty of Commerce and Accountancy, Economics, Education, Law, Science and Allied Science. These faculties were generally in the social science and science. In each faculty, students whose scores fell in the three proficiency groups were identified. Since there were many class sections in a faculty, only a few could be approached. Their English class schedules were checked and some arrangements with the class teachers were made. A total of 250 students in 14 sections were contacted and asked for their participation in the research study. The researcher explained to them the research purposes, the test, and procedure for data collection. Only 233 students agreed to participate in the study. However, during the data collection, some students could not come to the session although some rearrangements were made. Some data were missing or had poor recording quality. Thus, there were 158 students with complete data for the analyses of test scores, responses, and attitudes.

A total of nine examinees gave verbal reports on the processes and strategies they used when performing the CBST. The criteria in selecting were based on their CU-TEP scores and the faculties they were studying in. Three were from the advanced group, three the intermediate and three the lower intermediate. In each proficiency group, one was from the Faculty of Science or Allied Science, one from the Faculty of Education or Law and one from the Faculty of Commerce and Accountancy or Economics.

3.3 Stages of research

Three stages were involved in the study. First, the instruments were developed and validated by three content experts in the applied linguistic field. Second, the pilot study was conducted. Its purpose was to try out the instruments, data collection methods and data analytical methods, especially the qualitative approach. The participants in the pilot study was 30 first-year CU students in the academic year 2004. Although they were not those in Year 2005, it could be assumed that they shared some similar characteristics. That is, those in the pilot study came from similar faculties and had similar scores in the English oral proficiency. After the pilot study, some revisions were made to test administration, test tasks and rating scales. Finally, the main study was carried out. That is, the test and questionnaire were administered to the participants. After this phrase was complete, the verbal protocol data were collected through a retrospective stimulated report interview.

3.4 Research instruments

The instruments used in the study were developed by the researcher. These were a computer-based speaking test, computer-based speaking test rating scales, and a computer-based speaking test questionnaire. The development of the first two instruments was based on the framework of test development (Bachman & Palmer, 1996; Lazaraton & Riggenbach, 1990) and computer-based test design (Fulcher, 2003). The development of the third instrument followed the guidelines by Isaac and Michael (1995). After the development, the instruments were evaluated for their content relevance and appropriateness by three content experts in the applied linguistics field. Then, they were revised and tried out in the pilot study. The test was revised again before being used in the main study.

3.4.1 The computer-based speaking test

The CBST was a multi-media English oral proficiency test which was delivered by a computer (Appendix A). The computer presented test instructions, test tasks, controlled the preparation and response time, and stored test takers' responses. There was both text and audio input. The examinees wore a headset and spoke to a microphone. This CBST was administered in a language lab by the researcher. It lasted approximately 25 minutes.

There were four test tasks, representing four different task types: narrating the most enjoyable day on campus, expressing opinions towards studying for a Bachelor's degree abroad, describing an imaginary ideal world, and persuading foreigners to visit Thailand. In each task, the test topic, the person they were talking to, and sub questions were specified. These tasks required examinees to use a variety of language functions. The test topics covered general topical knowledge, that is, enjoyable personal experience as university students, opinions towards studying abroad, imagination about their surroundings, and general information about their country, especially tourist attractions.

After the test was developed, it was evaluated by three content experts as mentioned earlier. The experts were asked to give their comments on the appropriateness of the test as an instrument to measure the constructs (knowledge of syntax, vocabulary, pronunciation, cohesion and functions) as well as to elicit adequate speech samples to be rated, and the appropriateness of the test content, the instructions, the preparation time, the response time, and the interface of the program. At least two out of three experts agreed on the appropriateness of the CBST in all aspects. Thus, there was no major revision. A minor change in the sample test was made to make this part format parallel to the actual test tasks. The interface design in the paper-based form was also evaluated by the experts. There was no major suggestions. After the sample test format was revised, the design was reevaluated and no major comments were found.

During the test administration in the pilot study, most of the examinees spent about 30 seconds to one and a half minute responding to the tasks; the original response time was two minutes. Their answers were examined and it seemed to me that only one and a half minute response time would be enough. The answers were long enough to be rated and the examinees would not feel stressed waiting for the

response time to end. Another observation from their answers occurred when I was doing the speech function analysis on their responses. One of the speech functions in the original imaginary, opinion and persuasive tasks was the same, justifying. There was also a comment on the appropriateness of the imaginary task from the committee members. So I decided to revise this task so that it would generate different speech functions and be more related to the imaginary situation. The content of the task was kept similar to the original one since the examinees in the pilot study seemed to be able to talk about it. The latest version of the CBST was in Appendix A.

Finally, the comments the examinees gave about the test were checked. These comments were in the open-ended part of the questionnaire. The examinees seemed to like the interface of the program. The program seemed to work smoothly. Thus, the design was not changed.

To investigate the internal consistency of the CBST, the Cronbach's coefficient alpha was estimated. The standardized alpha for the 20 measures was .943, indicating high reliability.

3.4.2 Rating scales

Rating scales in the study were analytical scales for the descriptions of oral language proficiency (see Appendix B). The scales based on the model of language ability developed by Bachman and Palmer (1996) and Lazaraton and Riggenbach (1990) were used to score speech samples from the CBST. They consist of scales measuring knowledge of vocabulary, knowledge of syntax, knowledge of phonology, knowledge of cohesion and knowledge of functions.

After the rating scales were developed, they were evaluated by the experts, revised and tried out in the pilot study. At least two experts agreed with the use of the scales in all aspects; there was no major revision. After being tried out, the functional scales were revised to make the descriptors clearer and suit with the new imaginary task. The raters were also asked to give feedback on the use of the scales. No major comments were found.

3.4.3 A Computer-based Speaking Test Questionnaire

A Computer-based Speaking Test Questionnaire was a set of statements and questions which aimed to gather information about test takers' opinions towards the

computer-based speaking test (see Appendix C). It consisted of a Likert scale and open-ended questions. After the development, the questionnaire was evaluated by the experts, revised, and tried out in the pilot study.

There were some suggestions from the experts on the appropriateness of some questions in the questionnaire, especially the open-ended part. The questionnaire was then revised on that part and reevaluated by the experts. During the pilot study, there were no major comments from the examinees. As the imaginary task and the length of the response time were changed, the questionnaire was rewritten accordingly.

After the main study data collection, the Cronbach alpha was used to estimate its internal consistency. The Cronbach's coefficient alpha based on standardized items was estimated to be 0.776, indicating acceptable reliability (Hair, et al., 1995).

3.5 Data collection

3.5.1 Quantitative data collection

The quantitative data comprised of CBST test scores, and attitudes towards the CBST from the Likert-scaled questionnaire. The data were collected during the administration of the test and questionnaire which took place from mid August to mid September 2005. Due to the availability and capacity of the computer lab and the examinees' schedules, the data could not be collected in one time. Before the administration, the test order was counterbalanced to reduce task order effect and was given to the examinees randomly. The four test task orders and the number of examinees with complete data were shown in the table below.

Table 3.1: Orders of test tasks

Groups	N (158)	Tasks performed in order			
		Task 1	Task 2	Task 3	Task 4
Group 1	41	Narrative	Opinion	Imaginary	Persuasive
Group 2	37	Opinion	Narrative	Persuasive	Imaginary
Group 3	44	Imaginary	Persuasive	Narrative	Opinion
Group 4	36	Persuasive	Imaginary	Opinion	Narrative

Group 1 (41 examinees) performed the narrative task first, followed by the opinion, imaginary, and persuasive task, respectively. Group 2 (37 examinees)

performed the opinion task first and then the narrative, persuasive, and imaginary task, and so on.

After the test administration, the questionnaire was given out to the examinees and returned to the researcher in the same session. There were 152 examinees who completed both the CBST and the questionnaire. Since the attitudes would be compared among three oral language proficiency groups, these examinees were categorized into three groups. The criteria to assign them into groups were their CBST total score percentiles after the outliers were removed and equality of the number of examinees per group. The three groups were the lower intermediate group (LI), the intermediate group (INT), and the advanced group (ADV). The LI group consisted of 37 examinees whose percentiles ranged from 1-25, or scores of 55-67.5. The INT group were those whose percentiles between 35-65, or scores of 69-73. There were 44 students in this group. To maintain the equal number of 37 in all groups, 7 examinees at the top and bottom percentiles were randomly selected and discarded. Finally, the ADV group was those with the percentiles of 75-100, or scores 75.5-83. There were 39 examinees in this group; thus, two examinees at the percentile of 75 were randomly selected and discarded. The means of the total scores of the LI, INT, and ADV groups were 63.6, 70.9, and 78.1, respectively. A one-way ANOVA was performed to examine if the three groups were significantly different in their CBST proficiency levels. The results showed significant differences ($F_{(2,108)} = 429.756$, $p < .001$). Post hoc analyses (Tukey) indicated that all pairs were significantly different at $p < .001$. (see Table 3.2). That is, the mean of the LI group was lower than that of the INT group, which was lower than that of the ADV group.

Table 3.2: Post hoc analyses of the CBST total scores

	Group	N	Subset for alpha = .05		
			LI	INT	ADV
Tukey HSD	LI	37	63.6		
	INT	37		70.9	
	ADV	37			78.1
	Sig.		1.00	1.00	1.00

3.5.2 Qualitative data collection

The qualitative data consisted of verbal reports, responses to the test tasks, and attitude towards the CBST from the open-ended questions in the questionnaire.

3.5.2.1 Verbal reports

The selection of the participants for the verbal reports was based on the CU-TEP scores and their faculties as described earlier. The verbal reports were collected from a retrospective stimulated report interview in Thai. The data collection method was revised a few times. The initial procedure was adapted from the method used by Cohen and Olshtain (1993), and O'Loughlin (2001), and tried out in the pilot study (Appendix D). The method was presented and discussed with a group of a language assessment professor and Ph.D. students. There were comments that the questions used to elicit the test-taking processes and strategies may be too limited. It was then revised to include only the self reports without the questions from the researcher. The participants were first trained to give retrospective verbal reports before doing the test task and reporting their thoughts. After trying out the new method with a few participants, the researcher speculated that many participants may have forgotten to report several thoughts although in general they could report what they were thinking of during the preparation and response time. From the researcher's observation, there were some occurrences in which they paused or changed their responses to the test task but they did not mention the thoughts that might have happened during these periods. Therefore, after they finished reporting, the researcher asked them to go back to review these parts. It was found that they could recall more information. The data collection method was then revised.

In the latest procedure which was used in the main study, the participants in the protocol collection would be explained the purpose of the session and the description of the CBST. They were next trained how to give the verbal reports and did some practice tasks (see Appendix E). They could ask further questions about the procedures. After they said they were ready, the session began with the administration of the CBST. The participants did the test individually in a computer laboratory. The test procedures and environment were similar to the test takers who did not give verbal report. However, in addition to computer recorded, the responses were audio recorded so that they can be played back during the report.

After the test takers finished each test task, they would give their reports in Thai which was audio recorded. In order to help them recall the thoughts that happened during the preparation and response time, they could review their notes and listened to their own responses to the test task and stopped the tape any time. During

the report, the participants were first asked to start talking about the first thing they thought of. They were encouraged to continue their talk when they were silent. When the researcher speculated that there might be some missing information, she would help them recall by asking them to review these points. The researcher also asked them to clarify their thoughts if some were unclear. At the end of each report, she asked some fixed questions regarding their thoughts about English grammar, vocabulary and pronunciation.

The verbal reports in the main study were collected from the third week of December 2005 to the end of January 2006. They lasted from 45 minutes to two hours depending on how long the participants provided the reports.

3.5.2.2 Test responses

The test responses were recorded in the computer when the examinees were doing the test. A note was made regarding the topics of the narrative task the examinees produced. The prompt asked the examinees to talk about their most enjoyable day on campus. However, many responses were about enjoyable stories that did not actually take place on campus. For example, some were about going to a welcome camp with their classmates and seniors, and some about the extracurricular activities they did such as the Chula-Thammasart Football Competition. The researcher suspected that the examinees may interpret 'on campus' as 'university related' or they may pay more attention to the sub questions such as 'where it happened', 'what happened first', which did not restate the topic 'on campus story'. Although the topics varied, they shared the main characteristics of telling the most enjoyable experience. Also, they did not differ from other on campus stories in terms of genre, speech functions and grammatical features. This will be shown in the results of this research question in the next chapter. Thus, they were not discarded.

The total number of the responses to be analyzed in terms of their discourse was 36, that is, four responses from nine participants. These participants were selected from the three proficiency groups as categorized according to their CBST test scores described in the previous section. Three examinees were selected from the advanced group, three from the intermediate and another three from the lower intermediate group. In each proficiency group, the first selection criterion was their CBST subscores in syntax, vocabulary, cohesion, and function. Their pronunciation

subscores were not considered since the responses would not be analyzed in terms of prosodic features. The three advanced examinees had high scores in syntax, vocabulary, and cohesion subscores, the intermediate had scores in the middle range and the lower intermediate low scores. Another criterion was the selected participants should produce adequate speech to be analyzed, that is, their function subscores should be equal to or higher than 2.5 in each task. When possible, participants from different faculties were chosen.

3.5.2.3 Open-ended questions

The last qualitative data were the attitude towards the CBST from the open-ended questions in the questionnaire. The data came from the three proficiency groups described earlier.

3.6 Ratings of test scores

3.6.1 Scoring procedures

Two raters rated the CBST responses. The researcher was one of the raters; another was an experienced English language instructor who had taught several oral language communication classes. After the pilot study, the raters tried out the rating scales with some responses and discussed problems that occurred. As the rating scales were revised for the main study, the raters met, tried out the new scales with some responses and discussed the scores assigned and comments. After that, they rated the rest of the responses independently (see Appendix F for sample responses and their test scores).

3.6.2 Dependability of ratings

To investigate the reliability of scores from the two raters evaluating the performance of the examinees taking the CBST, multivariate generalizability analyses with a *person by rater design* were conducted for each test task. First, a $p\bullet$ x $r\bullet$ Generalizability study (G-study) was carried out to obtain estimates of variance-covariance matrices. The matrices were used in a Decision study (D-study), in which the number of rates was specified at two since in this study two raters evaluated the examinees' performance. The results of the D-study from each test task and the

generalizability coefficients which showed the degree of the dependability of the measures were reported below.

In Table 3.3, the estimated variance and covariance components and generalizability coefficients for narrative task scores were presented. The bold values in the diagonal are variances with the proportions of total variance accounted for by each facets of measurement in parentheses. The values in the lower diagonal are covariances, and the upper diagonal values are correlations among the five subscales.

Table 3.3: Estimated variance and covariance components and generalizability coefficients for narrative task scores (raters = 2)

Source of variation		Pro	Syn	Voc	Coh	Func
Person	Pro	**.182** (68.4%)	.730	.684	.532	.424
	Syn	.117	**.142** (68.1%)	.580	.614	.494
	Voc	.126	.094	**.187** (80.8%)	.727	.538
	Coh	.088	.090	.123	**.152** (71.1%)	.481
	Func	.100	.104	.130	.105	**.312** (83.2%)
Rater	Pro	**.028** (10.5%)				
	Syn	.021	**.016** (7.5%)			
	Voc	.011	.008	**.004** (1.6%)		
	Coh	.020	.015	.007	**.013** (6.3%)	
	Func	.006	.005	.002	.004	**.001** (.3%)
PR	Pro	**.056** (21.1%)				
	Syn	.003	**.050** (24.4%)			
	Voc	.004	-.003	**.040** (17.5%)		
	Coh	.004	.004	.013	**.048** (22.6%)	
	Func	.002	-.008	.006	.003	**.062** (16.5%)
Generalizability coefficient		.76	.74	.82	.76	.83

In each subscale, the variance components from persons accounted for the largest percentage (68.1%-83.2%). This showed that the examinees systematically differed in their knowledge of pronunciation, syntax, vocabulary, cohesion and function in the narrative task. The second largest percentages of variance were from the *p x r* interactions plus unidentified error (16.5%-24.4%). These relatively large

percentages of variance indicated that some proportion of the variance in the total scores on them were due to two sources, that is, the raters' differences in rank ordering of certain examinees and unexplained source. The smallest variance components were for the rater from all subscales (0.3%-10.5%) indicating that the two raters differed very little in their leniency when scoring the five subscales.

The covariance components provided further information about the sources of error covariation among the subscales. The relative large covariance components for persons (.088-.130) showed positive correlations among the five subscales. That is, the examinees who received high scores on one subscale also received high scores on the other subscales, and vice versa. The covariance components for raters were negligible (.002-.021). This means that the raters' difference in the degree of leniency on one subscale was not related to that on the other subscales. Finally, the residual covariance components were in general negligible also (.002-.013). This suggested that the disagreement in the ratings of some examinees on one subscale and the unexplained source of variation in that subscale did not contribute to these two sources of variation on the other subscales.

The universe score correlations among the five subscales (the upper diagonal values) were moderate (.424-.730). That is, the five subscales were related but distinct. The generalizability coefficients were acceptable since they were .70 or above (Ary, Jacobs & Razavieh, 2002): 0.76 for pronunciation, 0.74 for syntax, 0.82 for vocabulary, 0.76 for cohesion, and 0.83 for function scores.

The similar patterns were found in the other three tasks. Appendix G presents the estimated variance and covariance components and generalizability coefficients for opinion, imaginary and persuasive task scores, respectively. The patterns of the variance components for persons, raters and residuals were quite similar except that in the opinion task, the rater variance component in the function subscale was zero. This indicated that the raters were somehow different in their leniency when scoring the other four subscales but they were not different when scoring the function subscale in the opinion task.

The generalizability coefficients for most subscales in the three tasks were also acceptable, except two coefficients which were close to 0.70. They ranged from 0.67-0.89 in the opinion task, 0.73-0.81 in the imaginary task, and 0.68-0.83 in the persuasive task.

3.7 Data analyses

3.7.1 Data analyses for research question 1

The first research question asked: What is the trait structure of oral language ability in a Computer-based Speaking Test (CBST)? The data analyses for this question consisted of descriptive statistics, exploratory factor analyses (EFAs) and confirmatory factor analyses (CFAs).

First, the descriptive statistics were examined to check the assumptions of the CFA. After that, EFAs were carried out through SPSS for Windows Version 13.0 (SPSS Inc., 2004). Principal axis factor analysis was used initially to determine the eigenvalues of the covariance matrix (Tabachnick & Fidell, 2001). Based on the sizes of eigenvalues and the scree test (Hinton, 2004; Tabachnick & Fidell, 2001), a solution with three factors was initially suggested. Several principal axis factoring runs with orthogonal and oblique rotations specifying different numbers of factors were conducted (Tabachnick & Fidell, 2001). The oblique rotation was selected at the end because the orthogonal solutions provided unclear interpretations, the oblique solutions showed that the factors were correlated moderately (e.g. -.467), and several of the previous studies (Bachman & Palmer, 1981, 1982; Carroll, 1983; Fouly et al., 1990; Sawaki, 1993) showed that second language ability seemed to be multicomponential and correlated and the oblique solutions gave moderate factor correlations (e.g. -.467). The oblique EFA results were next compared to find the solution that was the most interpretable and had simple structure with factor loadings of at least 0.30 (Hair, et al., 1995), which was selected as one of the models to be tested in the CFA.

After the EFAs were carried out, the CFAs were performed in EQS for Windows Version 6.1 (Bentler & Wu, 2002), with the robust maximum likelihood estimation of the covariance matrix since the data were not multivariate normal. Four models were tested. The constructions of the models were based on the following reasons. The first model was based on the EFA results (Kunnan, 2002). It was a two-correlated factor model. Model 2 was based on the CBST design. It consisted of a second-order trait factor, five first-order trait factors and four method factors. Model 3 was a second-order factor underlying five first-order trait factors without the method factors. This model was based on the hypothesis that the method factors may not

exist. This was partly due to the results from the Wald test from Model 2 and the discourse analysis results which showed that the test tasks shared some similarities. The last model was a one factor model. This model was selected to test whether the speaking ability was unidimensional (Oller, 1983) or multidimensional (Bachman & Palmer, 1981, 1982; Carroll, 1983; Fouly et al., 1990; Sawaki, 1993). The goodness of fit results were compared and model modification was performed based on the Lagrange multiplier (LM) test and Wald test (Tabachnick & Fidell, 2001).

Model fit was evaluated using the following criteria: the Satorra-Bentler chi-square test statistic, its significance level, and its ratio to the degrees of freedom of the model; the non-normed fitted index (NNFI); the comparative fit index (CFI); and the root mean standard error of approximation (RMSEA). Other criteria included the average absolute and average off-diagonal absolute standardized residuals, the largest standardized residuals, the histogram of the distribution of standardized residuals, and the reasonableness and significance of the individual parameter estimates (Byrne, 1994; Schermelleh-Engel, Moosbrugger & Muller, 2003; Tabachnick & Fidell, 2001).

3.7.2 Data analyses for research question 2

The second research question asked: What processes do the examinees employ in taking the CBST? The data came from audio-recorded verbal reports. They were segmented into words, phrases or sentences, each of which represented a distinct process or strategy, and were assigned a code or taxonomy related to test-taking processes and strategies.

The coding schemes were developed from those suggested by previous research (Bachman & Palmer, 1996; Færch & Kasper, 1983) and by the data from the pilot study. They consisted of taxonomies of speaking processes and strategies to deal with communication problems. To try out the scheme, three verbal reports from the main study were selected. Each represented each proficiency group. One of the four sections from each report was analyzed by the researcher and another coder, an experienced applied linguistics researcher and lecturer. The two coders independently coded the reports and discussed the results. In general, they agreed on the coding but there were also some limitations and discrepancies. The coding scheme was then revised to include more taxonomies for the main study (Appendix H). As for the discrepancies, some were not resolved between the coders. These data could be

interpreted in more than one way because the context may not be specific or may be too little to pinpoint the exact code. Thus, they had to be left with two interpretations.

To investigate the consistency of the coding, the other three sections of the reports were coded independently by the codes using the revised scheme. Since one of the protocols of one participant was not clear in the last section, her three reports were replaced by a new participant's four sections. There were a total of 273 segments. The coders' agreement on identifying the segments was 254 segments (93.04%). Of the 254 segments, the coders agreement on coding was 178 segments (70.08%) and the disagreement was 76 segments (29.92%). The discrepancies may be due to the lack or inadequacy of the context in the verbal reports, leading to more than one interpretation of the processes or strategies the participant was employing.

3.7.3 Data analyses for research question 3

The third research question asked: To what extent do different types of the CBST tasks affect the test scores and characteristics of speech produced? This question was investigated by two means: quantitative and qualitative analyses. For the quantitative analyses, two methods were employed: CFA and repeated measures multivariate analysis of variance (RM MANOVA). The CFA was presented earlier. The RM MANOVA is presented in this section while the qualitative analysis, discourse analysis, is in the next section.

3.7.3.1 Quantitative data analyses for research question 3

A RM MANOVA was performed to test the hypothesis that different types of test tasks in the CBST had a significant effect on the test scores. The one between-subjects factor was the task order with four levels of information as presented in Table 3.1. There were five dependent variables: pronunciation, syntax, vocabulary, cohesion, and function subscores. Each subscore was measured by the four tasks. The one within-subjects factor was task type, that is, narrative, opinion, imaginary, and persuasive. The results would be examined in three aspects: the interaction effect between the task order and task type factor; given no interaction, the differences of each dependent variable among the task orders; and given no interaction, the differences of each dependent variable among the task types. The last aspect would be the focus of the research question.

After performing the RM MANOVA, it was found that one of the assumptions, the homogeneity of variance-covariance matrices, could not be checked. There was a warning message that "Box's Test of Equality of Covariance Matrices is not computed because there are fewer than two nonsingular cell covariance matrices." Therefore, the RM Analysis of Variance (ANOVA) was selected to be carried out instead of the multivariate one. Since it was a mixed design, the assumptions to be inspected were multivariate normality; the homogeneity of variance using the Levene's Test of Equality of Error Variances; and homogeneity of covariance using the Mauchly's Test of Sphericity (Tabachnick & Fidell, 2001).

As for the within-subjects effects, the F statistics were inspected to see if there were significant interaction and main effects for each dependent variable. In the case of the violation of sphericity, the significant tests which are adjusted for the violation such as Greenhouse-Geisser or Huynh-Feldt can be used. When the epsilon is less than .75, the value from the former test is selected. When the epsilon is greater than or equal to .75, that of the latter test is recommended (Barcikowski & Robey, 1994). If the univariate test showed significant main effects differences, the Bonferroni test was then applied to see which pair of means was different (Hinton, 2004). This test was selected because it can control for the overall Type I error rate when multiple comparisons are being made (Tabachnick & Fidell, 2001).

As for the between-subjects effects, the F statistics were examined to see if there were overall significance of task order on each dependent variable. If significant differences were found, post hoc analyses would be performed to see where the significance was located.

In addition, if significant task type effects were found for the 140 examinees (the number of the cases after the outliers were discarded), RM MANOVAs were next carried out for each group of examinees which performed the test in different orders to see if the significant effects were the same across the groups. In each group, the subscores were dependent variables, and the four tasks repeated measures. The multivariate test, Wilks' Lambda, was examined to see if overall mean differences existed. If they did, the sphericity test would be inspected. Then, the univariate tests which investigated each dependent variable individually would be checked. If mean differences in a dependent variable were found, the Bonferroni test would be conducted (Tabachnick & Fidell, 2001).

3.7.3.2 Discourse analyses for research question 3

The data were analyzed in terms of three discourse features as suggested by Lazaraton (2002). These features included genre, speech functions and grammatical elements. The first two features were appropriate in the current study because they can be applied with monologue which is the characteristics of the CBST responses. The grammatical features selected to be analyzed were tense, modal and conjunctions since these seemed to vary across the CBST tasks and shaped by the task conditions.

First, the data were transcribed by the researcher. To make the transcripts practical and appropriate for the analytical approaches intended, all the spoken words were included but the details of the sound of speech were not (see Appendix I).

Having been transcribed, the transcripts in each analysis were segmented into different units and assigned a code. The units of analysis were based on the suggestions given by Eggins and Slade (1997). The coding schemes were developed. They were based on the work by Hatch (1992), Martin (1989) and Plum (1988) cited in Eggins and Slade (1997), and the analysis of a few speech samples in the main study. The steps in the data analyses and coding schemes for each are presented in Appendix J.

To investigate coder consistency, six out of nine transcripts were segmented and coded by two coders independently. The first coder was the researcher. The second coder for the genre and speech function analyses was an English language university lecturer with an experience in ethnography and discourse analysis. The second coder for the grammatical features was an experienced English language lecturer. The researcher explained the purpose of the current research, procedures and explanation in segmenting, identifying, and coding the responses. The coders tired out the coding scheme and discussed any problems or questions they had. Then, they coded the material independently and discussed the results. The coding agreement was 82.1%, 71% and 91.2% for the three types of analyses, respectively. The majority of the discrepancies could be resolved after the discussion.

3.7.4 Data analyses for research question 4

The fourth research question aimed to explore the extent to which the examinees with different levels of speaking ability differed in terms of their attitudes toward the CBST. There were both quantitative and qualitative analyses to answer

this question. The quantitative one used independent measures MANOVA for the Likert-scaled questionnaire items, and the qualitative one used content analysis of the open-ended items.

3.7.4.1 Quantitative data analyses for research question 4

Before the analysis was carried out, negatively worded items were scored in the reverse. High scores then were indicative of positive views toward the CBST and low scores negative views. The mean scores were interpreted as follows:

1.00-1.49	Strong, negative view
1.50-2.49	Somewhat negative view
2.50-3.49	Somewhat positive view
3.50-4.00	Strong, positive view

Independent measures MANOVA was used to investigate the differences in the three groups' attitudes towards the CBST. The independent variable was the proficiency levels with three conditions: lower intermediate, intermediate, and advanced levels. The dependent variables were 22 items in the Likert scale questionnaire.

The assumptions of MANOVA to be tested included multivariate normality, and homogeneity of variance-covariance matrices. The multivariate test, Wilks' Lambda, would be examined to see if the effect of the proficiency levels on the answers to the items was significant. If significant, univariate ANOVAs for each dependent variable would be inspected. To control for the increased risk of Type I error, the Bonferroni correction would be made on the significant level (Tabachnick & Fidell, 2001). To maintain the overall alpha to be .05, the alpha for each dependent variable was set at .05 divided by the number of dependent variables, that is, .05/22 or .002 (Tabachnick & Fidell, 2001). When the Bonferroni test suggested significant differences, post hoc analyses would be performed to find where the significance was located.

3.7.4.2 Content analysis for research question 4

There were three open-ended questions in the questionnaire. The first one asked those who strongly disagreed or disagreed with the statement "I feel I had the opportunity to adequately demonstrate my speaking ability on the computer-based speaking test" to give their reasons. The second question asked what they liked about the CBST and the last one what they disliked about the test. The answers of the 111 students in the three proficiency groups were grouped and categorized according to the aspects of the CBST they expressed their views about. Next, frequency counts of these aspects were carried out. The data were not categorized according to the proficiency groups because the ANOVAs results in the previous part showed that the groups were not significantly different in their views to the CBST.

In conclusion, this chapter presents the research methodology for the study. The next chapter shows the research findings.

This chapter presents the results of the research questions 1-4.

4.1 Results of research question 1

Research Question 1 was: What is the trait structure of oral language ability in a Computer-based Speaking Test (CBST) for Thai first-year university students?

Descriptive Statistics

The complete data for the EFA and CFA were obtained from 158 examinees. Since the test scores were obtained from two raters, they were averaged and used in the data analyses. Table 4.1 shows the 20 measured variables with their names and labels used in the models. For example, ProNar was the mean pronunciation score in the narrative task, ProOp was the mean pronunciation score in the opinion task, and so on.

Table 4.1: Measured variables in the models

Measured variables		Labels
Pronunciation scores in	Narrative task	ProNar
	Opinion task	ProOp
	Imaginary task	ProIm
	Persuasive task	ProPer
Syntax scores in	Narrative task	SynNar
	Opinion task	SynOp
	Imaginary task	SynIm
	Persuasive task	SynPer
Vocabulary scores in	Narrative task	VocNar
	Opinion task	VocOp
	Imaginary task	VocIm
	Persuasive task	VocPer
Cohesion scores in	Narrative task	CohNar
	Opinion task	CohOp
	Imaginary task	CohIm
	Persuasive task	CohPer
Function scores in	Narrative task	FuncNar
	Opinion task	FuncOp
	Imaginary task	FuncIm
	Persuasive task	FuncPer

The descriptive statistics of all variables mentioned above are presented below.

Table 4.2: Descriptive statistics of all variables

Variables	Min	Max	Mean	Std. deviation	Skewness (std. error = .193)	Kurtosis (std. error = .384)
Pronunciation						
ProNar	2	4.5	3.364	.488	-.096	-.571
ProOp	2	4.5	3.370	.489	-.117	-.584
ProIm	2	4.5	3.367	.487	-.113	-.555
ProPer	2	4.5	3.377	.481	-.136	-.499
Syntax						
SynNar	2	4	3.165	.439	.218	.409
SynOp	2	4	3.266	.484	.121	-.461
SynIm	2	4	3.035	.348	.699	2.994
SynPer	2	4	3.253	.485	.277	-.626
Vocabulary						
VocNar	2	4	3.633	.477	-.993	.129
VocOp	2	4.5	3.627	.482	-1.123	.880
VocIm	2	5	3.532	.546	-.860	.672
VocPer	2.5	5	3.684	.473	-.563	-.134
Cohesion						
CohNar	2	4	3.699	.448	-1.723	3.219
CohOp	2	4.5	3.715	.412	-1.281	2.346
CohIm	2	4	3.525	.501	-1.131	1.196
CohPer	2.5	4	3.677	.403	-.872	-.455
Function						
FuncNar	2	4.5	3.690	.612	-.525	-.453
FuncOp	2	5	4.073	.482	-.766	4.245
FuncIm	2	4	3.044	.460	.465	.870
FuncPer	2	5	3.839	.548	-.967	1.038

As can be seen in Table 4.2, the lowest raw score was 2 and the highest score was 5. The test of normality (see Appendix K) showed that the distributions of these variables were not normally distributed. The Mardia's coefficient was 133.727, and the normalized estimate was 28.332, suggesting that the data were not multivariate normal.

There were some univariate outliers, those which were above or below 3 standard deviations (Hair et al., 1995). They included the cases with scores of 2 in ProNar, ProOp, ProIm, ProPer, SynNar, SynOp, SynPer, VocIm, and FuncNar and scores of 2.5 in VocPer, and CohPer. The Mahalanobis distance showed that nine

cases were identified as multivariate outliers (p < .001). The outliers were discarded, leading to 140 cases left.

Finally, the examination of the correlations among the 20 variables and the scatter plots revealed that the linearity assumption was not violated.

Exploratory Factor Analyses

To explore the number of factors underlying the measures, EFAs were conducted. To test the sampling adequacy and the correlations among the variables, the KMO and Bartlett's test was checked. It was found that the KMO was .914 and the Bartlett's test was significant (p < .05), indicating that it was appropriate to further the factor analysis (Hair et al., 1995).

The initial principal axis factor analysis with varimax rotation revealed four factors with eigenvalues greater than 1.0. The scree test suggested only three factors (Figure 4.1).

Figure 4.1:Scree plot

Scree Plot
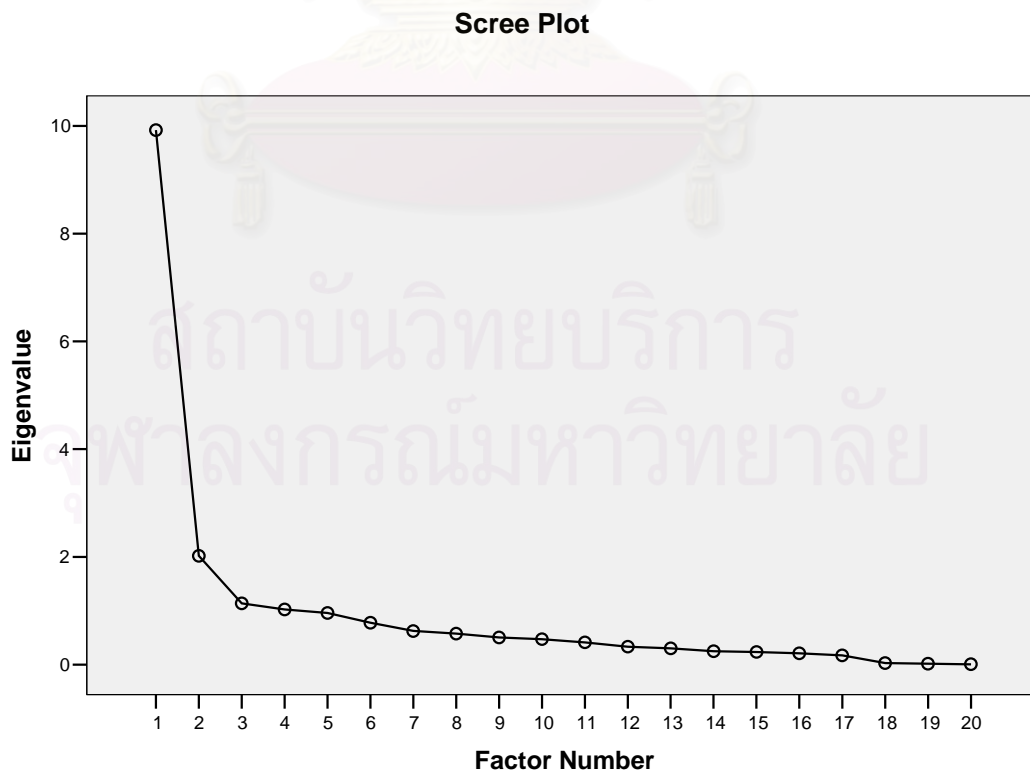
Therefore, solutions with two, three, four and five uncorrelated and correlated factors were compared. The uncorrelated solutions gave unclear interpretations and since the correlated factor solutions showed that the factors were moderately correlated (e.g. the correlations in the two factor solution was -.467), the correlated solutions were examined. In the three, four and five factor solutions, there were some cross loadings and unclear interpretation. For example, in the three factor solutions Factor 3 underlay all four syntax variables and one function variable. The two factor oblique solution (the Direct Oblimin) seemed to be the most interpretable and meaningful. As shown in the Table 4.3, the four pronunciation variables loaded highly on the second factor, which may be interpreted as the knowledge of pronunciation. The other variables had moderate to high loadings on the first factor, which may be named knowledge of form and function. The two factors correlated moderately ($r = -.467$).

Table 4.3: Factor loadings and correlation coefficient for a two-factor oblique solution

	Factor	
	1	2
ProNar	.207	**-.863**
ProOp	.170	**-.896**
ProIm	.153	**-.912**
ProPer	.174	**-.884**
SynNar	**.524**	-.221
SynOp	**.632**	-.161
SynIm	**.420**	-.166
SynPer	**.617**	-.133
VocNar	**.709**	-.101
VocOp	**.700**	-.040
VocIm	**.643**	-.160
VocPer	**.669**	-.094
CohNar	**.643**	-.079
CohOp	**.634**	-.067
CohIm	**.839**	.146
CohPer	**.764**	.123
FuncNar	**.353**	-.149
FuncOp	**.449**	-.078
FuncIm	.296	.143
FuncPer	**.384**	-.095
Factor correlation coefficient		**-.467**

Confirmatory Factor Analyses

The CFAs were next carried out. The four hypothesized models were tested using EQS 6.1 (Bentler & Wu, 2002). The first model which was based on the EFA results is presented in Figure 4.2. There are two factors which underlay the relationships among the 20 ratings. The first factor, Knowledge of Pronunciation (PRO KN), had four indicators, ProNar, ProOp, ProIm, and ProPer. The second factor, Knowledge of Form-Function (FF KN), had the rest of the variables as its indicators. The two factors were correlated.

To standardize the variances of the two factors, one of the factor loadings for each factor was fixed to one. To check if the model was identified, the number of data points and the number of parameters to be estimated were counted. With 20 variables, there were 210 data points. The number of parameters to be estimated was 41: 18 regression coefficients, 2 factor variances, 20 error variances, and 1 factor covariance. Thus, this was an overidentified model with 169 degrees of freedom.

Robust maximum likelihood estimation was used to estimate the model. Results show that the model was quite an adequate fit to the data (NNFI = .924, CFI = .933, RMSEA = .080). However, model testing continued to see if there would be a better fitting model.

Figure 4.2: Model 1
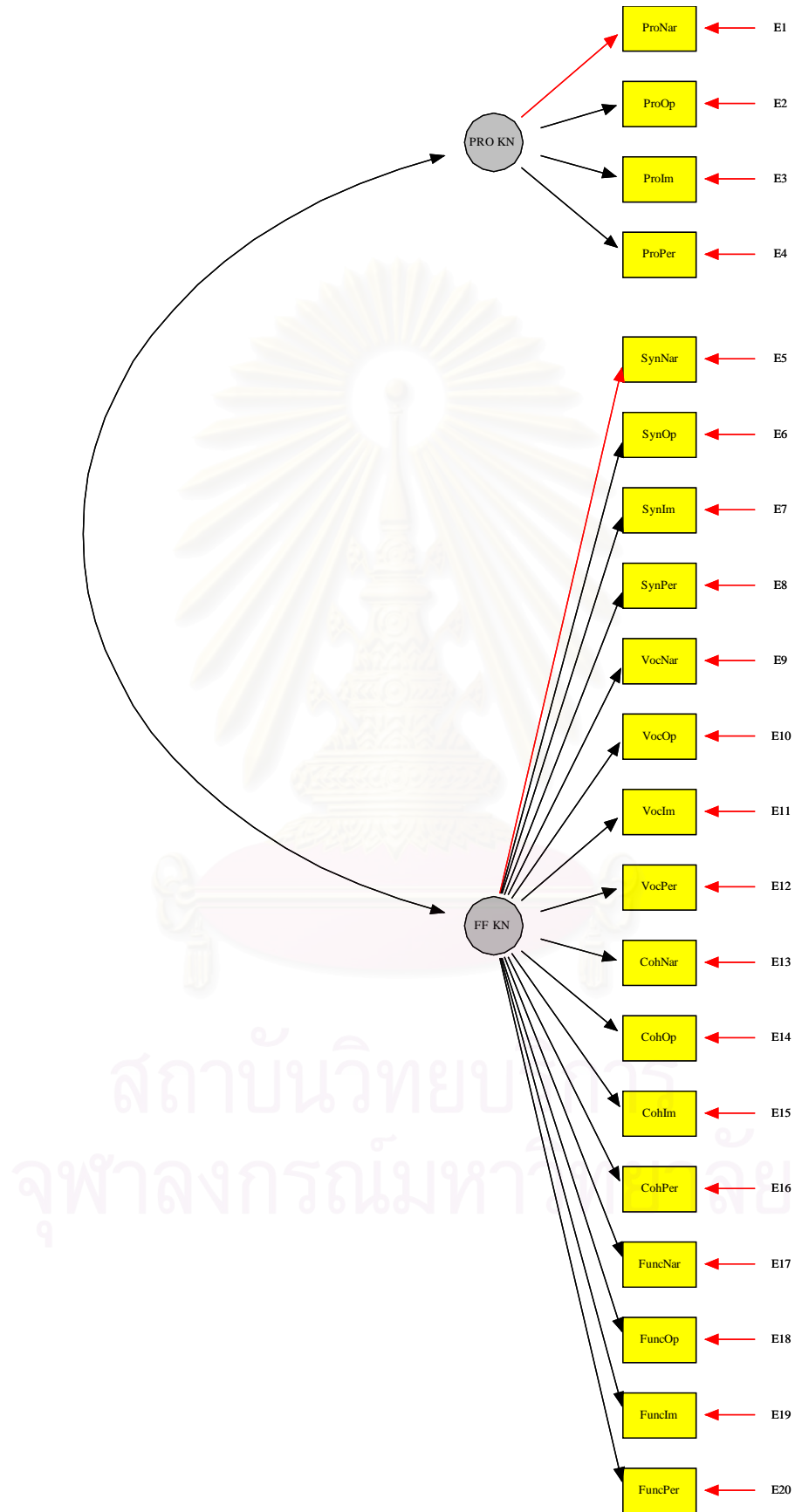
The next model was based on the design of the CBST. There were both trait and method factors (Figure 4.3). The trait factors consisted of one second-order factor, Speaking Ability or SPEAK, which accounted for the correlations among the five factors: knowledge of pronunciation (PRO KN), syntax (SYN KN), vocabulary (VOC KN), cohesion (COH KN), and function (FUNC KN). Also, there were four method factors corresponding to the four test tasks: narrative (NAR), opinion (OP), imaginary (IM), and persuasive (PER). In the first run, there were warning messages that the variance of E9 and E14 were constrained at lower bound. Thus, in the second run, these two variances were set at a low value, 0.01. The model ran successfully, with 147 degrees of freedom.

Results show that this model was a better fit model than the first one (NNFI = .956, CFI = .967, RMSEA = .027). That is, the NNFI and CFI were above .95; the RMSEA was below .05. The ratio of the chi-square to the degrees of freedom (1.102) was below 2. The Satorra-Bentler scaled chi-square was not significant (p > .05). The average absolute and off-diagonal absolute standardized residuals were quite small, and the frequency distribution of the residual covariances was symmetric and centered around zero. The largest standardized residual was .161. Taken together, these indicators showed that the model fit the data well. The Lagrange multiplier (LM) test and Wald test were examined to see if there were ways to improve the fit. Although the suggestions from the LM test were statistically significant, they did not seem to be justified substantively. For example, it suggested an addition of a path between E10 (error variance of VocOp) and E9 (error variance of VocNar). The Wald test suggested that 10 paths from the method factors should be deleted, for example, the path from ProNar on NAR, FuncOp on OP, ProPer on PER, and ProIm on IM. The results from the Wald test and the discourse analysis results (reported in the research question 3) which showed the four tasks shared some similarities indicated that the tasks may not be maximally dissimilar and may not account for the interrelationships among the variables measured by the same method. Thus, the third model which discarded the method factors was tested to see if the method factors were present.

Figure 4.3: Model 2
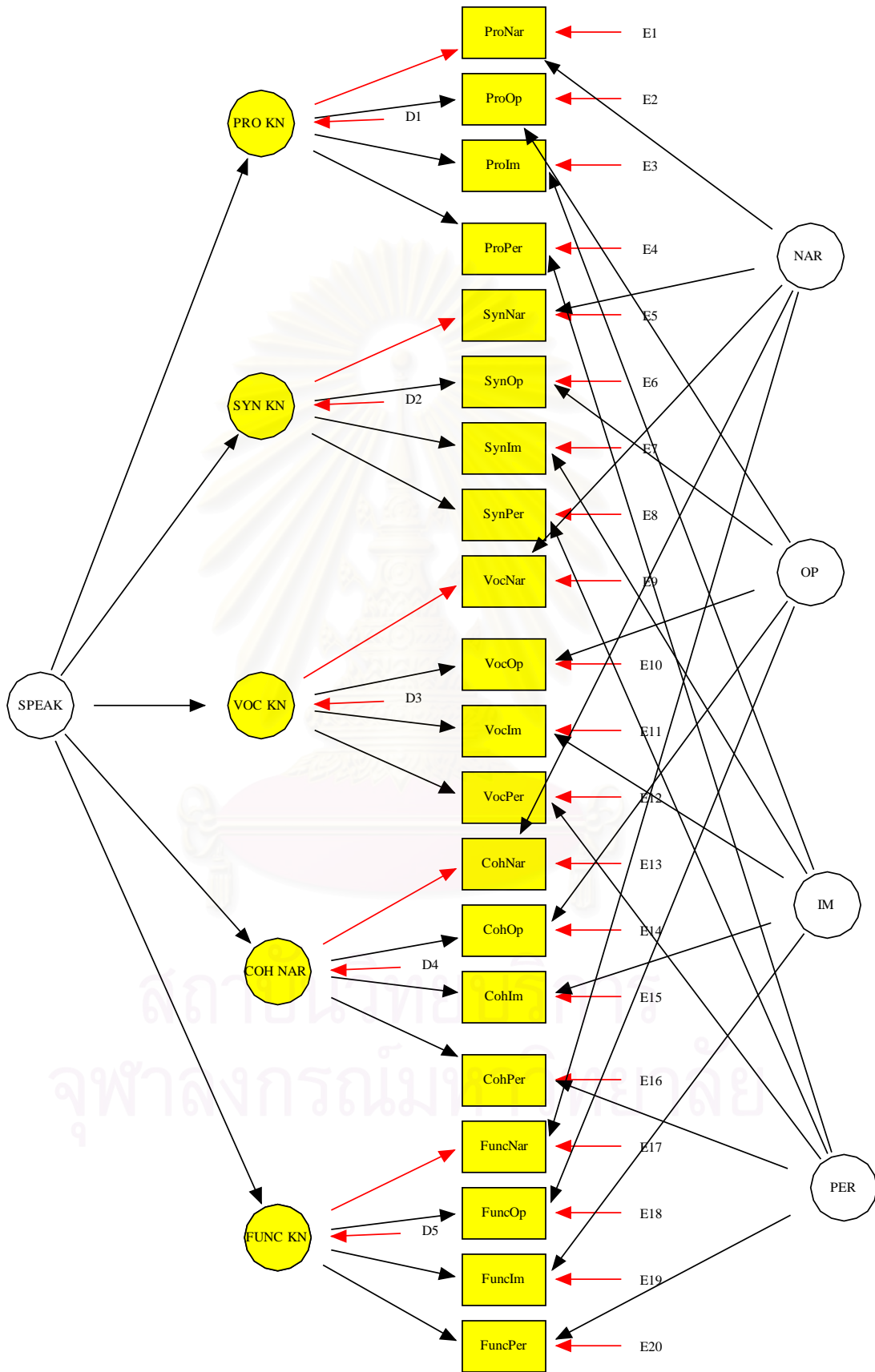
In Model 3 (Figure 4.4), there were only trait factors which included a second-order factor (SPEAK), and five first-order factors: PRO KN, SYN KN, VOC KN, COH KN, and FUNC KN. The model had 165 degrees of freedom. The results showed a very good fit (NNFI = .986, CFI = .988, RMSEA = .034). That is, the NNFI and CFI were above .95; the RMSEA was below .05 although one end of the confidence interval falling a little above .05. The ratio of the chi-square to the degrees of freedom (1.160) was below 2. The Satorra-Bentler scaled chi-square was not significant (p > .05). The average absolute and off-diagonal absolute standardized residuals were quite small, and the frequency distribution of the residual covariances was symmetric and centered around zero. The largest standardized residual was .157. Taken together, these indicators showed that the model fit the data well. None of the recommendations from the LM and Wald tests were substantive, so the model was not further revised.

Figure 4.4: Model 3
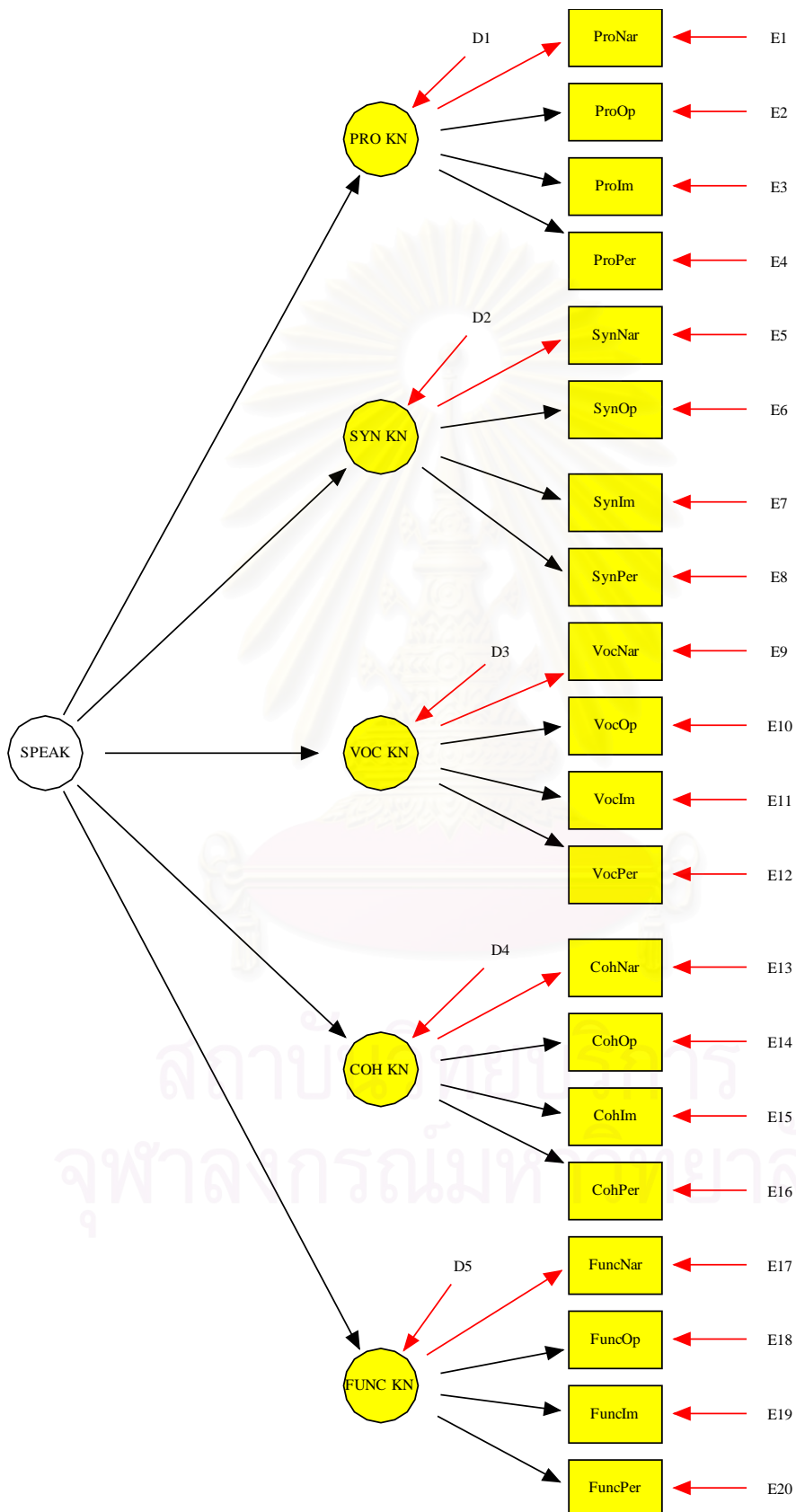
In order to test the dimension of speaking ability, the last model, a one factor model, was tested to compare its fit with that of the third model which showed the multidimensionality of the construct. In Model 4 (Figure 4.5), one factor was hypothesized to underlie the covariation among the 20 observed variables. The model had 170 degrees of freedom. The results showed that it was a poor fit to the data (NNFI = .727, CFI = .756, RMSEA = .151). Thus, the model testing finished here.

Since Models 2 and 3 are nested, the Chi-square difference test was performed to compare the two models. The $\chi 2$ difference test was 29.34 (191.4239-162.0837), df was 18 (165-147), p < .05. This indicates that Model 2 was a better fit model than Model 3.

Figure 4.5: Model 4
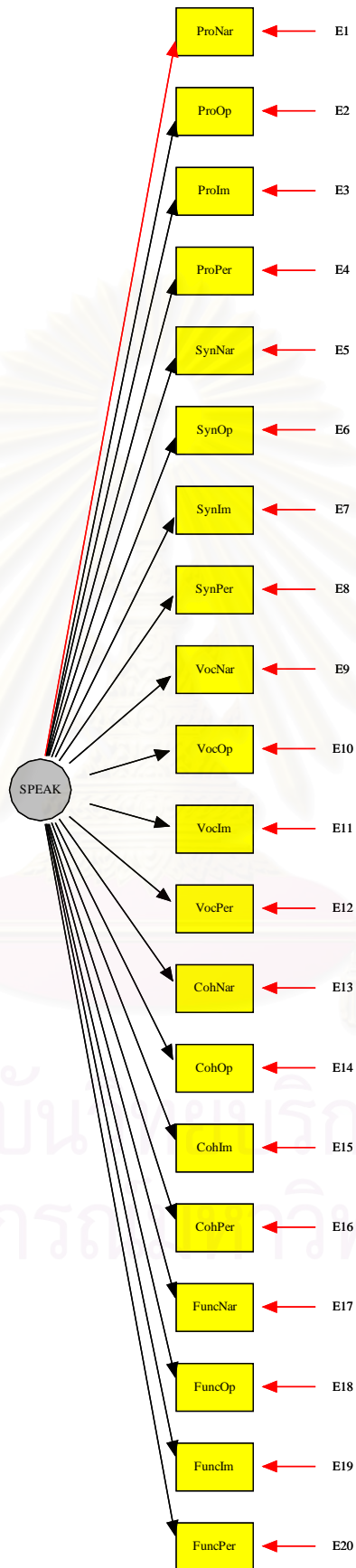
It was concluded that of the models tested, Model 2 provided the best fit to the data. This means the hypothesized relationship in the model could explain the relationship among the ratings very well. The results confirmed the hypothesis which stated that the trait factors of oral language ability for Thai first-year Chulalongkorn university students are multidimensional consisting of the factors of knowledge of pronunciation, syntax, vocabulary, cohesion and function.

The summary of the fit indices and other related information from the four models was presented below.

Table 4.4: Fit indices and other related information from the four models

	Satorra-Bentler χ^2	df	p	χ^2/df	NNFI	CFI	RMSEA	Ave. abs. st. residuals	Ave. off-diagonal abs. st. residuals
M1	317.668	169	.000	1.880	.924	.933	.080	.047	.052
M2	162.084	147	.188	1.102	.956	.967	.027	.037	.041
M3	191.424	165	.077	1.160	.986	.988	.034	.039	.043
M4	709.734	170	.000	4.175	.727	.756	.151	.120	.133

Interpretation of Model 2

The standardized parameter estimates and total proportions of variance explained by the predictors of Model 2 are presented in Table 4.5. The model structure is simple and the interpretation of the factors is straightforward. Each of the first-order trait factors had four corresponding measured variables loaded on it. That is, the knowledge of pronunciation underlay four pronunciation variables, the knowledge of syntax underlay four syntax variables, and so on. The five trait factors loaded on the second-order factor, hypothesized to be a general speaking ability. In addition, there were four method factors, each of which had five corresponding measured variables loaded on it. That is, the narrative method underlay five narrative subscores, the opinion method underlay five opinion subscores, and so on.

Table 4.5: Standardized parameter estimates and total proportions of variance explained by the predictors of Model 2

Predicted variables	Predictors										E	D	R^2
	SPEAK	PRO KN	SYN KN	VOC KN	COH KN	FUNC KN	NAR	OP	IM	PER			
ProNar		**.979**					.002				**.205**		.958
ProOp		**.992**						-.014			.127		.984
ProIm		**.996**							.037		.085		.993
ProPer		**.983**								.068	**.168**		.972
SynNar			**.796**				**-.200**				**.572**		.663
SynOp			**.836**					**.104**			**.538**		.710
SynIm			**.581**						.180		**.794**		.370
SynPer			**.809**							**-.102**	**.578**		.666
VocNar				**.836**			**.498**				**.231**		.947
VocOp				**.781**				.136			**.609**		.629
VocIm				**.758**					**.300**		**.580**		.664
VocPer				**.777**						.158	**.609**		.629
CohNar					**.727**		.124				**.676**		.543
CohOp					**.710**			**.645**			**.284**		.919
CohIm					**.810**				.121		**.574**		.670
CohPer					**.803**					.110	**.585**		.658
FuncNar						**.553**	.067				**.830**		.311
FuncOp						**.660**		.137			**.739**		.454
FuncIm						**.225**			.246		**.943**		.111
FuncPer						**.596**				.172	**.784**		.385
PRO KN	**.688**											**.726**	.473
SYN KN	**.821**											**.570**	.675
VOC KN	**.905**											**.425**	.819
COH KN	**.841**											**.541**	.707
FUNC KN	**.743**											**.669**	.553

(Bold values were significant, p < .05)

To answer the first research question, the trait factor loadings were examined in terms of the significance of the factor loadings and construct's reliability (Hair, et al., 1995). First, the trait factor loadings were all statistically significant (p < .05), meaning the variables were all significantly related to their traits as specified. That is, the pronunciation variables were very salient indicators of the knowledge of pronunciation factor (factor loadings = .979-.996). The indicators of the knowledge of syntax, vocabulary, and cohesion were also important, ranging from .581 to .836, .758 to .836, and .710 to .810, respectively. The indicators of the knowledge of function were moderate (.553 to .660), except FuncIm, which was low (.225). Among the indicators of the speaking ability, the strongest one was knowledge of vocabulary (.905), followed by knowledge of cohesion (.841), knowledge of syntax (.821), knowledge of function (.743), and knowledge of pronunciation (.688).

The construct reliability was estimated. The reliability of ProKN, SynKn, VocKn, CohKn, FuncKn and SPEAK were 0.964, 0.786, 0.830, 0.814, 0.557, and 0.845, respectively. These estimates were above .70 in most traits indicating that the indicators were sufficient in representing their traits (Hair, et al., 1995). However, the estimate of FuncKn was lower than the criterion, suggesting the indicators were not doing a good job representing this trait.

Thus, the answer to the first research question was the trait factors of oral language ability in the CBST were multidimensional, that is, it consisted of several distinct first-order trait factors: the knowledge of pronunciation, syntax, vocabulary, cohesion, and function. The relationships among these traits were accounted for by the second-order trait hypothesized to be a general speaking ability. However, the nature of the knowledge of function may be problematic since the construct reliability was low although the factor loadings were significant.

4.2 Results of research question 2

Research Question 2 asked: What processes do the examinees employ in taking the CBST?

The next section shows the test taking processes and strategies the examinees reported using while preparing for and responding to the CBST tasks. The results are organized by processes and strategies. The examinees' identification was assigned with the letter P and number from P1 to P9. The letter 'R' refers to the researcher. If

italics are presented, they correspond to the processes and strategies being discussed. The data in parentheses are information provided by the researcher to make the transcripts more understandable.

The main processes and strategies found were goal setting, assessment and planning. Other strategies used to cope with communication problems were also reported. The details are as follows:

1. Goal setting

The goal setting stage involves identifying the test tasks, choosing one or more tasks to do when given a chance, and making a decision whether to attempt to complete the task one chose. The verbal report analysis showed that none of the examinees reported using the first two processes. For the last one, only one examinee (P2) said that when she was doing the imaginary task, she did not understand the prompt "what this imaginary ideal world would be like in general", so she decided not to answer this part. Her report is presented below.

Im	R	What were you thinking at this point?
	P2	I was rereading the prompt.
	R	The first prompt?
	P2	(I was thinking) what it wanted me to do. *I didn't understand so I stopped.*
	R	Did you say anything else after that?
	P2	No.
	R	So your answer ended here?
	P2	Yes.

2. Assessment

The assessment stage included assessing the characteristics of the task, assessing one's own topical and language knowledge as well as affective schemata or feelings, and assessing the correctness or appropriateness of their responses.

2.1 Assessing the characteristics of the task

Most examinees reported they assessed the characteristics of test tasks to understand what the task requirements were. Some used the translation words provided on the screen to understand the prompt.

Example:

Op   P4:   During the preparation time which lasted 2 minutes, I read the questions to see what they were. Then, I looked at the translations. When I understood the question, I started thinking about the answer.

Another example showed that P3 saw the situation in the narrative task was talking to a close friend so she decided to use the phrase "long time no see" in her response to make it correspond to the task situation.

Nar   P3:   It said that "(imagine that) I am your close friend" so I put "long time no see" to make it like she was a friend I haven't seen for a long time.

Other examples from the imaginary and persuasive tasks revealed that some examinees paid their attention to the task requirements before performing them.

Im   P9:   (During the preparation time) I was thinking what the main issue of this task was. When I looked at the word 'world', I guessed it may be about the world in my imagination, something like that.

Per   P5:   I started to think of greeting words. I wanted to make it match the situation that I am an invited guest (to a TV show).

2.2 Assessing one's own topical and language knowledge as well as affective schemata

In this stage, speakers assessed themselves to see if topical and language knowledge related to the tasks was available for them. They also evaluated their feelings toward the tasks they were performing

Topical knowledge and affective schemata

The data showed that this process was employed when the examinees evaluated their knowledge to see if they had related world knowledge or similar experience to the task requirement. A few examinees mentioned that they were aware that they had previous experience or thoughts about the narrative and opinion tasks. In this case, the topical knowledge about these topics was already available for them. For example, P1 said that she had the most enjoyable day on campus so it was easy to think about the story.

Nar   P1:   Luckily, there was a real event like this that happened to me. So it was easy to think about the story.

P3 mentioned that she had thought about studying for a Bachelor's degree abroad before so when she was doing the opinion task, she recalled this thought.

Op   P3:   In this task I didn't have to think much because *it happened to me last year. It was about a scholarship. I was deciding whether to get a scholarship to study abroad, to study abroad or to study in Thailand* because I just got accepted to this university. So I could think of the content quickly. *The part that they asked whether it is a good idea or not. I used to think about it. I used to think about it a long time ago.*

Language knowledge and affective schemata

A few examinees evaluated their ability in speaking English when they did the narrative and imaginary tasks, and expressed their feelings to the experience in taking the CBST. One examinee (P1) felt comfortable and confident that she would be able to do a part of the test even though she didn't have enough time to prepare.

Nar   P1:   The (preparation) time was running out but this part "how you felt about it" *I felt that I would probably be able to say it* that I felt a bit sad, disappointed. But I didn't have time to note it down so I would just say it (without preparation).

Another examinee reported feeling worried when doing the imaginary task. She felt that she might not be able to do it because the vocabulary appropriate to the task may be formal and difficult to use.

Im   P6:   The first thing I thought of was the vocabulary must be difficult. I would not be able to think of the words.

She further explained that "I felt that the vocabulary must be at another level, not like those used in telling stories".

Another examinee said that she felt depressed about her English language ability after doing the narrative task.

Nar   P9:   *I felt that my English was very bad.* (I thought) Why do I feel I don't know much about English? If I don't study more, I will keep feeling like this. *I felt depressed.* I couldn't think of anything. It's like I had known these vocabulary before. *I should have been able to tell the story.*

2.3 Assessing the correctness or appropriateness of the responses

This process included monitoring and evaluating the correctness or appropriateness of one's answers to the test questions. It may occur before, during or after responding. All participants reported   that they evaluated their own responses in terms of content, grammar and pronunciation.

For example, P7 was assessing what the next content should be to make the response appropriate.

Per   P7:   When I finished speaking this part, I was deciding whether to continue with the reasons (to visit Thailand) or start talking about the example (of places to visit or activities to do).

Another example of assessing the appropriateness of the content was from P6. She stated that she used Phuket, Panga, and Koh Change as examples of tourist attractions in the South (in the persuasive task). After she gave the three examples, a thought came to her mind, "At this point I thought 'is Koh Change in the South?'"

In addition to content, the examinees monitored and evaluated their grammar and pronunciation. For example, P5 was concerned about the verb tense and form.

Im   P5:   I was worried about the verb tense, whether to use 'is' or 'was' or what.

Im   P5:   I didn't know whether I should use 'has' or 'have' with the word 'population'.

Another examinee was concerned about the structure of an utterance:

Op   P2:   In this part I thought whether 'because' and 'when' can be put next to each other. Later I felt it didn't matter because it was speaking.

Some examinees monitored their pronunciation:

Nar   P8:   I was thinking about how to say 'Saturday 14th'. I must have said it wrong earlier.

Although every examinee reported monitoring their speech, they did not do that in all tasks. For example, P4 and P7 mentioned that they did not pay attention to grammar or pronunciation in some tasks.

Nar   P4:   I didn't care about my grammar. I just wanted to give the answer, just to answer before the time was running out.

Per   P7:   In this task I didn't think much about grammar. I thought about the vocabulary instead. For the pronunciation it was automatic. I didn't pay specific attention to it.

3. Planning

The third stage, planning, included selecting elements from topical and language knowledge to be used in a plan, formulating a plan, and selecting a plan. Planning also included strategies the examinees employed when facing with difficulties in communication, for example, topic avoidance, formal avoidance, and paraphrase.

3.1 Selecting elements from topical and language knowledge to be used in a plan

The verbal reports revealed that all participants used this process when performing the tasks.

Topical knowledge

As for the topical knowledge, they decided to choose some specific elements from their previous experience or world knowledge to be used in their answers. The elements they selected were, for example, the specific story that happened to them (the narrative task), the knowledge about the world environment (the imaginary task), previous experience in giving the answer to the opinion task, and knowledge about Thai culture (the persuasive task). The excerpts below illustrated this point.

P3 selected specific experience she had earlier as her story in the narrative task:

Nar   P3:   I spent a lot of time thinking which day to talk about because there were a lot of (enjoyable) days but they were not quite remarkable. *So I chose the thank-you party for the seniors* because it was the most recent and remarkable.

P5 referred to her knowledge of the world and used it in her response to the imaginary task.

Im   P5:   After that I was thinking about the current environment: "now there is a lot of pollution. There should be some trees."

Next, P2 explained how she came up with the answer to the opinion task. She thought about her past experience and selected to answer this task the same way she did before.

Op   P2:   I was thinking that during my high school, a teacher once asked me to answer a question in English. The question was whether I would like to study abroad, something like that. My answer was "no."

Finally, P1 used her world knowledge to create the answer to the persuasive task:

Per   P1:   I was thinking about the Thai culture that was beautiful and well known.

Language knowledge

All participants in the verbal report mentioned that they selected some elements from their language knowledge to be used in a plan. These elements included knowledge of syntax, vocabulary, and cohesion. For example, some stated that they were selecting specific verb tense and modals to be used.

Examples:

Op   P7:   I was thinking about the tense I should use.

       R: In this task, which tense did you think of?

       P7:   The future and the present tense.

Im   P8:   I was thinking about the grammar, like "my ideal world would be".

       R   Which part that you were thinking about the grammar?

       P8:   The word 'would be'. I remembered that for an imaginary situation 'would be' should be used.

Some participants reported selecting specific words to be included in their responses. The examples are:

Nar   P1:   I used this phrase 'and you know' to make it like the way I talk to friends. So I added it (in my talk).

Per   P3:   I was thinking of words used in the advertisement about visiting Thailand like 'world with smile' so *I copied these words and used them in my talk.*

Some examinees stated that they deliberately selected only simple or common words in their responses.

Example:

Im   P4:   I wrote down simple words 'green' and 'clean'. I was trying not to use difficult words because I couldn't think of one.

In addition to content words, some examinees were selecting which function words to link their utterances:

Per   P5:   I was thinking of the word used after that (previous utterance). I thought if I spoke without this word, it would sound strange. I should find a word, a conjunction. The word which meant 'in addition to' or something like that.

Op   P7:   I was thinking about the conjunctions, for example, 'owing to', 'due to', 'by the way', something like that.

Two examinees noted that they took some language forms in the test prompt and used in their talk. For example, P2 said she copied 'good idea' from the prompt. P4 was thinking about the English word for 'abroad'. He felt that there may be many words which could express this meaning. He decided to use 'abroad' because it was already present in the prompt.

3.2 Formulating a plan to form a response to the task

After selecting which elements of the topical and language knowledge would be used, the examinees formulated a plan in which they made a decision about how to most successfully combine these elements to form a response. The plans they reported making involved content and language plans. For content planning, the examinees planned the message or topic that would be talked about. For language planning, they reported formulating English utterances from the message. For example, some planned the organization of the whole talk and some the word order in an utterance. Examples of making a content and language plan are presented below.

Examples of making content plans

An excerpt from P5 showed how she planned the reasons why foreigners should visit Thailand.

Per   P5:   I was thinking about the fact about things in Thailand that foreigners liked. I started writing the note because I was afraid I would forget. At first, I didn't think about Thai people (as one of the reasons). I was thinking of the nature or Thai food. After thinking a while, I thought the characteristics of Thai people would be more prominent so *I combined these two (nature and food) and make the reason about Thai people as a*

separate one.

Another examinee explained how she planned the organization of the story she wanted to talk about:

Nar   P7:   I was thinking of how to tell the story in English and about the story. (I was thinking of) the two things at the same time. *I was thinking which part of the story I should talk about first. What we did first and what next.* At the same time I was thinking of how to put it in English.

Similarly, P1 reported her thoughts about how she planned the organization of the story.

Nar   P1:   I would start with telling when the story happened.

Examples of making language plans

The language plans found in the verbal reports involved planning effective pronunciation, vocabulary, word order, and English utterances that could express the idea they had in mind. For instance, P1 planned to make her talk convincing by delivering it with appropriate rhythm (the persuasive task).

Per   P1:   (I was thinking about) the presentation style. How I should say it to make it interesting. I don't want to talk like (speaking slowly, word by word) "Thailand-is-a-". I want to make it more exciting, interesting than that.

She also made a plan about the choice of word:

Per   P1:   I was thinking that if I put 'welcome to', I mean, if I could complete my talk like what I intended and then say "so welcome to Thailand", it would sound good. It would be like a persuasion, something like that.

In another example, P4 planned how to put words into an appropriate sentence. Two excerpts illustrated this point:

Nar　P4:　I was thinking that ok I would say what happened was a football competition. The Thai language for this part was ฟุตบอลประเพณี (traditional football) but what is it in English? So I started thinking of each word. I know the word ฟุตบอล (football) but what is ประเพณี. Ok, 'tradition'. Then I put them in an order and made a sentence.

Im　P4:　I was writing the answers for the sub questions 1 and 2. I was thinking of the content in Thai. Then *I thought of how to put it in English* and wrote it down.

In addition to making the content and language plans, the examinees reported using several strategies when they faced with a communication problem. There were two options they chose: formulating a plan to solve it or avoiding the problem. The plans to solve the problems included impromptu, paraphrase and direct translation. However, some gave up; they avoided the difficult content or language forms; in other words, they employed the strategies of topic avoidance and formal avoidance, respectively.

Impromptu

There were times when some examinees did not have enough time to prepare for their talk. Therefore, they had to plan out the content and language while speaking. This strategy, impromptu, was used in all tasks. The examples are shown below.

Op　P7:　The first thing I did was writing my idea in Thai. After that I would put it in English if there was some time left. If I couldn't finish this before the (preparation) time ended, *I would think of what to say and speak at the same time.*

Paraphrase

Paraphrase includes using description, circumlocutions or exemplifications to express an idea when the speaker could not find the right words. Examples are:

Nar　P1:　I couldn't recall the word 'numb' so I said something like "I felt I can't control my body and had a stomachache", something like that. I couldn't think of the word. I just described other symptoms that I had to make it as close as possible.

In this case, P1 described other symptoms that really happened to her to express the idea that she felt numb. In another example, P6 described the word 'beggar' as 'someone who sits under the bridge'.

> Im   P6:   I wanted to say 'beggar' but I couldn't think of the word so I was thinking that beggars usually sit under the pedestrian bridge so I said 'someone who sits under the bridge' because I couldn't recall the word.

In addition to describing, the speaker may paraphrase by giving examples. An example was found in P4's excerpt. He used examples to describe the word 'scenery'.

> Per   P4:   I was thinking of words to use. But after a while, like the word 'scenery', suddenly I got stuck. So I gave examples instead. I used 'sea, mountain' instead.

Direct translation from Thai to English

There was one instance in which an examinee used direct translation to keep her continue the talk. She wanted to say 'senior' but she got stuck so she translated the Thai word รุ่นพี่ ('senior') as 'brother', which has a sense of 'someone older'.

> Nar   P2:   This word 'brother' I actually I wanted to say 'senior' but I didn't know what it was in English so I used the word 'brother' instead.

Topic avoidance

Some examinees stated that they avoided talking about a particular topic or message when they had difficulties expressing those thoughts. For example, P1 did not know how to describe a department store as she planned so she did not include the description in her talk.

> Nar   P1:   I wanted to say that the (Siam) Paragon was a grand department store but I didn't know how to say it so *I decided not to talk about it.*

In another example, P5 thought about the environmental problems in the world but she was not able to express the idea. So she skipped using it in her talk.

Im   P5:   I was thinking about the environment, the air. There were CFC, greenhouse effect but I didn't know how to say it *so I didn't mention it.*

Other examples from P7 and P8 showed they decided to avoid elaborating their idea because they got stuck.

Op   P7:   I was thinking about what I had noted down earlier. It was about (studying abroad would give you) new experience. *I wanted to elaborate this idea but I didn't know what to say* so I started the second reason. I just cut it short.

Per   P8:   After saying the phrase 'beautiful place' I was trying to elaborate it but I got stuck so I changed to talk about the first sub question. At first, I was planning to give some examples (of beautiful places) or elaborate how beautiful they were but I couldn't think of anything to say.

Formal avoidance

Some examinees mentioned that they avoided pronouncing some words and replaced them with others they thought easier to use. For example, P2 avoided telling the time '5.30 pm' because it was difficult to pronounce.

Nar   P2:   I was thinking of the time the event happened. It wasn't five o'clock. I felt lazy. Actually, I wanted to say 'five thirty' but it was difficult to say so I chose to say 'five o'clock' instead.

A similar example is from P1:

Nar   P1:   I remembered seeing word 'the Faculty of Pharmaceutical Science'. It was difficult to pronounce and I didn't know the correct pronunciation. So I said 'the pharmacist' instead.

3.3 Selecting one plan

The third stage in planning is selecting one plan. The data analysis showed that after formulating the plan as shown earlier, the examinees usually selected that plan as their response to the task. Thus, it may be seen that these two processes were not separable.

In summary, the examinees were engaged in several test-taking processes and strategies when preparing for and doing the CBST. The processes they reported included making a decision whether to attempt to complete the task; assessing the

characteristics of the test tasks; assessing one's own topical and language knowledge as well as one's feelings; assessing the correctness or appropriateness of the responses; selecting specific elements from topical and language knowledge; formulating and selecting a plan as a response to the tasks. Finally, they employed several strategies to deal with communication problems by impromptu, paraphrase, direct translation, and topic and formal avoidance.

4.3 Results of research question 3

Research Question 3 asked: To what extent do different types of the CBST tasks affect the test scores and characteristics of speech produced?

4.3.1 The effects on test scores

This part of the research question was examined by the CFA and RM ANOVA. The results from the CFA as presented earlier showed that some test tasks had an effect on the test scores. Four method factor loadings were significant, that is, the narrative loadings to syntax scores, and vocabulary scores; the opinion loadings to syntax and cohesion scores; and the imaginary loadings to vocabulary scores. This suggests that these methods had a significant effect on the respective subscores. The narrative method had a small and moderate influence on the syntax subscores (-.200) and vocabulary subscores (.498), respectively. Similarly, the opinion method had a small and moderate effect on the syntax subscores (.104) and cohesion subscores (.645), respectively. Finally, the imaginary method had a small effect on the vocabulary subscores (.300).

In addition to the CFA, RM ANOVA was conducted to answer this research question. First, the assumptions of RM ANOVA were examined. It was found that the data were not normally distributed as shown earlier. Two variables, CohNar and FuncIm, violated the assumption of the homogeneity of variance ($p < .05$; see Appendix L). However, ANOVA was robust to these two types of violation of the assumption (Glass & Hopkins, 1996). As for the assumption on the sphericity, the Mauchly's Test in the table below showed the test was significant in one dependent variables, pronunciation ($p < .05$), indicating that the sphericity assumption was violated. The epsilon was greater than .75; thus, the Huynh-Feldt correction method was used.

Table 4.6: Mauchly's Test of Sphericity

Within Subjects Effect	Scores	Mauchly's W	Approx Chi-Square	df	Sig.	Epsilon		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Scores	Pro	.699	48.243*	5	.000	.833	.869	.333
	Syn	.931	9.576	5	.088	.953	.998	.333
	Voc	.925	10.454	5	.063	.950	.993	.333
	Coh	.926	10.383	5	.065	.950	.993	.333
	Func	.953	6.538	5	.257	.971	1.000	.333

* p < .05

As can be seen in the following tables (Tables 4.7-4.11), the results of the RM ANOVAs showed no significant interaction between the task type and task order in all dependent variables. There were significant main effects for the repeated measures factor, task type, on all dependent variables with a small to large effect ($p < .05$; $\eta_p^2 =$ .036-.512), except the pronunciation subscores ($p > .05$). Finally, there were no significant effects of task order on all dependent variables ($p > .05$).

Table 4.7 shows that there were no significant main effects for the repeated measures factor, task type, on the pronunciation variables.

Table 4.7: F test of mean differences for the pronunciation scores

Pronunciation scores
Between-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Order		1.237	3	.412	.487	.692	.011
Error		115.049	136	.846			

Within-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Task type	Huynh-Feldt	.010	2.606	.004	.633	.572	.005
Task type x order		.039	7.818	.005	.796	.604	.017
Error		2.201	354.423	.006			

* p < .05

Table 4.8 shows that there were significant main effects for the repeated measures factor, task type, on the syntax variables.

Table 4.8: F test of mean differences for the syntax scores

Syntax scores
Between-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Order		1.127	3	.376	.809	.491	.018
Error		63.097	136	.464			

Within-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Task type	Sphericity Assumed	5.784	3	1.928	26.811*	.000	.165
Task type x order		1.459	9	.486	2.255	.085	.047
Error		29.339	408	.072			

* p < .05

Table 4.9 shows that there were significant main effects for the repeated measures factor, task type, on the vocabulary variables.

Table 4.9: F test of mean differences for the vocabulary scores

Vocabulary scores
Between-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Order		.709	3	.236	.424	.736	.009
Error		75.851	136	.558			

Within-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Task type	Sphericity Assumed	1.111	3	.370	5.084*	.002	.036
Task type x order		.531	9	.059	.810	.607	.018
Error		29.726	408	.073			

* p < .05

Table 4.10 shows that there were significant main effects for the repeated measures factor, task type, on the cohesion variables.

Table 4.10: F test of mean differences for the cohesion scores

Cohesion scores Between-subjects							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Order		.857	3	.286	.741	.529	.016
Error		52.402	136	.385			
Within-subjects							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Task type	Sphericity Assumed	2.585	3	.862	14.473*	.000	.096
Task type x order		.359	9	.040	.670	.736	.015
Error		24.292	408	.060			

* p < .05

Table 4.11 shows that there were significant main effects for the repeated measures factor, task type, on the function variables.

Table 4.11: F test of mean differences for the function scores

Function scores
Between-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Order		.873	3	.291	.676	.568	.015
Error		58.536	136	.430			

Within-subjects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2
Task type	Sphericity Assumed	75.771	3	25.257	142.93*	.000	.512
Task type x order		2.908	9	.323	1.829	.061	.039
Error		72.096	408	.177			

* p < .05

The line graph below shows the means of the pronunciation subscores in all tasks. The mean scores were not significantly different.

Figure 4.6: Means of pronunciation subscores in all tasks
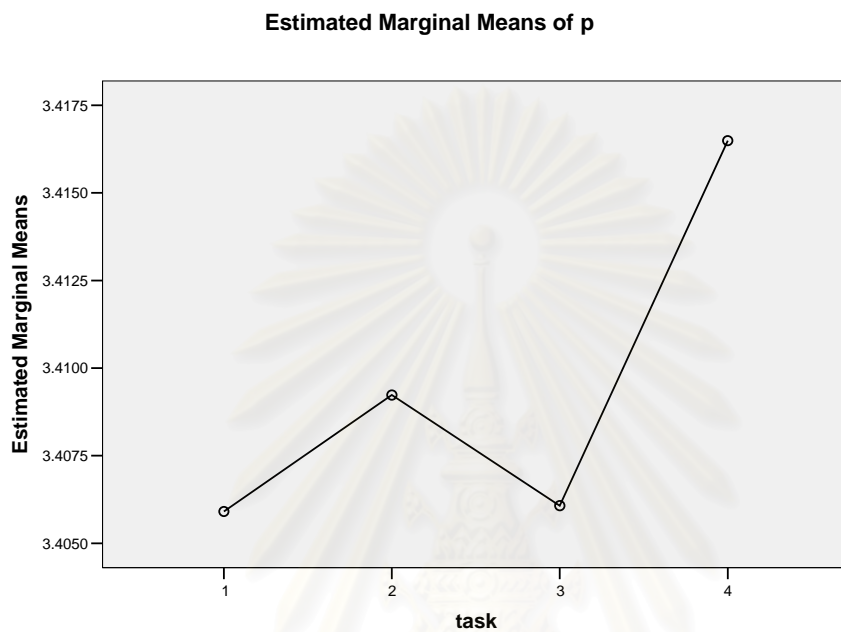
Estimated Marginal Means of p

(1 = narrative task; 2 = opinion task; 3 = imaginary task; 4 = persuasive task)

After significant differences in task types were found, multiple comparisons with the Bonferroni method were conducted to find where the significances were.

The next table (Tale 4.12) shows the syntax subscores were different from each other ($p < .05$) in 4 pairs. That is, the opinion mean was higher than the narrative and imaginary ones. The narrative and persuasive means were higher than that of the imaginary. Thus, it was clear that the opinion task was the easiest, and the imaginary task was the most difficult in terms of syntax.

Table 4.12: Pairwise Comparisons for the syntax scores (Bonferroni)

Scores	Task Type	Task Type	Mean Difference	Sig.
Syn	Nar	Op	-.106*	.002
		Im	.155*	.000
		Per	-.080	.068
	Op	Im	.261*	.000
		Per	.026	1.000
	Im	Per	-.236*	.000

* $p < .05$

The line graph below shows the means of the syntax scores in all task types.

Figure 4.7: Means of syntax subscores in all tasks
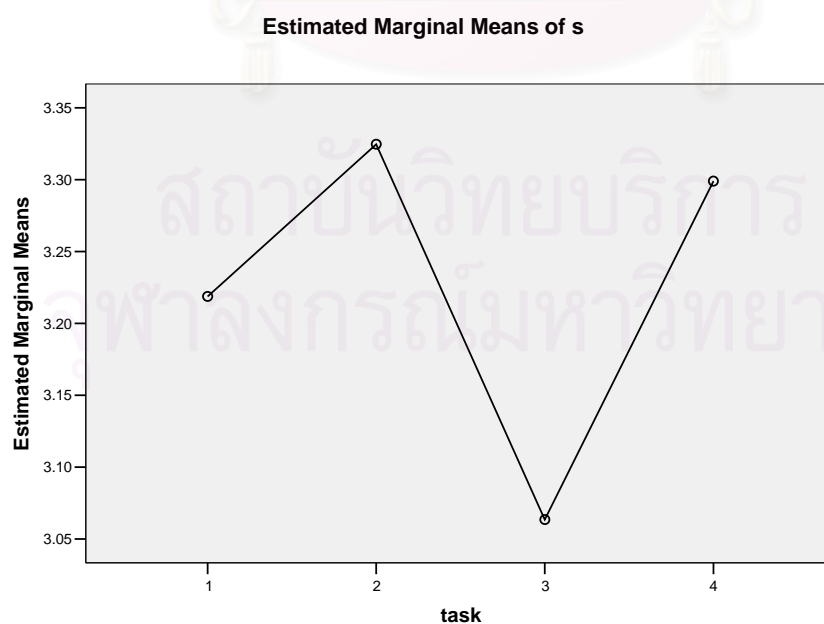
(1 = narrative task; 2 = opinion task; 3 = imaginary task; 4 = persuasive task)

Next, the mean of the vocabulary subscores were significantly different in one pair ($p < .05$) as shown below. The persuasive mean scores was higher than the imaginary one.

Table 4.13: Pairwise Comparisons for the vocabulary scores (Bonferroni)

Scores	Task Type	Task Type	Mean Difference	Sig.
Voc	Nar	Op	.018	1.000
		Im	.073	.196
		Per	-.052	.596
	Op	Im	.055	.721
		Per	-.070	.218
	Im	Per	-.125*	.001

* $p < .05$

The line graph below shows the means of the vocabulary scores in all task types.

Figure 4.8: Means of vocabulary subscores in all tasks
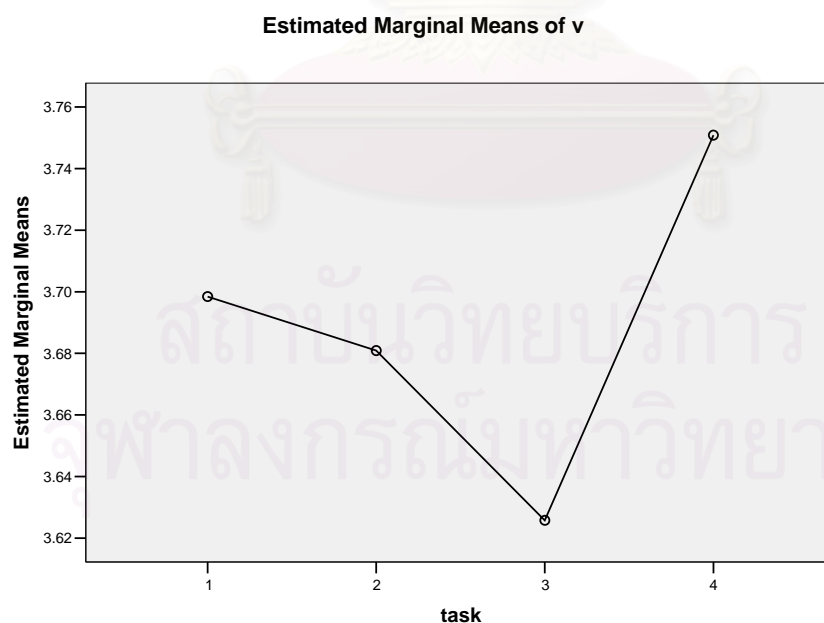
(1 = narrative task; 2 = opinion task; 3 = imaginary task; 4 = persuasive task)

The next table indicates the means of the cohesion subscores were significantly different in three pairs ($p < .05$). The narrative, the opinion and the persuasive mean scores were all higher than the imaginary score. Thus, the imaginary task was the most difficult one.

Table 4.14: Pairwise Comparisons for the cohesion scores (Bonferroni)

Scores	Task Type	Task Type	Mean Difference	Sig.
Coh	Nar	Op	.016	1.000
		Im	.174*	.000
		Per	.050	.451
	Op	Im	.158*	.000
		Per	.035	1.000
	Im	Per	-.123*	.000

* p < .05

The line graph below shows the means of the cohesion subscores in all task types.

Figure 4.9: Means of cohesion subscores in all tasks
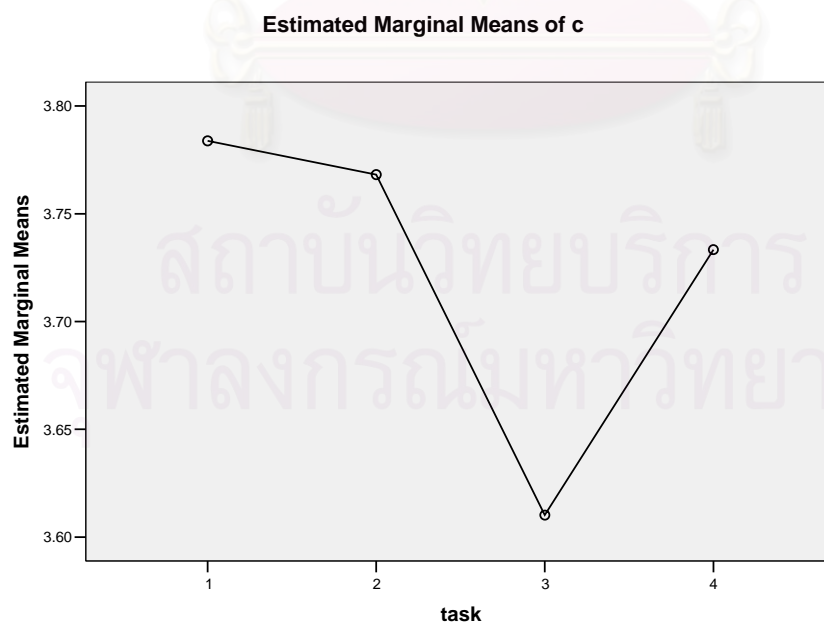
(1 = narrative task; 2 = opinion task; 3 = imaginary task; 4 = persuasive task)

As for the function subscores, the means were significantly different in five pairs ($p < .05$). The highest mean score was the opinion one, followed by the persuasive and narrative score, which were not different from each other. The lowest mean was the imaginary score. The easiest task was the opinion, followed by the persuasive and narrative; again the imaginary task was the most difficult.

Table 4.15: Pairwise Comparisons for the function scores (Bonferroni)

Scores	Task Type	Task Type	Mean Difference	Sig.
Func	Nar	Op	-.338*	.000
		Im	.655*	.000
		Per	-.111	.245
	Op	Im	.993*	.000
		Per	.227*	.000
	Im	Per	-.765*	.000

* $p < .05$

The line graph below shows the means of the function subscores in all task types.

Figure 4.10: Means of function subscores in all tasks
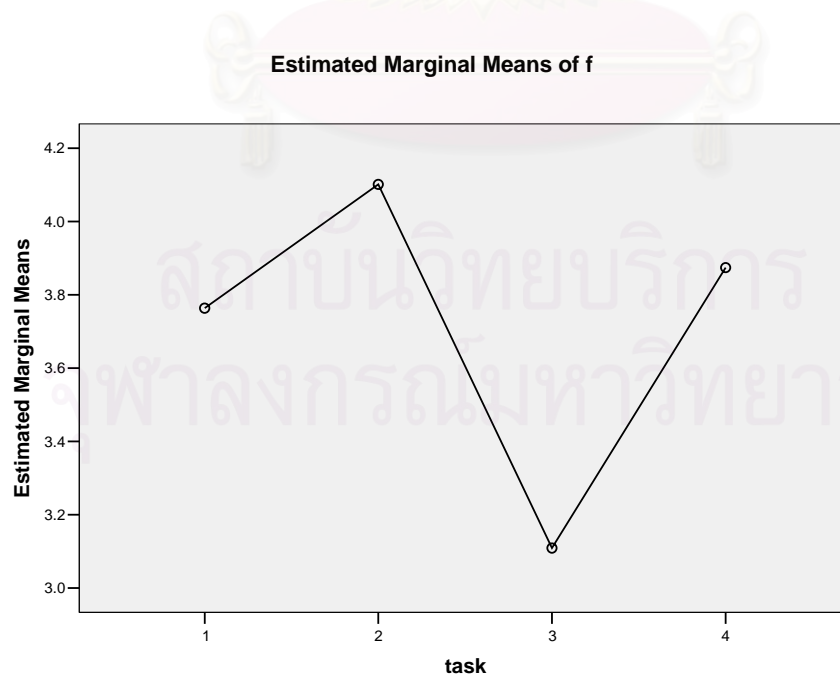
(1 = narrative task; 2 = opinion task; 3 = imaginary task; 4 = persuasive task)

The table below summarizes the results of the significant Bonferroni tests presented earlier. It can be seen that the imaginary task was the most difficult one in syntax, cohesion and function subscores; and the opinion task was the easiest in the function subscores.

Table 4.16: Summary of results of the significant Bonferroni tests

Task Type	Task Type	Significant Mean Difference			
		Syn	Voc	Coh	Func
Nar	Op	Op > Nar	-	-	Op > Nar
	Im	Nar > Im	-	Nar > Im	Nar > Im
	Per	-	-	-	-
Op	Im	Op > Im	-	Op > Im	Op > Im
	Per	-	-	-	Op > Per
Im	Per	Per > Im	Per > Im	Per > Im	Per > Im

* p < .05

As the mean differences of the syntax, vocabulary, cohesion and function subscores among the four task types were found to be significant, the differences were further investigated to see if they existed across the four task orders. Four one-way repeated measures MANOVAs were carried out for each group (see the descriptive statistics in Appendix M).

The results showed the multivariate test, Wilks' Lambda, was significant in every group ($p < .05$). The univariate tests were significant in some subscores (except the pronunciation subscores), and in some groups ($p < .05$). The Bonferroni tests for the significant univariate results were then conducted. The table below shows these differences across the four groups with different task orders.

Table 4.17: Pairwise Comparisons using the Bonferroni tests for the four groups

Scores	Task Type	Task Type	Mean Difference			
			Group 1	Group 2	Group 3	Group 4
Syn	Nar	Op	-.097	-.242*	-.115	.031
		Im	.167*	.091	.051	.313*
		Per	-.083	-.121	-.038	-.078
	Op	Im	.264*	.333*	.167	.281*
		Per	.014	.121	.077	-.109
	Im	Per	-.250*	-.212*	-.090	-.391*
Voc	Nar	Op	.042	.000	.013	.016
		Im	.056	-.015	.141	.109
		Per	-.083	-.045	.013	-.094
	Op	Im	.014	-.015	.128	.094
		Per	-.125	-.045	.000	-.109
	Im	Per	-.139	-.030	-.128	-.203
Coh	Nar	Op	-.083	.061	.038	.047
		Im	.111	.212*	.231*	.141
		Per	.000	.091	.064	.047
	Op	Im	.194*	.152	.192	.094
		Per	.083	.030	.026	.000
	Im	Per	-.111	-.121*	-.167	-.094
Func	Nar	Op	-.486*	-.242	-.295*	-.328*
		Im	.403*	.606*	.782*	.828*
		Per	-.319	-.152	.013	.016
	Op	Im	.889*	.848*	1.077*	1.156*
		Per	.167	.091	.308*	.344*
	Im	Per	-.722*	-.758*	-.769*	-.813*

* $p < .05$

As can be seen from the above table, in group 1 some of the syntax means were significantly different. The opinion, persuasive and narrative scores were not different from each other, and all were significantly higher than the imaginary score. The vocabulary mean scores were not found to be significantly different. As for the cohesion score, the opinion score was significantly higher than the imaginary one. Finally, in terms of the function scores, the opinion mean was higher than the narrative score. The imaginary one was lower than the other three. Thus, the imaginary task was the most difficult in syntax and function subscores. The opinion task was the easiest in the function subscores.

In group 2, for the syntax mean score, the opinion score was higher than the narrative and imaginary scores. Also, the persuasive mean was higher than that of the imaginary one. The vocabulary mean scores were not found to be significantly different. For the cohesion scores, both the narrative and persuasive scores were higher than the imaginary one. Finally, for the function score, the narrative, opinion and persuasive scores were not statistically different, and all were higher than the imaginary score. Thus, the imaginary task was the most difficult in the function subscores.

In group 3, the syntax and vocabulary mean scores were not significantly different among the pairs. As for the cohesion score, the narrative was higher than the imaginary score. Finally, in terms of the function scores, the opinion was the highest. The narrative and persuasive scores were the second highest, and the imaginary score was the lowest. Thus, the imaginary task was the most difficult in the function subscores.

In group 4, only the syntax and function mean scores were different. As for the syntax scores, the narrative, opinion and persuasive scores were not statistically different, and all were higher than the imaginary score. For the function scores, the opinion was the highest. The narrative and persuasive scores were the second highest, and the imaginary score was the lowest. Therefore, the imaginary task was the most difficult in the syntax and function subscores; and the opinion task the easiest in the function subscores.

The table below summarizes the significant differences between the mean scores in the four groups.

Table 4.18: Summary of results of the significant Bonferroni tests

Scores	Task Type	Task Type	Significant Mean Differences			
			Group 1	Group 2	Group 3	Group 4
Syn	Nar	Op	-	Op > Nar	-	-
		Im	Nar > Im	-	-	Nar > Im
		Per	-	-	-	-
	Op	Im	Op > Im	Op > Im	-	Op > Im
		Per	Per > Im	Per > Im	-	
	Im	Per	-	-	-	Per > Im
			-	-	-	-
Voc	Nar	Op	-	-	-	-
		Im	-	-	-	-
		Per	-	-	-	-
	Op	Im	-	-	-	-
		Per	-	-	-	-
	Im	Per	-	-	-	-
			-	-	-	-
Coh	Nar	Op	-	-	-	-
		Im	-	Nar > Im	Nar > Im	-
		Per	-	-	-	-
	Op	Im	Op > Im	-	-	-
		Per	-	-	-	-
	Im	Per	-	Per > Im	-	-
			-	-	-	-
Func	Nar	Op	Op > Nar	-	Op > Nar	Op > Nar
		Im	Nar > Im	Nar > Im	Nar > Im	Nar > Im
		Per	-	-	-	-
	Op	Im	Op > Im	Op > Im	Op > Im	Op > Im
		Per	-	-	Op > Per	Op > Per
	Im	Per	Per > Im	Per > Im	Per > Im	Per > Im

* $p < .05$

To summarize, the results of the effects of test tasks on subscores indicate that the tasks had significant effects on all scores except the pronunciation one. The results partly confirmed the hypothesis which stated that different types of test tasks in the CBST had a significant effect on the test scores. Although the results from all examinees and those from each group tested separately were different in some aspects, it can be seen that in general the opinion task seemed to be the easiest one in several types of subscores, while the imaginary task seemed to be the most difficult.

4.3.2 The effects of test tasks on the speech produced

The results from genre, speech function and grammatical features analyses are presented below.

4.3.2.1 Genre analysis

The results showed that the test tasks could be classified as different genres: the narrative task was a recount, the imaginary a description, and the opinion and persuasive tasks an exposition. The details and examples of the examinees' responses are presented below.

The narrative task

The narrative task consisted of the following stages: (abstract), (orientation), and record of events with evaluative comments occurring throughout the text. The excerpts below are examples of the narrative genre. They were separated into different stages with the codes put next to them. The examinees were identified by the numbers.

Examinee 110

Stage		Response
		oh I miss you so much I haven't seen you for a long time uh (1)
abstract		the most enjoyable day on my campus is uh freshy day and freshy night
orientation time	*to*	about (1) uh three months ago
orientation situation	*to*	it's it's a party and activity for the freshy
orientation place	*to*	um and it's um (2) was happen at my faculty of Law near Samyan you know
record events	*of*	uh at first we have to uh wear uniform of Chulalongkorn University and we have (1) have have been told to be many thing by the senior or the % ว้ากเกอร์ % like uh sing sing many songs in a very loud noise and (1)
[evaluative comments]		I was very tired I feel very very bored with this activity

orientation time	*to*	later on
record events	*of*	I I met friend playing many games
[evaluative comments]		which is very fun and (3) very interesting this was a was a very happy time
orientation time	*to*	and at about six o' clock
record events	*of*	we have a party and have a concert
[evaluative comments]		which is very fun and I was very impress for this activity I feel so glad to be

Examinee 110 started her story with an abstract which summarized the story to come. It stated what the story would be about, the most enjoyable day on campus. After the abstract, the examinee set the temporal, situational, and physical context and presented record of events. The evaluation about her feelings toward the story was inserted in many places as well as at the end of the story.

Another example from participant 59 shared the same stages although her story did not happen on campus. It was about camping with a friend. The first stage was an orientation to time, followed by an abstract, more orientation and record of events. There were an evaluation and some more orientation as she described the place, and record of events. The story ended with another evaluative comment about her feelings to the experience.

Examinee 59

Stage		Response
orientation time	*to*	last month
abstract		I I went to the camping with my friends and
orientation place	*to*	at Saraburi uh (2)
orientation situation	*to*	we we went to Saraburi by bus (1)

orientation time	*to*	before we before we before we arrive (1) the resort
record events	*of*	um my my friend close my eyes (1) and take me to resort
[evaluative comments]		that resort is very exciting
orientation place	*to*	there are there're there're on the mountain and (1) around the resort there there waterfall there have a waterfall
record events	*of*	we have many activity together just like uh we are walk we are walk ral.. rally
orientation place	*to*	on the mountain
Record events	*of*	and go to play go to play in waterfalls (1) go to the temple
orientation place and time	*to*	that on the mountain and in in that night
Record events	*of*	we are talking together
orientation time	*to*	until until morning (1)
[evaluative comments]		and I think this trip is very great in my life I have ev.. never go anywhere with

The imaginary task

The stages found in the imaginary task were description and (conclusion). The excerpt from examinee 53 showed the description of the characteristics of the ideal world, while that from examinee 110 consisted of both description and conclusion.

Examinee 53

Stage	Response
description	in the future I think our world will have many environment and lacking of natural source (2) then we should solve this problem (1) uh I want to live in (1) clean and full of natural source world (1) um people and people live together with natural our world will be like that every… everyone help to (2) to or do something not (1) destroy our environment

Examinee 110

Stage	Response
	good morning everyone today I would like to present my idea about an ideal world (3)
description	my ideal world's um the world with peace and happiness there's would be no war no crime everyone smile to each other and everyone befriend um no one took advantage from (2) other person and (2) besides I I think there shouldn't be a gap between the poor and the rich they should share everything because they are friend (1) what is more disaster is very (1) ver... (3) disaster like tsunami earthquake or storm is very bad thing I hope that there would be no disaster in this world then we can live in a (1) peaceful world which my ideal world and (4) I (1)
(conclusion)	if we live in a peaceful world I hope that everyone (?) to be happy and

The opinion and persuasive tasks

Both the opinion and persuasive tasks fell under the same genre, an exposition. However, they differed in that the function of the opinion task was to persuade the audience that what is being argued for, studying for a Bachelor's degree abroad is a good or bad idea, is well formulated. The aim of the persuasive task, on the other hand, was to persuade the audience to do as recommended, visiting Thailand. The stages in both tasks were thesis, argument and (conclusion). The first two stages were present in the opinion task but in some cases the thesis was not stated explicitly in the persuasive task. However, it could be implied from the test situation what the speaker was arguing for. The examples of the responses to the two tasks are presented below.

The Opinion task: Examinee 18

Stage	Response
thesis	I think it's is a good idea to study for a Bachelor's degree abroad (1)
argument	first I think it's makes learner know more thing to help to use in learner's life (1) um second I think it's can help learner to use more effective foreign language (2) to use in learner's life

Examinee 110

Stage	Response
thesis	well I think studying studying for a bachelor's degree abroad is very good idea for me (1) even it's costs a lot of money and may lack a time in uh with friend in the university in Thailand (1)
argument	first of all study studying abroad uh you will make you have a chance to use (1) um language like English which is uh international language and (1) it's uh access to source of knowledge you can um develop yourself (1) in many skill like speaking (1) reading (1) writing (1) and (2) what is more is when you studying abroad you you will be on your own it's inpedent independent so you have to be very responsible and you have to adapt yourself for the new social with many kind of person you you will be how to live on your own and (1)
conclusion	I think it's very (1) it going to be an experience for (1) real life than uh studying in our country

The excerpts from the opinion task showed that the thesis was stated clearly and followed by an argument with a few reasons. These stages were found in most data. However, some variation can be seen in Examinee 48's response. Her thesis was there were both good and bad points about studying for a Bachelor's degree abroad. The argument that followed consisted of two parts, a pro and a con. The pro was reasons why it was a good idea while the con was why it was bad. Both belonged to the same stage, an argument.

Examinee 48

Stage	Response
Thesis	about studying abroad in Bachelor degrees (1) I think there are good things and bad things too (1)
Argument	
pro	like (1) the good things (1) are the (1) if you go to the using English country of course you will get the great English back for sure but if you go to the (1) country which use different language of course you will get a new language and that will help you to find a job easier than the people with only two language (3) another things is the new experience (1) you will have to learn how to adapt yourself in a new place (1) and (2) learn about their culture

con	but the bad things for studying a bachelor degrees abroad is the (1) you will be far from home sometime you will be miss your family so much and miss your old friends (2) and if you are in the abroad (1) at the early age you had not many friends in Thailand and the h… pri…

In the persuasive task the example from examinee 110 clearly included the thesis followed by an argument, while that of examinee 18 was mainly reasons supporting the idea why foreigners should visit Thailand.

Examinee 110

Stage	Response
thesis	if you are thinking about where to spend your vacation Thailand would be your answer (1)
argument	Thailand's very (1) interesting (1) first (1) there are many tourist attractions in Thailand for example Wat Pra Kaew (2) which is ve… very very beautiful and gorgeous (2) floating market in Ratchaburi (1) JJ market which is perfect place to find some um souvenir for your friends your family (2) um Siam Square where's very fashionable clothe and many thing you can find here (3) and (2) what's more (1) Thailand is very well known for perfect food which very delicious tasty for example Tom Yum Kung papaya salad (2) and there many thing to do like um (3) and Thailand's (2) uh you can travel here in affordable price it's very important thing

The persuasive task: Examinee 18

Stage	Response
arguement	Thailand has many good place to visit (1) and almost people are kindly (6) when you visit Thailand you can see a beautiful sea at the south of Thailand (1) in Phuket or Krabi province (2) it has many manys many many beautiful beach (2) and (2) if you want to see (2) beautiful (1) beautiful (3) beautiful (2) beautiful forest you can go to the north (2) the north of Thailand to see the elephant (5) the elephant and (2) beautiful forest

To summarize, three genres were found in the four test tasks: a recount, description and exposition. Each genre had different internal structures. The structure of a recount, which was the genre of the narrative task, consisted of (abstract), (orientation), and record of events with evaluative comments; that of the imaginary task was description and (conclusion); and that of the opinion and persuasive task was thesis, argument and (conclusion). Some variation was observed in the last genre. The

thesis may be optional in the persuasive task and the argument in the opinion task may consist of one or two subparts, a pro and a con.

4.3.2.2 Speech function analysis

The analysis of the speech functions in the CBST tasks showed some similarities and differences. Some speech functions were found across tasks: expressing opinion, feelings or wish, and elaborating. Some were task specific. That is, greeting and stating the purpose of the talk were found in all tasks except the opinion task. The functions that were commonly identified in the narrative task were setting the temporal and physical context, describing the situation, and narrating. Those found in the imaginary task were predicting, describing the characteristics of the ideal world, justifying and suggesting. Finally, those generally present in the opinion and persuasive tasks were expressing opinion, and justifying. The excerpts below illustrated these features.

The narrative task

The responses from examinee 53, 110 and 92 revealed the speech functions that frequently occurred in the narrative task: describing the temporal and physical context, narrating, and expressing feelings. Some variation also existed such as expressing opinion and justifying (Examinee 53), and expressing wish (Examinee 92). The similarities in speech functions in the three stories were observed although they differed in topic: the first meeting with friends and seniors (Examinee 53), Freshy Day and Night (Examinee 110), and a group trip (Examinee 92).

The examples of the narrative task were:

Examinee 53

Speech Functions	Response
Setting temporal, physical context	story is on first day of my first meeting friends in this university (1)

narrating	senior are located former year to many group (1) first senior divided all new (1) all former year in twelve group (1) a… after that they advise many things many rule in this faculty such as (1) uh curriculum (3)
Describing the situation	they advise us all day
narrating	and in the evening (1) they (1) get me to my group (1) and I partici… participated with my new friends and (1) my senior (3)
Expressing opinion	I (1) I think it's a good thing to have senior
Justifying	because (3) it's make me (1) uh unity (3) it's it's make everyone are unity

Examinee 110

Speech Functions	Response
Greeting	oh I miss you so much I haven't seen you for a long time uh (1)
Setting temporal context	the most enjoyable day on my campus is uh freshy day and freshy night about (1) uh three months ago
Describing the situation	it's it's a party and activity for the freshy
setting physical context	um and it's um (2) was happen at my faculty of Law near Samyan you know
narrating	uh at first we have to uh wear uniform of Chulalongkorn University and we have (1) have have been told to be many thing by the senior or the % ว้ากเกอร์ % like uh sing sing many songs in a very loud noise and (1)
expressing feelings	I was very tired I feel very very bored with this activity
narrating	later on I I met friend playing many games which is very fun and (3) very interesting
expressing feelings	this was a was a very happy time
Narrating	and at about six o' clock we have a party and have a concert which is very fun
expressing feelings	and I was very impress for this activity I feel so glad to be

Examinee 92

Speech Functions	Response
Setting the temporal, physical context	Last week uh I had a group trip at Prachuebkirikhan um yeah
Narrating	the first day we play sport for example uh beach volleyball football yeah and in the afternoon we play we play game and yeah in the evening we had a little party um and the second day we did the group activities yes (1) just like um (2) play game together again and (1) takes photo uh sorry took photos and in the afternoon we went back to Bangkok
Expressing feelings	uh I was very (1) uh I was very tired but I really impress and enjoy this trip I think (1) I think this trip is is is really nice
Expressing wish	and and it would be it would be more fun if you had join our trip

Finally, it can be seen that some participants also greeted the audience (her friend in this situation) before telling the story, and asked the audience a question at the end. For example, examinee 48 began her talk with "hello my friends we haven't seen each other for a long time how have you been I have been great (1)", and ended it with "how about you have you had anything great happen to you recently".

The imaginary task

The speech function that was typical in the imaginary task was describing the characteristics of the ideal world. Other functions found were predicting, justifying, elaborating and concluding. In addition, some participants greeted the audience and stated the purpose of their talk clearly. For example, Examinee 110 started her talk saying, "good morning everyone today I would like to present my idea about an ideal world". Another function frequently found was 'suggesting ways to make this world a better place', which was not expected to be performed from the task requirement. Examples were provided to show the speech functions in the task.

Examinee 42

Speech Functions	Response
describing	my imaginary ideal worlds (2) would be like (2) people live together with peaceful and (3) help to reserve the nature (4) um (2) I the ideal world (1) um have two characteristic is (1) peace and everyone love each other (3)
suggesting	and the way in which this world would be a better place for everyone is (1) everyone help each other to s… to (3) to save the world by reserve the nature and (4) love each other (4)
concluding	that's will make (2) and (1) that will make the world (3) be a better place (1) for everyone

Examinee 53

Speech Functions	Response
Predicting	in the future I think our world will have many environment and lacking of natural source (2)
Suggesting	then we should solve this problem (1)
describing	uh I want to live in (1) clean and full of natural source world (1) um people and people live together with natural our world will be like that every… everyone help to (2) to or do something not (1) destroy our environment

Examinee 48

Speech Functions	Response
Describing	my imaginary world will be like there no poverty and wars
Justifying	because you know poverty is um (1) very bad thing
Elaborating	at the moment people suffer from poverty so much and some people die from the poverty (2)
Describing	the next thing is the war (1)
Elaborating	as you can see from the world war one and world war two

	people die and die and died and some poverty also cause from the war (3)
Justifying	of course if we live with wars our world won't be any happy (2)
Suggesting	and the way in which the world would be a better place for everyone is everyone should be like consider more about (1) the other people's feelings (2)
Elaborating	so that we can help each other (1) and the world will be (1) a better place for sure (3)
suggesting	and sometimes we should (1) not be selfish (1) we should give to the others more than we do right now

The opinion and persuasive tasks

Some speech functions were performed in both tasks: expressing opinion and justifying. Some were found in the persuasive task only: greeting, stating the purpose of the talk, and emphasizing one's point. The excerpts below illustrated these features.

The opinion task: Examinee 42

Speech Functions	Response
Expressing opinion	I think studying for Bac… for a Bachelor degree abroad is a good idea
Justifying	because you have (?) experience and meet a lot of kind of people (1) and you will make a lot of friends (1) uh (4) and the second reasons is (2) when you studying for Bachelor degree abroads (2) it's very good way for you to study English by yourself because (1) you must speak English every time in your (1) daily life
elaborating	that's will make you develop all your (2) listening speaking reading skill

Examinee 42 started her talk with an explicit opinion to the idea. A few reasons to support her idea followed, and she ended with an elaboration of the last reason. In Examinee 48's response, similar functions were present: expressing opinion

justifying, and elaborating. However, the latter provided more elaboration of each reason.

The opinion task: Examinee 48

Speech Functions	Response
Expressing opinion	about studying abroad in Bachelor degrees (1) I think there are good things and bad things too (1)
Justifying	like (1) the good things (1) are the (1) if you go to the using English country of course you will get the great English back for sure but if you go to the (1) country which use different language of course you will get a new language
Elaborating	and that will help you to find a job easier than the people with only two language (3)
Justifying	another things is the new experience (1)
Elaborating	you will have to learn how to adapt yourself in a new place (1) and (2) learn about their culture
Justifying	but the bad things for studying a bachelor degrees abroad is the (1) you will be far from home
Elaborating	sometime you will be miss your family so much and miss your old friends (2)
Justifying	and if you are in the abroad (1) at the early age you had not many friends in Thailand and the h… pri…

As for the persuasive task, one examinee greeted and stated the purpose of the talk before persuading the audience to visit the country: "hello everyone we are now going to (1) uh introduce you about Thailand" (Examinee 48).

Another example of the persuasive task was from Examinee 110. She stated clearly her suggestion for the audience and provided justification to her suggestion as well as some elaboration. She ended the talk by emphasizing her point.

The persuasive task: Examinee 110

Speech Functions	Response
Suggesting	if you are thinking about where to spend your vacation Thailand would be your answer (1)
justifying	Thailand's very (1) interesting (1)
Elaborating	first (1) there are many tourist attractions in Thailand for example Wat Pra Kaew (2) which is ve… very very beautiful and gorgeous (2) floating market in Ratchaburi (1) JJ market which is perfect place to find some um souvenir for your friends your family (2) um Siam Square where's very fashionable clothe and many thing you can find here (3)
Justifying	and (2) what's more (1) Thailand is very well known for perfect food which very delicious tasty for example Tom Yum Kung papaya salad (2) and there many thing to do like um (3) and Thailand's (2) uh you can travel here in affordable price
Emphasizing one's point	it's very important thing

Similarly, Examinee 42 started with her suggestion and reasons.

The persuasive task: Examinee 42

Speech Functions	Response
Suggesting and justifying	the reason why they why you should come to Thailand is because Thailand is an interesting places (1) and um such as Pattaya and Koh Samui (1) that has white sand and beaches for you to come to see it Thailand has a lot of interesting culture in each part of Thailand (2) and the second is (1) Thai people is very kind and friendly (1) to the foreigner (3)
Elaborating	the example of what you can do or see in Thailand is you can go to an interesting places and take photos (2) or when you come to when you go to (1) s… to Pattaya or Koh Samui you can (3) you can (4) swimming in the sea and (4) and see the coral reef (2)
Expressing opinion	that's very beautiful (3)
Elaborating	or you can eat some Thai food

In the last example of the persuasive task, Examinee 53 began her response with the reason why foreigners should visit Thailand; there was no explicit suggestion. The rest of her talk was an elaboration of interesting places in Thailand.

The persuasive task: Examinee 53

Speech Functions	Response
justifying	there are many interest in interested in interested places in Thailand (1)
elaborating	uh in the east there are there is a Chonburi province there they it has (1) a beautiful beach and many entertainment to do at the night such as um (2) night market uh in the south there are many beautiful beach too and they have (1) uh (2) beautif… and they have a long beach (1) and (3) have a (4) and they has many (2) h… many hotel or in Bangkok there are many entertaining entertainment things to do such as department store or JJ market because (2) they they buy a many cheap things

To conclude, some speech functions were common in many tasks such as greeting, stating purpose of that talk, and elaborating. Some were found to be task specific such as narrating, expressing wish and predicting.

4.3.2.3 Analysis of grammatical features

The features investigated here were verb tense, modal and conjunctions. Typical errors were also noted. The data analysis showed that certain grammatical features were associated with certain tasks. In addition, some conjunctions were common in all tasks: additive (adding further information, e.g. 'and), reason (e.g. 'because'), and example (e.g. 'like'). The conjunction 'and' was used with different purposes but only the uses related to the characteristics of the tasks were concerned here. That is, the uses of 'and' to give further information (additive), and to indicate that an event took place after another (temporal) were noted. The details are presented task by task.

The narrative task

The grammatical elements typically found in the narrative task were verbs and modals in the past form and temporal and additive conjunctions. Temporal conjunctions were such as 'after', 'then', and 'before'. The frequently used additive conjunction was 'and'. Some errors in using present tense rather than the past form were observed (* indicates errors). The examples below illustrated these points.

In the example from Examinee 48, the verb tense used in the response below was the present perfect and the incorrect use of present tense instead of the past tense. No modals were used here. Finally, the conjunctions were varied: temporal (first, and then, then, until, after), additive (too, also), contrast (but), and reason (because).

Examinee 48

Response	Tense (* only)	Modal	Conjunction
hello my friends			
we haven't seen each other for a long time			
how have you been			
I have been great (1)			
and recently we have the shareball competition last week at the faculty (2)	Have*		
I am in the shareball team too (1)	Am*		too
you know first we practice and practice (1)	Practice*		First
and then (1) we went to the competition (2)			And then
then we won the first round			then
we were so glad			
and then we went through the competition			And then
we won and won and won			
and that's (1) a great feelings			
until the final round and we get a bit stred stress	Get*		Until
because (1) another team is also a tough team too	Is*		Because Also Too
but in the end we won the gold medal (2)			but
and I (1) felt very glad that we won the gold medal (2)			
and I'm glad that I play the shareball too	I'm* Play*		too
and I also (1) got some new friends (1)			And also
and after we won we went to celebrate (1)			after
and that a great feelings (3)			

how about you			
have you had anything great happen to you recently			

* errors

Similar grammatical features were seen in another example below. There was also a conjunction which showed examples (such as) of a part of the event being told.

Examinee 53

Response	Tense (* only)	Modal	Conjunction
story is on first day of my first meeting friends in this university (1)	Is*		
senior are located former year to many group (1)	Are located*		
first senior divided all new (1) all former year in twelve group (1)			First
a… after that they advise many things many rule in this faculty such as (1) uh curriculum (3)	Advise*		After that Such as
they advise us all day	Advise*		
and in the evening (1) they (1) get me to my group (1)	Get*		And
and I partici… participated with my new friends and (1) my senior (3)			*And* And
I (1) I think it's a good thing to have senior			
because (3) it's make me (1) uh unity (3)	it's make*		because
it's it's make everyone are unity	it's it's make*		

* errors

The imaginary task

The verb form and modals in the imaginary task were the past form such as 'would' and verbs in the past simple tense showing that the situation was hypothetical or imaginary. Other modals which showed prediction and possibility were also present. Many errors were observed in tense and modal. As for the conjunctions, the most frequent one was an additive; others included those of condition, reason, result, and example.

In the response from Examinee 48, several tense and modal errors were found, for example, the present tense of 'die' was used instead of the past tense, and 'will' was used instead of 'would' in the first utterance. In addition to 'would' indicating an unreal situation, other modals were also applied in the task. That is, they included

'can' (possibility), 'won't', 'will' (prediction), and 'should' (weak obligation). The conjunctions varied: 'and', 'also' (additive), 'because' (reason), 'if' (condition), and 'so that' (purpose).

Examinee 48

Response	Tense (* only)	Modal	Conjunction
My imaginary world will be like there no poverty and wars		Will*	And
because you know poverty is um (1) very bad thing			because
at the moment people suffer from poverty so much			
and some people die from the poverty (2)			And
the next thing is the war (1)			
as you can see from the world war one *and* world war two people die and die and died	Die*	can	*And*
and some poverty also cause from the war (3)	Cause*		also
of course if we live with wars our world won't be any happy (2)		Won't	if
and the way in which the world would be a better place for everyone is everyone should be like consider more about (1) the other people's feelings (2)	Is*	Would should	
so that we can help each other (1)		can	So that
and the world will be (1) a better place for sure (3)		will	
and sometimes we should (1) not be selfish (1)		should	And
we should give to the others more than we do right now		should	

* errors

Additional conjunctions were 'or' (alternative), 'like' (example), and 'then' (result). They can be seen from the following excerpt.

Examinee 110

Response	Tense (* only)	Modal	Conjunction
good morning everyone			
today I would like to present my idea about an ideal world (3)			
my ideal world's um the world with peace and happiness	world's*		And
there's would be no war no crime everyone	there's*	would	And

smile to each other and everyone befriend um no one took advantage from (2) other person	smile* befriend*		
and (2) besides I I think there shouldn't be a gap between the poor and the rich		Shouldn't	And besides
they should share everything		should	
because they are friend (1)			because
what is more disaster is very (1) very (3) disaster like tsunami earthquake or storm is very bad thing			what is more* like or
I hope that there would be no disaster in this world			
then we can live in a (1) peaceful world which my ideal world		can	Then
and (4) I (1) if we live in a peaceful world I hope that everyone (?) to be happy and	to be*		And if

* errors

The opinion and persuasive tasks

Both tasks shared several similarities in the types of verb tense, modal and conjunctions. For example, the tense was typically the present form, and modals were that of possibility ('can').

The responses to the opinion task below showed modals of possibility ('can), and prediction ('will'), conjunctions used to introduce the points ('first' and 'second'), reason ('because'), and condition ('when').

The opinion task: Examinee 18

Response	Tense (* only)	Modal	Conjunction
I think it's is a good idea to study for a Bachelor's degree abroad (1)			
first I think it's makes learner know more thing to help to use in learner's life (1)	it's makes*		First
um second I think it's can help learner to use more effective foreign language (2) to use in learner's life	It's*	Can	Second

* errors

The opinion task: Examinee 42

Response	Tense (* only)	Modal	Conjunction
I think studying for Bac… for a Bachelor degree abroad is a good idea			
because you have (?) experience and meet a lot of kind of people (1)	Have* Meet*		Because And
and you will make a lot of friends (1)		Will	And
uh (4) and the second reasons is (2) when you studying for Bachelor degree abroads (2) it's very good way for you to study English by yourself			And When
because (1) you must speak English every time in your (1) daily life		Must*	because
that's will make you develop all your (2) listening speaking reading skill		will	

* errors

As for the persuasive task, the responses revealed the use of present tense form, modals of possibility ('would' and 'can'), conjunctions of condition ('if'), introduction of points to be discussed ('first'), example ('for example', 'such as'), addition ('and', 'what's more', 'too'), and alteration ('or').

The persuasive task: Examinee 110

Response	Tense (* only)	Modal	Conjunction
if you are thinking about where to spend your vacation Thailand would be your answer (1)		Would	if
Thailand's very (1) interesting (1)			
first (1) there are many tourist attractions in Thailand for example Wat Pra Kaew (2) which is ve very very beautiful and gorgeous (2) floating market in Ratchaburi (1) JJ market which is perfect place to find some um souvenir for your friends your family (2) um Siam Square where's very fashionable clothe *and* many thing you can find here (3)		can	First For example And *And*
and (2) what's more (1) Thailand is very well known for perfect food which's very delicious tasty for example Tom Yum Kung papaya salad (2)			And What's more For example
and there many thing to do like um (2)			And
and Thailand's (2) uh you can travel here in		can	And

| affordable price | | | |
| it's very important thing | | | |

* errors

The persuasive task: Examinee 53

Response	Tense (* only)	Modal	Conjunction
there are many interest in interested in interested places in Thailand (1)			
uh in the east there are there is a Chonburi province			
there they it has (1) a beautiful beach and many entertainment to do at the night such as um (2) night market			And Such as
uh in the south there are many beautiful beach too			Too
and they have (1) uh (2) beautif…			And
and they have a long beach (1) and (3) have a (4)			And
and they has many (2) h… many hotel			And
or in Bangkok there are many entertaining entertainment things to do such as department store or JJ market			Or Such as Or
because (2) they they buy a many cheap things			Because*

* errors

To conclude, some similarities and differences were present across tasks in terms of the tense form, modal and conjunctions. The features generally present in the narrative task were, for example, past tense and modal, and temporal conjunctions. Those in the imaginary task were tense and modal in an unreal form, and conditional conjunctions. Those in the opinion and persuasive tasks were, for instance, present tense, modal of possibility and conjunctions of condition and example.

4.4 Results of research question 4

4.4.1 Views of examinees in the lower intermediate, intermediate, and advanced groups towards the CBST obtained from the Likert-scaled items

The assumption of homogeneity of variance-covariance matrices was checked. The Box's *M* test was found to be significant ($p (M) < .05$) as the table below shows. This indicates that the assumption was violated.

Table 4.19: Results of Box's *M* test for homogeneity of variance-covariance matrices

Box's M	806.789
F	1.125
df1	506
df2	30202.152
Sig.	.027

However, the outcome of Box's *M* test can be disregarded since the robustness of the Wilks' Lambda test can be expected when the sample sizes are equal (Tabachnick and Fidell, 2001). The multivariate test (Table 4.20) was not significant (Wilks' Lambda = .621; $F_{(44, 174)}$ = 1.065, $p > .05$; η_p^2 = .212). This means that there was no significant effect of the proficiency levels on all dependent variables.

Table 4.20: Results of multivariate test

Effect		Value	*F*	Hypothesis df	Error df	Sig.	η_p^2
Intercept	Wilks' Lambda	.006	696.962	22	87	.000	.994
Group	Wilks' Lambda	.621	1.065	44	174	.377	.212

The means and standard deviation of the scores of all items were presented in the table below. The lowest mean of the total score was 1.88 from item 17, and the highest mean was 3.65 from item 3.

Table 4.21: Mean level of view towards the CBST by all proficiency groups

Question items	LI (N = 37)		INT (N = 37)		ADV (N = 37)		Total (N = 111)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. I feel I had the opportunity to adequately demonstrate my speaking ability on the computer-based speaking test (CBST).	2.78	.71	2.65	.68	2.84	.73	2.76	.70
2. I felt comfortable taking the speaking test via computer.	2.65	.68	2.49	.77	2.78	.59	2.64	.69
3. I did not understand the instructions on how to take the test.	3.51	.61	3.68	.53	3.76	.43	3.64	.53
4. I think the warm-up part was useful.	3.46	.56	3.51	.65	3.49	.56	3.49	.57
5. I think the CBST would give a rater an accurate idea of my speaking ability.	3.08	.83	2.86	.71	2.62	.86	2.86	.82
6. The computer is a useful instrument to test speaking ability.	3.22	.53	3.00	.67	3.05	.71	3.09	.64
7. The instructions on how to take the CBST are easy to follow.	3.49	.51	3.62	.49	3.57	.50	3.56	.50
8. The procedures for taking the test were too complicated.	3.19	.46	3.49	.56	3.40	.55	3.36	.54
9. A speaking test should not be delivered by computer.	2.70	.74	2.40	.69	2.59	.83	2.57	.76

Question items	LI (N = 37)		INT (N = 37)		ADV (N = 37)		Total (N = 111)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
10. Taking the oral language test delivered by computer was an unpleasant experience.	3.02	.64	2.81	.46	2.97	.83	2.94	.66
11. The narration task was too difficult.	2.56	.55	2.48	.77	3.02	.50	2.69	.66
12. The opinion task was too difficult.	2.73	.51	2.78	.53	3.00	.62	2.84	.56
13. The imagination task was too difficult.	2.24	.72	2.27	.80	2.46	.80	2.32	.78
14. The persuasion task was too difficult.	2.78	.63	2.75	.55	2.86	.59	2.80	.58
15. The preparation time allowed was too short for the narrative task.	2.10	.81	1.95	.70	2.38	.82	2.14	.80
16. The preparation time allowed was too short for the opinion task.	2.22	.71	2.05	.66	2.41	.83	2.23	.75
17. The preparation time allowed was too short for the imaginary task.	1.86	.67	1.76	.60	2.03	.73	1.88	.67
18. The preparation time allowed was too short for the persuasive task.	2.00	.75	1.95	.62	2.14	.82	2.03	.73
19. The response time allowed for was too short for the narrative task.	2.78	.67	2.76	.60	2.70	.66	2.74	.64

Question items	LI (N = 37)		INT (N = 37)		ADV (N = 37)		Total (N = 111)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
20. The response time allowed for was too short for the opinion task.	2.89	.57	2.81	.62	2.68	.63	2.79	.60
21. The response time allowed for was too short for the imaginary task.	2.78	.67	2.78	.67	2.67	.62	2.74	.65
22. The response time allowed for was too short for the persuasive task.	2.73	.65	2.70	.70	2.56	.72	2.67	.69

The results showed that all three proficiency groups were not significantly different in their views towards the CBST. The interpretations of their views are as follows. First, the examinees in all groups seemed to have a positive view to the CBST as a measure of their speaking ability. They felt that they had a chance to demonstrate their speaking ability in this test (questionnaire item 1) and that the CBST would give a rater an accurate view of their speaking ability (item 5). They had positive attitudes to the usefulness (item 6) and the appropriateness of using the computer to deliver a speaking test (item 9). All groups also felt comfortable (item 2) and had a pleasant experience (item 10) while taking the test. In terms of the test instruction and warm-up (items 3, 4, 7, 8), the examinees found the instructions easy to understand, the procedures not too complicated, and the warm-up useful. They felt that the narrative, opinion and persuasive tasks were not too difficult but the imaginary task was (items 11-14). Finally, they felt that the preparation time was too short for all tasks, but the response time was not too short for all tasks.

To summarize, the views of the three proficiency groups were not significantly different in all questionnaire items. This confirmed the hypothesis that there are no significant differences in terms of attitudes towards the CBST among examinees with different levels of speaking ability. In general, they had positive views towards the

CBST. However, there were some negative opinions in terms of the difficulty of the imaginary task and the adequacy of the preparation time for all four tasks.

4.4.2 Views of examinees in the lower intermediate, intermediate, and advanced groups towards the CBST obtained from the open-ended questions

For the first open-ended question, there were one examinee and 38 examinees who selected 'strongly disagreed' and 'disagreed', respectively, to the statement "I feel I had the opportunity to adequately demonstrate my speaking ability on the computer-based speaking test". As can be seen from Table 4.22, the examinees' reasons were categorized into two main parts: the reasons about using the computer to test English speaking ability, and the features of the CBST. The most frequent reason to this statement was about the inappropriateness of the preparation time. Most of the examinees who gave the reasons to this question (41 %) felt that the preparation time was too short. The next frequently found reasons were they felt nervous when taking the test (25.6 %), and they thought the testing situation was unauthentic (20.5 %) because they had to speak to the computer, not to a real person. Other reasons were given by a few examinees such as the inappropriateness of the test topics and test rooms.

Table 4.22: Answers to question 1

Reasons (from 39 examinees)	Frequency	Percent
1. Using the computer to test English speaking ability		
- feeling nervous when using the computer	10	25.6%
- feeling shy	2	5.1%
- unauthentic test situation	8	20.5%
- unfamiliar with the use of computer-based testing	1	2.6%
- unable to redo the answer	1	2.6%
- test room should have given more privacy	3	7.7%
2. The CBST		
2.1 Test tasks		
- more word translation should have been given	1	2.6%
2.2 test topics		
- unfamiliar test topics	4	10.3%
2.3 Preparation and Response time		
- preparation time was too short	16	41%
- limitation of time causing nervousness	2	5.1%
- response time was too short	1	2.6%
- response time was too long	1	2.6%
- there should have been no preparation time	1	2.6%
- seeing the timer causing stress	1	2.6%

There were 106 examinees who answered the next two questions. First, the answers to the question what the examinees liked about the CBST can be categorized into similar topics: feelings towards using the computer to test speaking ability, and the features of the CBST in terms of the test instructions, tutorials, test tasks and program design. From the answers given (Table 4.23), the most frequent one was they felt less stress (29.7%) and more confident (27.9%) than taking the test with a person. The next reasons were they liked the design of the computer program (20.7%). Many felt the program convenient to use. Many also liked the tutorials of how to take the CBST (10.8%), and test prompt (9.9%). Some felt they were able to concentrate more when using the computer (9.0%) and they believed the test was able to assess their speaking ability (8.1%). Other good points about the CBST were, for example, some examinees felt free to express their opinion (6.3%), the preparation and response time was appropriate (6.3%) and the test was convenient and time saving (5.4%).

Table 4.23: Advantages of the CBST

Advantages of the CBST (from 106 examinees)	Frequency	Percent
1. Using the computer to test English speaking ability		
- feeling confident	31	27.9
- less stress	33	29.7
- able to concentrate more	10	9.0
- trendy and interesting	4	3.6
- time saving and convenient	6	5.4
- felt free to express one's opinion	7	6.3
- privacy	5	4.5
- having an opportunity to show one's speaking ability	2	1.8
- testing environment was the same for everyone	2	1.8
- corresponding to the use of the computer in the classroom	1	0.9
2. The CBST		
2.1 Test instructions		
- easy to understand	9	8.1
2.2 Tutorials	12	10.8
- useful	3	2.7
- easy to follow and flexible	8	7.2
- giving us a chance to ask questions about the test	1	0.9
2.3 Test tasks		
2.3.1 test prompt		
- easy to understand	3	2.7
- not too difficult	2	1.8
- specifically stated	2	1.8
- liked the translations given	2	1.8
- the number of questions was appropriate	1	0.9
- could be applied in real life	1	0.9
2.3.2 able to assess one's speaking ability	9	8.1
2.3.3 preparation and response time were appropriate	7	6.3
2.4 program design		
- convenient, easy to use, clear	23	20.7

What the examinees disliked about the CBST can be categorized into similar groups: using the computer to test speaking ability, and the features of the CBST in terms of the test instructions, tutorials, test tasks and program design, and others (Table 4.24). The majority of the examinees answering this question felt the preparation time was too short (44.1%). Many viewed using the computer to test speaking ability was unauthentic because there was no interaction with another speaker (34.2%). Some felt that the response time was too short (9.9%), and some felt stressed being timed (9.0%). Other features that the examinees disliked about the test were, for example, feeling nervous when using the computer (4.5%); the tutorials

were too long (4.5%); the user should have more control navigating (4.5%); there was no privacy in the test room (3.6%); and the test could not assess one's speaking ability (3.6%).

Table 4. 24: Disadvantages of the CBST

Disadvantages of the CBST	Frequency	Percent
1. Using the computer to test English speaking ability		
- feeling nervous	5	4.5
- speaking to the computer was an unauthentic test situation		
	38	34.2
- no privacy in the test room	4	3.6
2. The CBST		
2.1 Test instructions		
- the Thai version of the instructions should have been deleted	1	0.9
- the audio input should have been deleted if the Thai instructions were provided	1	0.9
2.2 Tutorials		
- too long	5	4.5
- should have been explained by a person	1	0.9
2.3 Test tasks		
2.3.1 Test prompt		
- some were too difficult (e.g. the imaginary task)	2	1.8
- some were not clear	2	1.8
- some situations were not appropriate	2	1.8
- some were too broad	1	0.9
- some were too easy	1	0.9
- more translation should have been provided	1	0.9
- too many questions	1	0.9
2.3.2 Unable to assess one's speaking ability	4	3.6
2.3.3 Preparation and response time		
- preparation time was too short	49	44.1
- timing causing stress	10	9.0
- response time was too short	11	9.9
- response time was too long	3	2.7
- preparation and response time were not set according to the difficulty of task	3	2.7
- there should have been no preparation time allowed	1	0.9
- the speakers should have allocated the time themselves	1	0.9
- more time should have been allowed before the preparation started	1	0.9
2.3.4 Program design		
- the user should have had more control to move from one screen to another	5	4.5
- too complicated	1	0.9
3. Others		
- feeling uncomfortable when wearing the headset	1	0.9
- not familiar with the audio input	1	0.9
- sample answers should have been provided	1	0.9

To summarize, the findings from the open-ended questions gave additional information about the examinees' views to the CBST. They felt that there were both advantages and disadvantages of the CBST. Many felt less stress and more confident using the computer as a speaking test, and they found the test program convenient and easy to use. The major disadvantages were the preparation time was too short and some examinees viewed the test as unauthentic because there was no interaction with others.

In conclusion, this chapter presents the findings from all four research questions. The next chapter will present the discussion of the results as well as the conclusion.
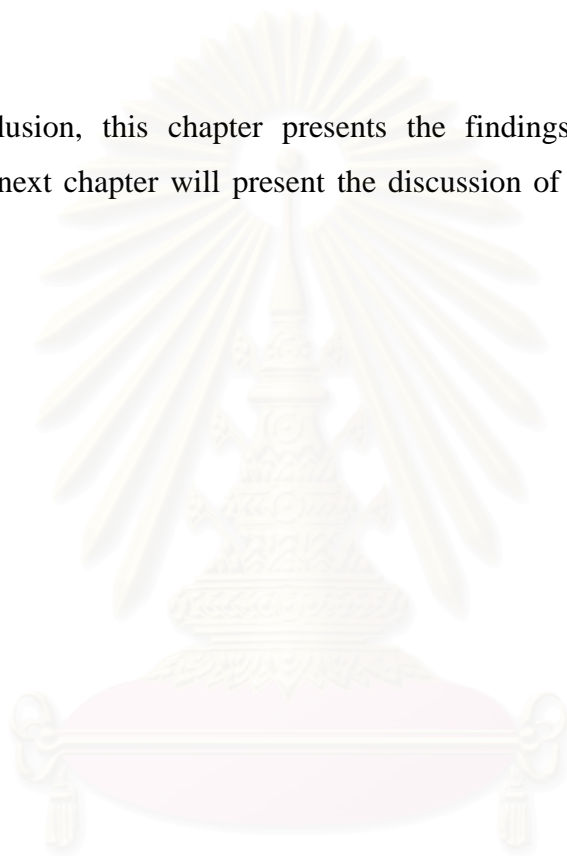
CHAPTER V
DISCUSSIONS AND CONCLUSIONS

This chapter presents the summary of the study, discussions on the findings, conclusions, implications and recommendations for further studies and test use.

5.1 Summary of the study

The study aimed to investigate the trait structure of oral language ability in a computer-based speaking test (CBST), the test-taking processes and strategies of the examinees taking the test, the effects of test tasks on test scores and speech produced, and the attitudes of the examinees to the test. The participants were 167 Thai first-year university students. The research instruments included the CBST, rating scales, and questionnaire. The traits of interest were knowledge of pronunciation, syntax, vocabulary, cohesion and function. The test tasks comprised the narrative, opinion, imaginary and persuasive tasks which generated planned monologic responses. The data were analyzed through both quantitative and qualitative approaches. The quantitative approaches were multivariate generalizability theory, confirmatory factor analysis, and MANOVAs. The qualitative approaches consisted of content analysis of verbal protocols, and attitudes to the test, and discourse analysis of the test responses: analyses of genre, speech functions and grammatical features. The limitations of the study concerned the homogeneous nature of the participants in terms of their native language and educational level, the use of purposive sampling method to select the participants as well as the relatively small sample size.

The research findings were as follows. First, the trait structure of oral language ability in the CBST was multidimensional, consisting of the knowledge of pronunciation, syntax, vocabulary, cohesion and function. However, the last trait may be problematic since it had low construct reliability. In addition, all of these five traits were explained by the general speaking ability. The second result was examinees employed several types of test-taking processes: goal setting, assessment and planning; and communication strategies when they faced with difficulties in performing the tasks. Third, the effects of test methods on test scores and speech produced were present. The CFA results showed that the narrative had an influence on syntax and vocabulary subscores; the opinion had an impact on syntax and cohesion

subscores; and the imaginary on vocabulary subscores. The RM ANOVAs revealed that the imaginary task seemed to be the most difficult one in all subscores, except the pronunciation scores. The discourse analysis also found some similarities and differences among the test task responses in genre, speech functions and grammatical elements. Finally, the examinees at different speaking proficiency levels were not different in their views towards the CBST. They generally had positive views towards the test but they also had negative attitudes, for example, in terms of test preparation time and difficulty of the imaginary task.

The findings seemed to provide evidence which supported and challenged the validity of the CBST score interpretations. Several kinds of supportive evidence may allow us to draw an inference about the examinees' speaking ability from their test scores. First, the CFA results showed most of the trait factor loadings were significant ($p < .05$) and ranged from moderate to high and most construct reliability was above the criterion of 0.70. This suggests that most measures were good indicators of the construct they were designed to measure, indicating that the CBST was a valid measure of most of the speaking ability as defined. Second, the examinees engaged in test-taking processes and strategies relevant to the constructs of interest. This implies that the CBST allowed the examinees to make use of their language knowledge, topical knowledge, strategic competence and communication strategies when performing the test tasks. Thus, their test scores may be interpreted as a product of the use of the ability that the CBST aimed to measure. In addition, the genre, speech functions and grammatical features found in the test responses also corresponded largely with the task requirements. Finally, the examinees generally had positive attitudes to the test. However, evidence that may threaten the validity of the score interpretation came from the low reliability of the functional knowledge construct. This may be caused by the characteristics of the imaginary task. The task was seen more demanding and problematic to many examinees. The difficulty and ambiguity of the task prompt has led to the lowest mean score in almost all areas, unintended speech functions and negative attitude to the test. Moreover, the lack of interaction with another interlocutor caused some examinees to view the test unauthentic although many agreed that the test could measure their speaking ability. Despite these weak points, with some revisions and careful interpretations made from the test

scores, the CBST may serve as a potentially useful instrument in oral language assessment.

5.2 Discussions

The following section presents the discussion of the findings. It should be noted that it is important to take into account the sample size, the homogeneity of the sample in terms of their first language and educational level, the number of factors investigated, and the characteristics of the CBST in the interpretation of the results. Also, since the purposive sampling method was used in the study, the results may not be generalized to the population.

5.2.1 The trait structure of oral language ability in a Computer-based Speaking Test

The first research question asked "What is the trait structure of oral language ability in a Computer-based Speaking Test?" To answer this research question, only the trait structure in the CFA results was examined. The findings showed the trait structure consisted of a general speaking ability underlying first-order factors of knowledge of pronunciation, syntax, vocabulary, cohesion, and function. Each of the first-order factors related to the ratings of the corresponding knowledge from all four test tasks. The trait factor loadings were all significant and ranged from very high to moderate, with an exception of one relatively weak loading from the function scores in the imaginary task on the knowledge of function. The higher order factor accounted for the correlations among the five factors to a great extent. In addition, the construct reliability of all traits was high, except the knowledge of function. This suggests that the trait structure of oral language ability in the CBST was multidimensional, consisting of a general speaking ability underlying first-order factors of knowledge of pronunciation, syntax, vocabulary, cohesion, and function. However, the nature of the knowledge of function may not be clear because of its low construct reliability although it had significant loadings to the respective variables and on the general speaking ability factor. The low reliability of the construct indicates that the indicators of this trait have some measurement errors associated with them (Hair et al,. 1995). The sources of the measurement errors could range from inaccurate responses to problematic definition of the construct (Hair et al., 1995). That is, the variables may

not represent the construct or the examinees may not understand how to answer or may interpret the tasks differently from the intended purposes. As the factor loadings from most variables associated with the construct were moderate (0.553-0.660) except that to the FuncIm (0.225), it may be speculated that this variable may cause the low reliability of the functional knowledge. The data from the discourse analysis may provide some explanation to the problematic functional scores in the imaginary task. A speech function that was not a part of the test prompt, making suggestions, was performed by several examinees. This shows one source of measurement errors in that the examinees interpreted the task requirement in a different way than what was intended, introducing low reliability of the variable.

The trait structure found in the study is consistent with previous research (Bachman & Palmer, 1982; Carroll, 1983; Fouly et al., 1990). That is, oral language ability is multidimensional, rather than unidimensional. The factors of knowledge that were investigated are distinct but related to each other to some extent and are significant indicators of a general speaking ability. Finally, the high coefficients from the speaking ability factor to the five traits indicate high validity of this trait. Thus, the findings in this part of the study provide some evidence to support the theory of multidimensionality of language ability.

5.2.2 The test taking processes and strategies used in the CBST

The verbal reports analyses show that the examinees employed various processes and strategies when doing the CBST. They went through a process of goal setting, assessment and planning in which their topical and language knowledge were involved. Several strategies were used to solve problems in communication such as paraphrase and topic avoidance. These findings correspond to the strategic competence proposed by Bachman and Palmer (1996) and previous research on communication strategies (for example Færch & Kasper, 1983). As defined by Bachman and Palmer (1996), the strategic competence is executive processes which operate in language use in a non-test and test setting as well as other cognitive activities. The communication strategies as defined by Færch and Kasper (1983) have come into play when the speaker has difficulties in reaching a particular communicative goal. Similarly, they can be utilized in any language use situation. Thus, the presence of these processes and strategies in the CBST may suggest that the

tasks could create a situation in which the examinees are allowed to demonstrate their oral language proficiency. As the verbal reports revealed, the examinees had a particular communicative goal in mind when performing the tasks. They employed both their language, topical and strategic ability to accomplish their goals.

However, there were some observations about the strategic competence the examinees reported using. That is, not all components were found, especially in the goal setting stage, and formulating and selecting a plan were inseparable. The reason why all areas in the goal setting process did not appear in the verbal report may be due to the characteristics of the task prompts (Bachman & Palmer, 1996). The test tasks did not allow the examinees to choose the task to perform, and the tasks were clearly specified as four different items. Thus, the examinees need not to choose which task to do or identify them. The reason why formulating and selecting a plan cannot be separated may be because of the time limitation. Since the examinees had limited preparation time, they may not be able to create several plans to choose from. They may have to decide to make only one plan to complete the tasks in time.

In addition, the test-taking processes and strategies employed seem to be associated with the constructs the CBST aimed to measure. That is, the CBST was designed to elicit language performance in which we can draw inferences about the speaker's five areas of language knowledge: knowledge of pronunciation, syntax, vocabulary, cohesion and function. The verbal report analyses showed that the examinees did make use of these types of knowledge as expected. Their test scores may allow us to make inferences about their speaking ability in the five areas. Thus, the results seem to yield evidence that supports the construct validity of the CBST interpretation.

5.2.3 The effects of test tasks on test scores and speech produced

The presence of the effects of the test tasks on the test scores and characteristics of the responses may be explained as follows. First, language use or in this case oral language test is influenced largely by the context in which it occurs (Hymes, 1972). This context comprises several components, fore example, the setting, topic and participants involved in the situation. The way context of the situation may shape language performance may be described through the use of strategic competence (Bachman & Palmer, 1996). That is, language users use the strategic

competence to understand and produce speech appropriate to the situation at hand by setting a communicative goal; assessing the task characteristics, their own knowledge, and correctness and appropriateness of the responses; and planning how to perform the task effectively. Thus, one component that may influence language use or test performance may be the situation or test task characteristics (Bachman & Palmer, 1996; Hymes, 1972).

The specified task requirements and the interpretation of the examinees about what and how they should perform may illustrate the effects of the test method on test performance. The narrative task required the examinees to tell a story to a friend; the subtopics provided were what, when, where it happened, how it ended and what their feelings were. The genre of the expected responses was then a recount containing an abstract (a summary of the story), orientation (the context of the event), and record of events (series of events that happened and feelings of the speaker). The speech functions in the task also corresponded to the task requirement, for example, describing the time, place and situation, narrating, and expressing feelings. Since the task situation was talking to a friend, some examinees greeted the audience the same way they would talk to a friend before they started their story, and ended with a question asking the friend to share a story with them. The tense was mostly past tense because the story is a past event. No modals were expected from the task prompt. The conjunctions were appropriate to story telling. Most of them were temporal and additive. Some errors in verb tense were quite common in many responses. The data revealed that the examinees understood the task situation well although some stories were not about an event that happened on campus. They may pay more attention to the subtopics of the story or they understood that the word 'on campus' meant 'university related' since some were talking about going to a camp with university friends and some about the Chula-Thammasart Football Competition.

The imaginary task required the examinees to describe an unreal situation. The subtopics were to describe the general and specific characteristics of the ideal world and the ways in which this world would be a better place for everyone. The genre of the test responses was as expected, a description. Most of the speech functions found in the responses were appropriate to the task such as describing and elaborating the characteristics. However, the function 'suggesting ways to make this world become an ideal world' was found in many responses. This may be because the examinees

misinterpreted the third subtopic so they made suggestions instead of describing the living conditions of the people in the ideal world. Other functions, greeting and stating the purpose of the talk, were expressed in some responses. This may be due to the fact that the test situation was presenting the idea in front of the class. The examinees who used these functions may be aware of this feature so they added the functions in their speech. Most of them also made errors in verb tense and modals. They frequently used present simple tense rather than 'would' or unreal past form. The misinterpretation of the subtopic and frequent errors in verb may explain why the means subscores in this task were the lowest. As for the conjunctions, they were fit to the task requirement. Additive conjunctions were used to give additional characteristics, and those expressing condition, reason, result, and example were used to elaborate the description.

The opinion and persuasive tasks shared several similarities because they required the examinees to justify their point of views. In the opinion task, the examinees should express their view about "studying for a Bachelor's degree abroad" and give reasons to support it. In the persuasive task, they should give reasons to support "visiting Thailand" and examples of interesting activities or places for foreigners. The genre of the two tasks was then similar; it consisted of thesis in which the speakers expressed their view or suggestion, and argument in which they justified their point. In addition, similar speech functions were used such as expressing opinion, justifying and elaborating as the task requirements suggested. In expressing these functions, present tense, modals of possibility, prediction and weak obligation, and conjunctions of condition, introduction of idea, reason, and addition were appropriate. One difference that can be seen between the two tasks was how some examinees started their talk. In the opinion task, all explicitly stated their view toward the studying abroad. However, in the persuasive task, some left their view implicit because there was no option of point of view to choose. The task was to give reasons and examples. In addition, none of the opinion responses were found to start with greeting the audience and stating the purpose of the talk like some of the persuasive responses. This is because the test situation of the opinion task was talking to a friend while that of the persuasive task was being a guest in a television show and persuading the viewers.

Finally, the results from the discourse analyses could explain why the imaginary task was frequently the most difficult and the opinion task the easiest in one or many types of subscores as the results from the quantitative analyses suggested. This may be because the imaginary task was more demanding than the other tasks. To perform the task well, one needs to use appropriate verb tense and modals to express unreal situations. Also, many misinterpreted one of the sub questions; their function subscores then were lower than the other tasks because they did not perform it. As for the opinion task, the task was relatively less demanding. The examinees could use the present form of the verb tense and modals well; and only a few speech functions were required to fulfill the task. Thus, it was easier than other tasks.

5.2.4 The attitudes of examinees towards the CBST

The results from the Likert scaled and open-ended questionnaire showed that in general the examinees had positive views towards the CBST as a measure of their speaking ability. They found the test useful and appropriately designed. However, they agreed that the preparation time should have been longer and the imaginary task was too difficult.

The positive attitudes may be due to the appropriate interface design and the characteristics of the test task. The clear and consistent interface did not distract the examinees from the task or create difficulties while taking the test. Thus, the examinees viewed the test well designed. The test environment was less stressful. This may be because in many oral tests that the examinees may have taken before they may be monitored by a teacher or had to interact with a teacher. Therefore, they may feel more stressed and nervous when interacting with someone superior. In a computer lab, however, the examinees may feel more confident since there was no one watching and judging their performance right in front of them.

The findings of positive attitudes towards the CBST provide evidence contradicting the concern that computer-based testing may negatively influence examinees' views of a language test (Chapelle, 2001; Dunkel, 1999). Thus, the results seem to support the use of the computer in testing speaking ability.

The negative attitudes mainly had to do with the imaginary task. As discussed earlier about the weak points of this task, the examinees felt the task was too difficult.

Another negative view that should be addressed was some felt the CBST was unauthentic although many agreed that it could assess their speaking ability. The unauthenticity seems to come from the lack of interlocutors. The CBST task situation generates planned monologues which may be different from many real life situations which creates unplanned dialogues or multilogues. Thus, it is important for both test users and examinees to be aware of making appropriate test inferences that are shaped not only by the constructs being measured but also the task characteristics themselves.

5.3 Conclusions

The present study investigated the trait structure of oral language ability in the CBST; the test taking processes and strategies the examinees used when performing the tasks; the effects of test tasks on the scores and speech produced; and attitudes of the examinees towards the CBST. Since these aspects provide us with complementary information about the CBST, the conclusion that can be drawn from the study is that there is evidence that both supports and challenges the validity of the CBST interpretation. That is, as the CFA results showed, the CBST seems to be able to measure the speaking ability as defined as the knowledge of pronunciation, syntax, vocabulary, cohesion and function although the last knowledge may be problematic due to its low construct reliability. In addition, to fulfill the task requirements, the examinees revealed in their verbal reports that they need to employ not only their language knowledge but also their topical knowledge through the operation of the strategic competence. The speaking product has shown a variety of macro structures and micro speech functions and grammatical features. This may reflect different ranges of oral language ability coming into play. Finally, in general the examinees reported having positive attitudes to the CBST. However, there are also some factors that may weaken the validity of the CBST interpretation. That is, the function variables may be confounded with some measurement errors. Also, some test methods had an important effect on the test scores. The ambiguity of the imaginary task prompt may contribute to the difficulty of the task in all subscores except the pronunciation scores, and the irrelevant speech functions and negative attitudes to the test. All of these may introduce some construct-irrelevant variance to the interpretations of the test scores. With some revision, the CBST could be useful in several language test situations.

5.4 Implications of the study

This study can offer some theoretical and practical implications to the language testing and learning fields. The theoretical implications are, first, the findings support the multidimensional theory of language ability. That is, oral language proficiency in the CBST includes several distinct but related knowledge of pronunciation, syntax, vocabulary, cohesion and function but the last knowledge needs further revision and investigation because of its low construct reliability. In addition, the study shows evidence which supports the operation of strategic competence in a language test situation. The examinees employed a variety of test taking processes and strategies when completing the tasks as shown in the verbal reports. Finally, the study illustrates how to use several quantitative and qualitative approaches to validate a language test interpretation. These procedures can complement each other since they offer different views to help us better understand the test.

The study also offers some practical implications. For instance, test users who would like to develop a computer-based test should bear in mind that the constructs they aim to measure should be clearly defined and the characteristics of the test tasks should be carefully designed since both can have an effect on the interpretation of the test scores. Secondly, language teachers may use the verbal protocol training methods in their classroom. This can help the learners to be aware of different processes and strategies they may employ to successfully perform a language task. Finally, the teachers may teach the learners to use the discourse analytical methods to assess their own speech. This may help learners to improve their speaking by observing the speech produced and compared them with the task requirements.

5.5 Recommendations

Some recommendations could be made for oral language testing, teaching and further studies. For oral language testers, it is recommended that in order to measure oral language ability it is important to incorporate different types of language ability which should be assessed separately. This suggests the use of analytical rating scales which could inform us the speaking performance in several types of knowledge. Also, it is suggested that oral language test developers who are interested in using Item Response Theory (IRT) should check its dimensionality assumption by performing

EFA and CFA. If multidimensionality of constructs is found like that in oral language ability in this study, they should employ multidimensional IRT models rather than the unidimensional one since the former would correspond with the assumption of multicomponential oral language ability better. Finally, speaking test users should keep in mind that test task characteristics may have an influence on test performance and interpretation of the scores. Thus, they should investigate the evidence which may support and challenge the test usefulness before deciding to use or develop a speaking test.

Regarding language teaching, teachers should pay attention to each of the ability components in teaching SL/FL speaking. That is, they should design a curriculum that aims to develop students' speaking ability in several areas. It should not focus on only some components such as grammar and vocabulary since other components also pay a role in using and delivering successful oral communication. Also, teachers should draw students' attention to examining speaking task characteristics and its requirement to produce speech appropriate to specific situations.

As noted earlier that one of the limitations of this study is the constructs being investigated include only five areas. Thus, it is recommended that future research should examine other types of oral language knowledge which may constitute fluency, for example, knowledge of how to use appropriate speech rate for a particular situation and knowledge of communicative devices such as fillers to make the talk smooth and flowing. These features should be incorporated into the construct definition and the way to operationalize them in a speaking test should be examined. In addition, the tasks in this study generate monologic responses. Thus, there should be other studies which investigate the effects of computer-based speaking tests on other types of speech such as dialogues and multilogs. Another research area that would be very interesting is to investigate the similarities and differences between a computer-based speaking test and other direct speaking tests such as group oral testing. Finally, the replication of this study with other groups of examinees would provide insights into the trait structure of oral language ability to see if the model in the present study holds in other situations.

REFERENCES

Alderson, J. C. (1996). Do corpora have a role in language assessment? In J. Thomas & M. Short (Eds.), Using corpora for language research. (pp. 248-259). Harlow: Longman.

Alderson, J. C. (2000). Technology in testing: The present and the future. System, 28, 596-603.

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, 41-66.

Anderson-Hsieh, J. & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. Language Learning, 38, 561-613.

Ary, D., Jacobs, L. C., & Razavieh, A. (2002). Introduction to research in education. USA: Wadsworth.

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

Bachman, L. F. & Eignor, D. (1997) In C. Clapham & D. Corson (Eds.), Encyclopedia of language and education volume 7: Language testing and assessment. (pp. 227-242).

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. Language Learning, 31, 67-86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. TESOL Quarterly, 16, 449-465.

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. Language Testing, 14-29.

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

Banerjee, J. & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham & D. Corson (Eds.), Encyclopedia of language and education volume 7: Language testing and assessment. (pp. 275-287).

Barcikowski, R. S. and Robey, R. R. (1994). Use and misuse of repeated measures designs. In L. Collins and L.A. Seitz (Eds.), Advances in data analysis for prevention intervention research. NIDA Research Monograph (pp. 302-341).

Bentler, P. M. & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), <u>Multivariate analysis VI</u> (pp. 9-42). Amsterdam: North-Holland.

Bentler, P. M. & Wu, E. J. C. (2002). <u>EQS for Windows: User's guide.</u> CA: Multivariate Software.

Bhatia, V. K. (1993). <u>Analysing genre: language use in professional settings</u>. London: Longman.

Bhatia, V. K. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), Academic discourse. England: Longman. (pp. 21-39).

Bongaerts, T. & Poulisse, N. (1989). Communication strategies in L1 and L2: Same or different? <u>Applied Linguistics, 10,</u> 253-268.

Brown, A. (1993). The role of test-taker feedback in the test development process: test takers' reactions to a tape-mediated test of proficiency in spoken Japanese. <u>Language Testing, 10,</u> 277-304.

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. <u>Language Testing, 12,</u> 1-15.

Brown, H. D. (2001). <u>Teaching by principles: Interactive language teaching methodology.</u> Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J. D. (1997). Computers in language testing: Present research and some future directions. <u>Language Learning & Technology, 1,</u> 44-59. Available from: http://llt.msu.edu/vol1num1/brown/default.html[2001, January 15]

Bunton, D. (2002). Generic moves in Ph.D. thesis introductions. In J. Flowerdew (Ed.), Academic discourse. England: Longman. (pp. 57-75).

Bygate, M. (1987). <u>Speaking.</u> Oxford: Oxford University Press.

Byrne, B. M. (1994). <u>Structural equation modeling with EQS and EQS/Windows: Basic concepts, application and programming.</u> Thousand Oaks: Sage Publications.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. <u>Psychological Bulletin, 56,</u> 81-105.

Carroll, J. B. (1983). Psychometric theory and language testing. In J. W. Oller, Jr. (Ed.) <u>Issues in language testing research.</u> Rowley, MA: Newbury House.

Chalhoub-Deville, M. (1996). Perfromance assessment and the components of oral construct across different tests and tater groups. In M. Milanovic & N. Saville (Eds.), Studies in language testing 3: Performance testing, cognition, and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem. (pp. 55-73)Cambridge: Cambridge University Press.

Chalhoub-Deville, M. & Deville, C. (1999). Computer-adaptive testing in second language contexts. Annual Review of Applied Linguistics, 19, 273-99.

Chapelle, C. (2001). Computer applications in second language acquisition: Foundations for teaching, testing, and research. Cambridge, UK: Cambridge University Press.

Child, D. (1990). The essentials of factor analysis. London: Cassell.

Clark, J.L.D. (1985). Development of a tape-mediated, ACTFL/ILR scale-based test of Chinese speaking proficiency. In C. W. Stansfield. (Ed.), Technology and language testing. Princeton, NJ: Education Testing Service.

Clark, J. L. D. (1987). A study of the comparability of speaking proficiency interview ratings across three government language training agencies. In K. M. Bailey, T. L. Dale, & R. T. Clifford. (Eds.), Language testing research: selected papers from the 1986 colloquium. Monterey, DA: Defense Language Institute.

Clark, J. & Hooshmand, D. (1992). "Screen-to-screen" testing: an exploratory study of oral proficiency interviewing using video teleconferencing. System, 20, 293-304.

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

Cohen, A. D. (1998). Strategies in learning and using a second language. London: Longman.

Cohen, A. D. (2000). Exploring strategies in test taking: Fine-tuning verbal reports from respondents. In G. Ekbatani &H. Pierson., (Eds.), Learner-directed assessment in ESL. (pp. 127-150). London: Lawrence Erlbaum Associates.

Cohen, A. D. & Dornyei, Z. (2002). Focus on the language learner: Motivation, styles and strategies. In N. Schmitt (Ed.), An Introduction to applied linguistics. London: Arnold.

Cohen, A. D. & Olshtain, E. (1993). The production of speech acts by EFL learners. TESOL Quarterly, 27, 33-56.

Cohen, A. D., & Weaver, S. F., & Li, T. (1998). The impact of strategic-based instruction on speaking a foreign language. In A. D. Cohen (Ed.), Strategies in learning and using a second language (pp. 107-56). London: Longman.

Coniam, D. (1997). A computerized English language proofing cloze program. Computer Assisted Language Learning, 10, 83-97.

Dandonoli, P. & Henning, G. (1990). An investigation of the construct validity of the ACTFL Proficiency Guidelines and oral interview procedure. Foreign Language Annals, 23, 11-22.

Davies, C. E. (1998). Maintaining American Face in the Korean Oral Exam: Reflections on the Power of Cross-Cultural Context. In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency (pp. 271-296). Philadelphia: John Benjamins.

Dunkel, P. (1991). Computerized Testing of Nonparticipatory L2 Listening Comprehension Proficiency: An ESL Prototype Development Effort. The Modern Language Learning, 75, 64-73.

Dunkel, P. (1999). Research and development of a computer-adaptive test of listening comprehension in the less commonly-taught language Huasa. In M. Chalhoub-Deville (Ed.), Development and research in computer adaptive language tesing (pp. 91-121). Cambridge University of Cambridge Examinations Syndicate/Cambridge University Press.

Eggins, S. and Slade, D. (1997). Analysing Causal Conversation. London and Washington: Cassell.

Færch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Færch, & G. Kasper (Eds.), Strategies in interlanguage communication (pp. 20-60). London: Longman.

Flewelling, J. & Snider, D. (2001). "CARLA" (Computer-assisted recorded language assessment)-software for evaluating student oral responses. Available from: http://www.icte.org/T01_Library/T01_166.PDF[2001, January, 25]

Fouly, K.A., Bachman, L. F., & Cziko, G.A. (1990). The divisibility of language competence: a confirmatory approach. Language Learning, 40, 1-21.

Fulcher, G. (2000). Computers in language testing. In P. Brett & G. Motteram (Eds.) A special interest in computer: learning and teaching with information and communications technologies. Kent: IATEFL.

Fulcher, G. (2003). Interface design in computer-based language testing. Language Testing, 384-408.

Glass, G. V. & Hopkins, K. D. (1996). Statistical methods in education and psychology. Boston: Allyn & Bacon.

Green, A. (1998). Verbal protocol analysis in language testing research: A handbook. UK: Cambridge University Press.

Hair, J. F., Anderson, R.E., Tatham, R. L. & Black, W.C. (1995). Multivariate data analysis with readings. New Jersey: Prentice Hall.

Hatch, E. (1992). Discourse and language education. Cambridge: Cambridge University Press.

Hatcher, L. (1994) A step-by-step approach to using the SAS system for factor analysis and structural equation modeling: Focus on the CALIS procedure. Cary, NC: SAS Institute.

He, A. W. (1998). Answering Questions in LPIs: A case study. In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency. (pp. 101-116). Philadelphia: John Benjamins.

He, A. W. & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency. (pp. 1-26). Philadelphia: John Benjamins.

Henning, G., Anbar, M., Helm. E. E. & D'Arcy, S. J. (1993). Computer-assisted testing of reading comprehension: comparisons among multiple-choice and open-ended scoring methods. In D. Douglas & C. Chapelle, (Eds.), A new decade of language testing research: selected papers from the 1990 Language

Testing Research Colloquium. (pp. 123-131). USA: Teachers of English to Speakers of Other Language Inc.

Hill, K. (1998). The effect of test-taker characteristics on reactions to and performance on an oral English proficiency test. In A. Kunnan (Ed.). Validation in Language assessment: Selected papers from the 17ᵗʰ Language Testing Research Colloquium. Long Beach. (pp. 209-229). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Hinton, P. R. (2004). Statistics explained. London: Routledge.

Hinton, P. R., Brownlow, C., McMurray, I. & Cozens, B. (2004). SPSS explained. London: Routledge.

Hoekje, B & Linnell, K. (1994). "Authenticity" in language testing: Evaluation spoken language tests for international teaching assistants. TESOL Quarterly, 28, 103-126.

Hymes, D. H. (1972). Models of interaction of language and social life. In J. J. Gumperz & D. H. Hymes (Eds.). Directions in sociolinguistics: The ethnography of communication. (pp. 35-71). New York: Holt, Rinehart and Winston.

Isaac, S. & Michael, W. B. (1995). Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design and evaluation of studies in education and the behavioral sciences (3ʳᵈ edition). San Diego, CA: EdITS/Educational and Industrial Testing Services.

Jenkins, S. & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decision. The Modern Language Journal, 87, 90-107.

Johnson, M. & Tyler, A. (1998). Re-analyzing the OPI: How much does it look like natural conversation? In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency. (pp. 27-52). Philadelphia: John Benjamins.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models. (pp. 294-316). Newbury Park, CA: Sage.

Kasper, G., & Kellerman, E. (1997). Introduction: approaches to communication strategies. In G. Kasper, & E. Kellerman (Eds.), Communication strategies: Psycholinguistic and sociolinguistic perspectives. (pp. 1-14). London: Longman.

Katona, L. (1998). Meaning negotiation in the Hungarian Oral Proficiency Examination of English. In R. Young & A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency (pp. 239-270). Philadelphia: John Benjamins.

Kellerman, E. (1991). Compensatory strategies in second language research: A critique, a revision, some (non-) implications for the classroom. In R., Phillipson, E. Kellerman, L. Selinker, M. S. Smith & M. Swain (Eds.), Foreign/second language pedagogy research. (pp. 142-161). Clevedon, Avon: Multilingual Matters,.

Kenyon, D.M. & Malabonga, V. (2001). Comparing examinee attitudes towards computer-assisted and other oral proficiency assessments. Language Learning & Technology, 5, 60-83.

Kim, K. & Suh, K. (1998). Confirmation Sequences as Interactional Resources in Korean Language Proficiency Interviews. In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency (pp. 297-332). Philadelphia: John Benjamins.

Koike, D. A. (1998). What happens when there's no one to talk to? Spanish foreign language discourse in simulated oral proficiency interviews. In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency (pp. 69-98). Philadelphia: John Benjamins.

Kormos, J. (1999). Simulating conversations in oral –proficiency assessment: A conversation analysis of role plays and non-scripted interviews in language exams. Language Testing, 16, 163-188.

Lazaraton, A. (1992). The structural organization of a language interview: A conversational analytic perspective. System, 20, 373-386.

Lazaraton, A. (1996). A qualitative approach to monitoring examiner conduct in CASE. In M. Milanovic & N. Saville (Eds.), Studies in language testing 3: Performance testing, cognition, and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem. (pp. 18-33). Cambridge: Cambridge University Press.

Lazaraton, A. (2002). A qualitative approach to the validation of oral language tests. New York: Cambridge University Press.

Lazaraton, A. and Riggenbach, H. (1990). Oral skills testing: A rhetorical task approach. Issues in Applied Linguistics, 1, 196-217.

Long, J. S. (1983). Confirmatory factor analysis: A preface to LISREL. USA: Sage Publications.

Lumley, T. (2002). Assessment criteria in a large scale writing test: What do they really mean to the raters? Language Testing, 19, 246-276.

MacCallum, R. C., Widaman, K. F., Preacher, K. J. & Hong, S. (2001). Sample size in factor analysis: The role of model error. Multivariate Behavioral Research, 36, 611-637.

Madsen, H. (1991). Computer-adaptive testing of listening and reading comprehension: The Brigham Young University approach. In P. Dunkel (Ed.), Computer-assisted language learning and testing: Research issues and practice. (pp. 237-258). New York: Newbury House.

Martin, J. R. (1989). Factual Writing: Exploring and Challenging Social Reality. Oxford: Oxford University Press.

Martin, J. R. (1992). English Text: System and Structure. Amsterdam: Benjamins

McNamara, T.F. & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. Language Testing, 14, 140-156.

Norris, J. M. (2001). Concerns with computerized adaptive oral proficiency assessment. Language Learning & Technology, 5, 99-105.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. Language Testing,12, 217-237.

O'Loughlin, K. (2001). The equivalence of direct and semi-direct speaking tests. Cambridge: Cambridge University Press.

Ordinate. (1998). The PhonePass Test. Available from: http://www.ordinate.com[2001, January, 6]

Paribakht, T. (1985). Strategic competence and language proficiency. Applied Linguistics, 6, 132-146.

Plum, G. A. (1988). Text and contextual conditioning in spoken English: a genre-based approach. PhD thesis, University of Sydney.

Poulisse, N., Bongaerts, T., & Kellerman, E. (1984). On the use of compensatory strategies in second language performance. Interlanguage Studies Bulletin, 8, 70-105.

Powers, D. E., Fowles, M. E., Farnum, M. & Ramsey, P. (1994). Will they think less of a handwritten essay if others wordprocess theirs? Effects on essay scores of intermingling handwritten and word-processed essays. Journal of Educational Measuremnt, 31, 220-233.

Ranney, S. (1992). Learning a new script: An exploration of sociolinguistic competence. Applied Linguistics, 13, 25-50.

Ross, S. (1998). Divergent Frame Interpretations in Oral Proficiency Interview Interaction. In R. Young &A.W. He (Eds.), Talking and testing: Discourse approaches to the assessment of oral proficiency. (pp. 333-354). Philadelphia: John Benjamins.

Satorra, A. & Bentler, P. M. (1988a). Scaling corrections for chi square statistics in covariance structure analysis. American Statistical Association 1988 Proceedings of the Business and Economic Sections (pp. 308-313). Alexandria, VA: American Statistical Association.

Satorra, A. & Bentler, P. M. (1988b). Scaling corrections for statistics in covariance structure analysis. (UCLA Statistics Series 2). Los Angeles: University of California at Los Angeles, Department of Psychology.

Sawaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. Language Learning, 43: 313-344.

Schermelleh-Engel, K. Moosbrugger, H. & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research[Online]. Available from: http://www.mpr-online.de [2006, October 1]

Shavelson, R. J. & Webb, N. M. (1991). Generalizability theory: A primer. Newbury Park, CA: Sage.

Shohamy, E. (1982). Affective considerations in language testing. Modern Language Journal, 66, 13-17.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. Language Testing, 99-123.

Skehan, P. (1998). A cognitive approach to language learning. London: Oxford University Press.

SPSS Inc. (2004). SPSS for Windows Release 13.0. Chicago: SPSS, Inocrporated.

Stansfield, C. W. (1991). A comparative analysis of simulated and direct oral proficiency interviews. In Anivan, S. (Ed.), Current developments in language testing. Singapore: Regional English Language Center.

Stansfield, C. W., Kenyon, D. M. (1992). Research on the comparability of the Oral Proficiency Interview and the Simulated Oral Proficiency Interview. System, 20, 347-364.

Stansfield, C. W., Kenyon, D. M., Paiva, R., Doyle, F., Ulsh, I. & Cowles, M. A. (1990). The development and validation of the Portuguese speaking test. Hispania, 73, 641-51.

Tabachnick, B. & Fidell, L. S. (2001). Using multivariate statistics. Boston: Allyn and Bacon.

Tarone, E. (1977). Conscious communication strategies in interlanguage. In H. Brown, C. Yorio, & R. Crymes (Eds.), On TESOL'77 (pp. 194-203). Washington, DC: TESOL.

Tarone, E. (1983). Some thoughts on the notion of 'communication strategy'. In C. Færch & G. Kasper (Eds.), <u>Strategies in interlanguage communication</u> (pp. 61-74). London: Longman.

Tarone, E., & Yule, G. (1989). <u>Focus on the learner.</u> Oxford: Oxford University Press.

Taylor, C., Kirsch, U., Eignor, D. & Jamieson, J (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. <u>Language Learning, 49,</u> 219-274.

Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. <u>TESOL Quarterly, 21</u> (4), 713-29.

Underhill, N. (1987). <u>Testing spoken language: A handbook of oral testing techniques.</u> Cambridge: Cambridge University Press.

Wiersma, W. & Jurs, S. G. (2005). <u>Research methods in education: an introduction.</u> Boston: Pearson.

Wigglesworth, F. (1997). An investigation of planning time and proficiency level on oral test discourse. <u>Language Testing, 14,</u> 85-106.

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants<u>. TESOL Quarterly, 26</u> (4), 693-711.

Yoshida-Morise, Y. (1998). The use of communication strategies in language proficiency interviews. In R. Young &A.W. He (Eds.), <u>Talking and testing: Discourse approaches to the assessment of oral proficiency.</u> (pp. 205-238). Philadelphia: John Benjamins.

Young, R. & Halleck, G. (1998). "Let Them Eat Cake!" of How to Avoid Losing Your Head in Cross-Cultural Conversations. In R. Young &A.W. He (Eds.), <u>Talking and testing: Discourse approaches to the assessment of oral proficiency.</u> (pp. 355-382). Philadelphia: John Benjamins.

Zeidner, M. & Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as a foreign language. <u>Language Testing, 5,</u> 100-114.
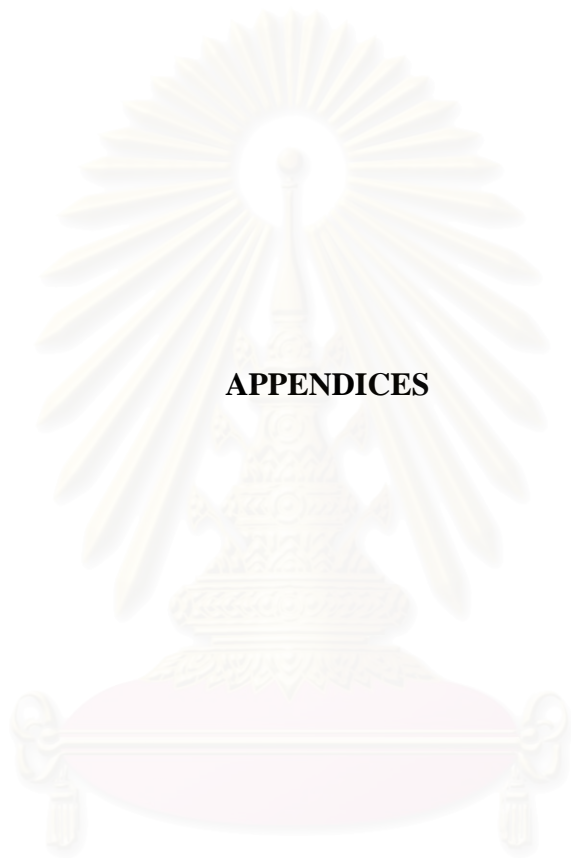
APPENDICES

Appendix A
Computer-based Speaking Test

Instructions: This test consists of four parts and it will last approximately 25 minutes. In each part, you will speak for **1:30 minutes** about a general topic. Before you speak, you will have **two minutes** to prepare. During the preparation time, you can make some notes on the paper given if you wish.

Part 1

Imagine that I am **your close friend**. We haven't seen each other for some time. Today, we met and you want to tell me what has happened to you recently.

Situation: Talking to a friend

Topic: The most enjoyable day on campus

Please be sure to talk about:

- *when the story happened*
- *where it happened*
- *what happened first*
- *what happened next*
- *how it ended*
- *how you felt about it.*

Part 2

Imagine that I am **one of your friends**. I asked you what you think about studying for a Bachelor's degree abroad.

Situation: Talking to a friend

Topic: What do you think about studying for a Bachelor's degree abroad?

Please be sure to talk about:

- *whether you think studying for a Bachelor's degree abroad is a good idea or not*

- *reasons to support your opinion.*

Part 3

Imagine that you are invited to **a television show**. The host asks you to persuade the viewers to visit Thailand. The viewers are **foreigners**.

Situation: Persuading foreigners to visit Thailand

Topic: How would you persuade foreigners to spend their vacation in Thailand?

Please be sure to talk about:

- *at least two reasons why they should come to Thailand*

- *and give at least two examples of what they can do or see in Thailand.*

Part 4

Suppose that your teacher asked you to present your idea in class about an imaginary situation.

Situation: Talking in front of the class

Topic: What would your ideal world be like?

Please be sure to talk about:

- *what this imaginary ideal world would be like in general*

- *at least two characteristics of this ideal world*

- *the ways in which this world would be a better place for everyone.*

End of the test.

Thank you very much for your participation.

Appendix B

Scoring Scales

After the examinees finished the computer-based speaking test, their speech was rated by two raters. The rating scales below describe different ability levels within the various areas of language knowledge. The examinees were scored from 1-5 on knowledge of pronunciation, syntax, vocabulary, cohesion and functions.

Theoretical construct definition	Operational construct definition:
Knowledge of pronunciation	Evidence of pronunciation that is comprehensible, and as rated on the following scales.
Knowledge of syntax	Evidence of accurate use of a variety of syntactic structures, as demonstrated in the context of the specific test tasks, and as rated on the following scales.
Knowledge of general purpose vocabulary items, including cultural references	Evidence of accurate use of a variety of general purpose vocabulary items, including cultural references, as demonstrated in the context of the specific test tasks, and as rated on the following scales.
Knowledge of language forms for explicitly marking cohesive textual relationships	Evidence of accurate use of a variety of language forms for explicitly marking cohesive textual relationships, as demonstrated in the context of the specific test tasks, and as rated on the following scales.
Knowledge of various functions	Evidence of appropriate use of various functions (e.g. narrating, expressing opinions), as demonstrated in the context of the specific test tasks* and as rated on the following scales.

Levels of ability/mastery	Pronunciation
5 **Excellent**	easily understood; uses stress and intonation appropriately; no errors
4 **Good**	generally easily understood; uses stress and intonation appropriately; few errors but they do not cause difficulty in understanding
3 **Moderate**	somewhat comprehensible; noticeable pronunciation errors which may cause some difficulty in understanding
2 **Poor**	accented; frequently incomprehensible
1 **Very poor**	highly accented; almost always incomprehensible

Levels of ability/mastery	Syntax
5 **Excellent**	uses a wide range of syntactic structures highly accurately; no errors
4 **Good**	uses a wide range of syntactic structures generally accurately; produce few errors but they do not lead to miscomprehension
3 **Moderate**	uses some syntactic structures accurately; errors occasionally lead to miscomprehension
2 **Poor**	frequently uses syntactic structures inaccurately
1 **Very poor**	almost always uses syntactic structures inaccurately

Levels of ability/mastery	Vocabulary
5 **Excellent**	uses a wide range of vocabulary accurately; no errors
4 **Good**	uses a wide range of vocabulary accurately; few errors in selecting word choices
3 **Moderate**	uses some words or phrases accurately; sufficient range of vocabulary to complete the task; errors may sometimes lead to miscomprehension
2 **Poor**	frequently uses inaccurate words or phrases; unable to complete the task due to limited vocabulary
1 **Very poor**	almost always uses inaccurate words or phrases; unable to complete the task due to very limited vocabulary

Levels of ability/mastery	Cohesion
5 **Excellent**	uses a wide range of explicit devices highly accurately; relationships between sentences are clear; no errors
4 **Good**	uses a wide range of explicit devices accurately; relationships between sentences generally clear; few errors in cohesion but they do not lead to confusion
3 **Moderate**	uses some explicit devices to connect ideas; relationships between sentences somewhat clear; noticeable errors in cohesion
2 **Poor**	uses few markers of cohesion; relationships between sentences frequently confusing
1 **Very poor**	uses very few markers of cohesion to connect utterances; relationships between sentences confusing

Levels of ability/mastery	Function*
5 **Excellent**	performs all the required functions appropriately. Elaborates well. Gives relevant and adequate information to fulfill the task requirements.
4 **Good**	performs all the required functions but gives little or some extraneous information.
3 **Moderate**	performs at least half but not all of the required functions. May also give some inadequate or irrelevant information.
2 **Poor**	performs less than half of the required functions. May also give some inadequate or irrelevant information.
1 **Very poor**	does not perform any functions required.

Please see below the lists of the functions required to carry out each test task.

Tasks	Functions required to complete the tasks.
Part1: Narrative (The most enjoyable day on campus)	1. Describes the time. 2. Describes the place. 3. Describes the beginning of the story. 4. Describes the middle of the story. 5. Describes the ending of the story. 6. Expresses feelings toward the story.
Part 2: Opinion (Studying for a Bachelor's degree abroad)	1. States one's opinion whether agreeing or disagreeing with the topic. 2. Gives at least 2 reasons to support one's opinion. (In the pilot study, one reason is enough to fulfill the task requirement.)
Part 3: Persuading (Persuading the viewers to visit one's country)	1. Recommends the audience to visit the country. 2. Gives at least 2 reasons to support the recommendation. 3. Gives at least 2 examples of places or activities for the audience.
Part 4: Imaginary	1. Describes what this ideal world would be like in general. 2. Describes 2 characteristics of the ideal world. 3. Describes how this world would be a better place for everyone.

Appendix C

Computer-based Speaking Test Questionnaire

This survey aims to gather information about your opinion towards the computer-based speaking test (CBST) that you have just taken. The information you give will help us better understand the reactions of test takers to the test, which in turn will help evaluate the appropriateness of the use of such test in measuring speaking ability of first-year university students. Your answers to any or all questions will be confidential. Please feel free to express your opinions. Thank you very much for your help.

There are three parts in this survey.

Part 1 Please complete these details.

ID Number: _____

Faculty: _____

Part 2 Please indicate your opinion after each statement by checking the space in the table that best indicates the extent to which you agree or disagree with the statement.

	Strongly agree 4	Agree 3	Disagree 2	Strongly disagree 1
1. I feel I had the opportunity to adequately demonstrate my speaking ability on the computer-based speaking test (CBST).				
2. I felt comfortable taking the speaking test via computer.				
3. I did not understand the instructions on how to take the test.				
4. I think the warm-up part was useful.				
5. I think the CBST would give a rater with an accurate idea of my speaking ability.				
6. The computer is a useful instrument to test speaking ability.				
7. The instructions on how to take the CBST are easy to follow.				
8. The procedures for taking the test were too complicated.				
9. A speaking test should not be delivered by computer.				

	Strongly agree 4	Agree 3	Disagree 2	Strongly disagree 1
10. Taking the oral language test delivered by computer was an unpleasant experience.				
11. The narration task was too difficult.				
12. The opinion task was too difficult.				
13. The imagination task was too difficult.				
14. The persuasion task was too difficult.				
The preparation time allowed was too short for 15. the narration task				
16. the opinion task				
17. the imagination task				
18. the persuasion task				
The response time allowed for was too short for 19. the narration task				
20. the opinion task				
21. the imagination task				
22. the persuasion task				

Part 3 Please write down your answers in the space provided.

23. Please look at statement 1 (I feel I had the opportunity to adequately demonstrate my speaking ability on the computer-base speaking test.) If you checked "strongly disagree", or "disagree", please explain why. Write your answer in the space provided.

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

24. Please write down what you <u>liked</u> about the test and give reasons in the space provided.

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

25. Please write down what you <u>did not like</u> about the test and give reasons in the space provided.

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

Thank you very much for your cooperation. ☺

Appendix D

Verbal Report Collection Procedures For the Pilot Study

Verbal Report Collection Procedures in the Pilot Study

The participants who would give verbal report did the test individually in a computer laboratory. The test procedures and environment were similar to the participants who did not give verbal report. However, in addition to computer recorded, the responses were audio recorded so that they could be played back during the interview.

The participants were informed about the purpose of the research study, the description of the test and the process of verbal report collection. They did the first test task and gave verbal reports in Thai, did the second task, gave reports, and so on. In order to help them recall, they could review their notes and listen to their own responses to the test asks. The researcher also asked her some fixed and probing questions about their test-taking processes and strategies.

The interview, which was audio recorded, consisted of two parts. In the first part, the researcher asked questions about how the participants prepared the responses, and the researcher also added probes according to what the participants had written in the notes during their test preparation period (List A). In the second part, the questions focused on how they actually responded to the tasks. In addition, the researcher asked probes based on the recorded responses (List B).

The questions in the lists were adapted from Cohen & Olshtain, 1993:

List A

1. What was the first thing you thought about after you read the test topic and situation?
2. What else did you think about?
3. What did you write in your notes?
4. How did you come up with the vocabulary and phrases in your responses?
5. Did you think in Thai or English while writing down your notes? Which parts?

List B

1. When you were answering, did you use your notes only? Did you think out the response as you went along?

2. Did you have a number of alternatives in vocabulary and phrases? Why did you choose that response?

3. Were you thinking about grammar while responding?

4. Did you think about pronunciation while you were answering the question?

Appendix E

Procedures for verbal protocol collection in the main study

The instructions for the tutorial of the stimulated verbal report procedures were adapted from Ericsson and Simon (1993) and Green (1997). They were read to and given to the participants in print as follows:

In this study I am interested in your thoughts when you work on the tasks I am going to give you. To do this, I am going to ask you to do the first task. After you finish, I would like you to tell me all that you can remember about your thinking from the time you started the task until you completed it.

- Please tell me about your memories in the sequence in which they occurred while preparing and giving the answers. Please start your report saying "I first thought of…"
- You can talk in English, Thai or a mixture of the two.
- To help you remember your thoughts, you can look at the question and your notes that you took when you prepared for the answer. Also, you can listen to your response that will be recorded. You can play and pause it as you like.
- If you are uncertain about any of your memories, please let me know.
- I don't want you to work on the task again, just report all that you can remember thinking about from the time you read the question until you gave the answer. Also, don't plan out or try to explain to me why you thought in a certain way.
- If you are silent for any long period, I will ask you to continue your talk.
- Your talk will be audio recorded so please speak loudly.

Do you have any questions?

Now let's do some practice tasks.

Practice Task 1

Add these two numbers. After finishing the task, tell me all that you can remember about your thinking.

- 15 and 27

Practice Task 2

Judge whether two letters are in alphabetical order or not. After finishing the task, tell me all that you can remember about your thinking.

- B S

Judge whether two letters are in alphabetical order or not. After finishing the task, tell me all that you can remember about your thinking.

- M L

Practice Task 3

Answer the question below. You have 15 seconds to prepare your response and 30 seconds to answer. You can use the paper provided. Your answer will be recorded. After finishing the task, tell me all that you can remember about your thinking when you prepared and gave the answer. You can use your notes and play back the tape to help refresh your memories.

- What do you plan to do after you graduate from Chulalongkorn University?

Now are you ready to start the first test task?

Appendix F
Sample responses and their test scores

No. 135 (Lower Intermediate)

1. Narrative task
um (4) on freshy days (4) at (3) uh (5) on (1) freshy day at (3) Sci 25 (3) building (2)
um (2) I (3) dancing and sing a song (4) with my friend (3) uh (5) uh (1) in the end (4)
I (2) in the end I (3) in the end I (1)in the end I (9) in the end I

2. Opinion task
I think that (2) studying for (1) uh (2) bachelor's degree abroad is a goods idea
because it make me um (4) in trend and (7) it make me in trend (2) update (3) um (7)
and (5) it make me in trend and update (?) (22) and find (1) and find new friend

3. Imaginary task
in my imaginary imaginary in my imaginary I would like (2) uh (3) not (2) not have
(1) a war (3) in the world (1) and (2) not have um (6) and not have a (6) not have a (8)
not have a (6) %ภัยธรรมชาติ ภัยธรรมชาติ% (6) I want (6) I want (2) uh (4) the world is uh
beautiful and uh

4. Persuasive task
Thailand is a beautiful country (1) um have many fruit and good um (1) Thailand has
a beautiful building (1) um I think that (2) um (6) Thailand (1) Thailand is (7) um the
people in Thailand is sincere (1) and lively (1) um I want uh (2) I want (2) um (5) um
(3) the (1) traveler to (1) travel in Thailand (9) because (1) they (2) can find

No. 135	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	3	3	3	3
Syntax	3	3	2.5	3
Vocabulary	3	3	2.5	3
Cohesion	3	3	2.5	3.5
Function	3	4	2.5	3

No. 18 (Lower Intermediate)

1. Narrative task
in fourth of August (2) I (?) to join music for freshy night (?) this freshy night at my faculty it's uh in Sala Pra Keaw (1) in party I joy (1) I enjoy with my friend eat a lot of food and it has fashion show from each model (4) and

2. Opinion task
I think it's is a good idea to study for a Bachelor's degree abroad (1) first I think it's makes learner know more thing to help to use in learner's life (1) um second I think it's can help learner to use more effective foreign language (2) to use in learner's life

3. Imaginary task
I think my ideal worlds must have many factor such as good people and good environment (1) that people can do (1) that people and environment can live together (1) people should protect about environment by many method (1) such as (2) um (2) to (1) don't don't cut (1) don't cut (1) forest (2) and (1) save the forest (10) and people (2) live together (1) in (1) in (2) in (1) in kindly (1) in kindly (2) and don't be jealous (2) or selfish (3) or (5) for this I think it's (1) a better place for everyone to live together

4. Persuasive task
Thailand has many good place to visit (1) and almost people are kindly (6) when you visit Thailand you can see a beautiful sea at the south of Thailand (1) in Phuket or Krabi province (2) it has many manys many many beautiful beach (2) and (2) if you want to see (2) beautiful (1) beautiful (3) beautiful (2) beautiful forest you can go to the north (2) the north of Thailand to see the elephant (5) the elephant and (2) beautiful forest

No. 18	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	3	3	3	3
Syntax	3	3	3	3
Vocabulary	3	3.5	3.5	3.5
Cohesion	3	3	3	3
Function	3	4	4	3.5

No. 53 (Intermediate)

1. Persuasive task
there are many interest in interested in interested places in Thailand (1) uh in the east there are there is a Chonburi province there they it has (1) a beautiful beach and many entertainment to do at the night such as um (2) night market uh in the south there are many beautiful beach too and they have (1) uh (2) beautif… and they have a long beach (1) and (3) have a (4) and they has many (2) h… many hotel or in Bangkok there are many entertaining entertainment things to do such as department store or JJ market because (2) they they buy a many cheap things

2. Imaginary task
in the future I think our world will have many environment and lacking of natural source (2) then we should solve this problem (1) uh I want to live in (1) clean and full of natural source world (1) um people and people live together with natural our world will be like that every… everyone help to (2) to or do something not (1) destroy our environment

3. Opinion task
I think it's unnecessary to going abroad because there are many international program in many university in Thailand (1) and (2) uh (3) and it's not want to (2) pay many expense (1) if you (1) if you study in national program in Thailand you will have many Thai friends because of when you graduate from another (1) another city you (1) you won't have many Thai friend like student in studying in Thai

4. Narrative task
story is on first day of my first meeting friends in this university (1) senior are located former year to many group (1) first senior divided all new (1) all former year in twelve group (1) a… after that they advise many things many rule in this faculty such as (1) uh curriculum (3) they advise us all day and in the evening (1) they (1) get me to my group (1) and I partici… participated with my new friends and (1) my senior (3) I (1) I think it's a good thing to have senior because (3) it's make me (1) uh unity (3) it's it's make everyone are unity

No. 53	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	3.5	3.5	3.5	3.5
Syntax	3.5	3	3	3
Vocabulary	4	3.5	3.5	3.5
Cohesion	3.5	3.5	3.5	3.5
Function	3	4	3	4

No. 42 (Intermediate)

1. Imaginary Task
my imaginary ideal worlds (2) would be like (2) people live together with peaceful and (3) help to reserve the nature (4) um (2) I the ideal world (1) um have two characteristic is (1) peace and everyone love each other (3) and the way in which this world would be a better place for everyone is (1) everyone help each other to s… to (3) to save the world by reserve the nature and (4) love each other (4) that's will make (2) and (1) that will make the world (3) be a better place (1) for everyone

2. Persuasive task
the reason why they why you should come to Thailand is because Thailand is an interesting places (1) and um such as Pattaya and Koh Samui (1) that has white sand and beaches for you to come to see it Thailand has a lot of interesting culture in each part of Thailand (2) and the second is (1) Thai people is very kind and friendly (1) to the foreigner (3) the example of what you can do or see in Thailand is you can go to an interesting places and take photos (2) or when you come to when you go to (1) s… to Pattaya or Koh Samui you can (3) you can (4) swimming in the sea and (4) and see the coral reef (2) that's very beautiful (3) or you can eat some Thai food

3. Narrative task
the story happen (1) is the day that has freshy day and freshy night (1) I have to do a lot of activity that day (4) I met a lot of new friends and spoke with them (2) I also saw concert and eating some food (4) it's happen at my faculty (6) the first we met (1) we m… I meet (1) all my friends at my faculty and the next is we do a lot of (1) activities together (2) I spent a lot of time with my friends to do a lot of activities and (2) this day is end (1) ended (2) about (4) twenty two o'clock (2) and I felt very happy about this activities and I

4. Opinion task
I think studying for Bac… for a Bachelor degree abroad is a good idea because you have (?) experience and meet a lot of kind of people (1) and you will make a lot of friends (1) uh (4) and the second reasons is (2) when you studying for Bachelor degree abroads (2) it's very good way for you to study English by yourself because (1) you must speak English every time in your (1) daily life that's will make you develop all your (2) listening speaking reading skill

No. 42	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	3.5	3.5	3.5	3.5
Syntax	3	3	3	3
Vocabulary	4	4	3.5	4
Cohesion	3.5	4	3.5	3.5
Function	4.5	4.5	3	4.5

No. 110 (Advanced)

1. Persuasive task
if you are thinking about where to spend your vacation Thailand would be your answer (1) Thailand's very (1) interesting (1) first (1) there are many tourist attractions in Thailand for example Wat Pra Kaew (2) which is ve… very very beautiful and gorgeous (2) floating market in Ratchaburi (1) JJ market which is perfect place to find some um souvenir for your friends your family (2) um Siam Square where's very fashionable clothe and many thing you can find here (3) and (2) what's more (1) Thailand is very well known for perfect food which very delicious tasty for example Tom Yum Kung papaya salad (2) and there many thing to do like um (3) and Thailand's (2) uh you can travel here in affordable price it's very important thing

2. Imaginary task
good morning everyone today I would like to present my idea about an ideal world (3) my ideal world's um the world with peace and happiness there's would be no war no crime everyone smile to each other and everyone befriend um no one took advantage from (2) other person and (2) besides I I think there shouldn't be a gap between the poor and the rich they should share everything because they are friend (1) what is more disaster is very (1) ver… (3) disaster like tsunami earthquake or storm is very bad thing I hope that there would be no disaster in this world then we can live in a (1) peaceful world which my ideal world and (4) I (1) if we live in a peaceful world I hope that everyone (?) to be happy and

3. Opinion task
well I think studying studying for a bachelor's degree abroad is very good idea for me (1) even it's costs a lot of money and may lack a time in uh with friend in the university in Thailand (1) first of all study studying abroad uh you will make you have a chance to use (1) um language like English which is uh international language and (1) it's uh access to source of knowledge you can um develop yourself (1) in many skill like speaking (1) reading (1) writing (1) and (2) what is more is when you studying abroad you you will be on your own it's inpedent independent so you have to be very responsible and you have to adapt yourself for the new social with many kind of person you you will be how to live on your own and (1) I think it's very (1) it going to be an experience for (1) real life than uh studying in our country

4. Narrative task
oh I miss you so much I haven't seen you for a long time uh (1) the most enjoyable day on my campus is uh freshy day and freshy night about (1) uh (1) three months ago it's it's a party and activity for the freshy um and it's um (2) was happen at my faculty of Law near Samyan you know uh at first we have to uh wear uniform of Chulalongkorn University and we have (1) have have been t… told to be many thing by the senior or the %วากเกอร์% like uh sing sing many songs in a very loud noise and (1) I was very tired I feel ve… very bored with this activity later on I I met friend playing many games which is very fun and (3) very interesting this was a was a very happy time and at about six o' clock we have a party and have a concert which is very fun and I was very impress for this activity I feel so glad to be

No. 110	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	3.5	3	3	3
Syntax	4	4	4	4
Vocabulary	4	4	4	4
Cohesion	4	4	4	4
Function	4	4	4.5	4.5

No. 48 (Advanced)

1. Imaginary task

my imaginary world will be like there no poverty and wars because you know poverty is um (1) very bad thing at the moment people suffer from poverty so much and some people die from the poverty (2) the next thing is the war (1) as you can see from the world war one and world war two people die and die and died and some poverty also cause from the war (3) of course if we live with wars our world won't be any happy (2) and the way in which the world would be a better place for everyone is everyone should be like consider more about (1) the other people's feelings (2) so that we can help each other (1) and the world will be (1) a better place for sure (3) and sometimes we should (1) not be selfish (1) we should give to the others more than we do right now

2. Persuasive task

hello everyone we are now going to (1) uh introduce you about Thailand Thailand know as land of smile and that's true when you come to Thailand you will be welcome warmly for sure (2) people here are so kind and generous (2) Thailand also has many tourist attractions for example the Emerald Buddha temple (2) you can walk and see the Grand Palace too (2) and if you tired from seeing the temple you can visit at the south part of Thailand which is very well known about the beach for example in Krabi (1) you can go and diving uh y… and you can also eat many seafoods (2) if you come to Thailand (3) uh (2) you will see many different things in which your country don't have (1) and that's one of the most a… amazing experience

3. Narrative task

hello my friends we haven't seen each other for a long time how have you been I have been great (1) and recently we have the shareball competition last week at the faculty (2) I am in the shareball team too (1) you know first we practice and practice (1) and then (1) we went to the competition (2) then we won the first round we were so glad and then we went through the competition we won and won and won and that's (1) a great feelings until the final round and we get a bit stred stress because (1) another team is also a tough team too but in the end we won the gold medal (2) and I (1) felt very glad that we won the gold medal (2) and I'm glad that I play the shareball too and I also (1) got some new friends (1) and after we won we went to celebrate (1) and that a great feelings (3) how about you have you had anything great happen to you recently

4. Opinion task

about studying abroad in Bachelor degrees (1) I think there are good things and bad things too (1) like (1) the good things (1) are the (1) if you go to the using English country of course you will get the great English back for sure but if you go to the (1) country which use different language of course you will get a new language and that will help you to find a job easier than the people with only two language (3) another things is the new experience (1) you will have to learn how to adapt yourself in a new place (1) and (2) learn about their culture but the bad things for studying a bachelor degrees abroad is the (1) you will be far from home sometime you will be miss your family so much and miss your old friends (2) and if you are in the abroad (1) at the early age you had not many friends in Thailand and the h… pri…

No. 48	Narrative	Opinion	Imaginary	Persuasive
Pronunciation	4	4	4	4
Syntax	4	4	3.5	4
Vocabulary	4	4	4	4
Cohesion	4	4	4	4
Function	4.5	4	3	4.5

Appendix G

Estimated variance and covariance components and generalizability coefficients

Appendix G1: Estimated variance and covariance components and generalizability coefficients for opinion task scores (raters = 2)

Source of variation		Pro	Syn	Voc	Coh	Func
Person	Pro	**.183** (68.6%)	.693	.613	.582	.344
	Syn	.125	**.178** (73.7%)	.7	.759	.381
	Voc	.112	.126	**.183** (73.8%)	.745	.423
	Coh	.084	.108	.107	**.114** (58.9%)	.488
	Func	.068	.074	.083	.075	**.21** (88.6%)
Rater	Pro	**.028** (10.4%)				
	Syn	.014	**.006** (2.7%)			
	Voc	.02	.01	**.013** (5.4%)		
	Coh	.025	.012	.018	**.023** (11.8%)	
	Func	-.002	-.001	-.001	-.001	**.000** (0%)
PR	Pro	**.056** (21%)				
	Syn	.004	**.057** (23.6%)			
	Voc	.01	-.003	**.051** (20.8%)		
	Coh	.006	.005	.016	**.056** (29.3%)	
	Func	.003	.000	-.002	-.003	**.027** (11.4%)
Generalizability coefficient		.77	.76	.78	.67	.89

Appendix G2: Estimated variance and covariance components and generalizability coefficients for imaginary task scores (raters = 2)

Source of variation		Pro	Syn	Voc	Coh	Func
Person	Pro	**.178** (67.6%)	.593	.659	.485	.149
	Syn	.074	**.088** (69.8%)	.629	.592	.198
	Voc	.136	.09	**.238** (73.5%)	.76	.438
	Coh	.088	.075	.159	**.184** (63.1%)	.474
	Func	.026	.024	.088	.084	**.172** (80.7%)
Rater	Pro	**.027** (10.1%)				
	Syn	.013	**.006** (4.4%)			
	Voc	.026	.012	**.025** (7.6%)		
	Coh	.033	.015	.032	**.04** (13.9%)	
	Func	.007	.003	.007	.009	**.002** (.8%)
PR	Pro	**.059** (22.3%)				
	Syn	.000	**.032** (25.8%)			
	Voc	.01	.002	**.06** (18.8%)		
	Coh	.001	.002	.019	**.067** (23%)	
	Func	.002	.006	.004	.004	**.039** (18.5%)
Generalizability coefficient		.75	.73	.80	.73	.81

Appendix G3: Estimated variance and covariance components and generalizability coefficients for persuasive task scores (raters = 2)

Source of variation		Pro	Syn	Voc	Coh	Func
Person	Pro	**.168** (65.9%)	.567	.613	.513	.38
	Syn	.1	**.187** (76.6%)	.638	.662	.512
	Voc	.106	.117	**.179** (77.9%)	.7	.47
	Coh	.07	.095	.098	**.11** (61.3%)	.436
	Func	.079	.112	.101	.074	**.258** (82.8%)
Rater	Pro	**.024** (9.3%)				
	Syn	.015	**.009** (3.6%)			
	Voc	.012	.007	**.005** (2.4%)		
	Coh	.02	.013	.01	**.017** (9.6%)	
	Func	.006	.004	.003	.005	**.001** (.36%)
PR	Pro	**.063** (24.8%)				
	Syn	.006	**.048** (19.8%)			
	Voc	.009	.001	**.045** (19.7%)		
	Coh	.005	.006	.011	**.052** (29.1%)	
	Func	-.003	-.005	-.001	0	**.053** (16.9%)
Generalizability coefficient		.73	.79	.80	.68	.83

Appendix H

Coding schemes for the verbal reports

The coding schemes for the verbal reports consists of several taxonomies which are based on the literature in the model of language ability by Bachman and Palmer (1996), and the research on communication strategies by, for example, Færch and Kasper (1983), Cohen and Olshtain (1993) and Dornyei (1995). The taxonomies in **bold** or regular are those found in the literature, while those in *italic* are found in my pilot study as well as in the preliminary study of a few data in my main study. Some of the categories found in the studies are grouped as subtypes of those in the literature. The last section, other test-taking behaviors, was additional data which might belong to the processes and strategies categories.

Taxonomies of speaking test taking strategies and processes	Definition/description
Goal setting	Deciding what one is going to do
1. Identifying the test tasks	
2. Choosing one or more tasks to do when given a choice	
3. Deciding whether or not to attempt to complete the task(s)	
- *Deciding whether to continue to respond to the task*	
Assessment	Taking stock of what is needed, what one has to work with, and how well one has done

1. Assessing the characteristics of the test task	Identifying the characteristics of the test task to determine
	   - the desirability and feasibility of successfully completing the task, and
	   - what elements of topical

		knowledge and language knowledge likely to require the test taker to do
-	*Identifying the meaning of vocabulary in the prompt*	
2.	Assessing one's own topical knowledge	Determining the degree of relevant topical knowledge that is available, and if available, which might be used. Also, determining one's affective schemata for coping with the task requirement.
3.	Assessing one's own language knowledge	Determining the degree of relevant language knowledge that is available, and if available, which might be used. Also, determining one's affective schemata for coping with the task requirement.
4.	Assessing the correctness or appropriateness of the response to the test task	The areas to be assessed include the grammatical, textual, functional and sociolinguistic features of the response and its topical content.
-	Monitoring grammar	
-	Monitoring pronunciation	
-	Afterthoughts	Thinking to oneself afterwards that one could have said something else
-	*Commenting on topical knowledge*	Making comments on the topical knowledge that has been used earlier
Planning		Deciding how to use what one has
1.	Selecting specific elements from topical knowledge that will be used in a plan	

- *Drawing on topical knowledge*	*Using real world knowledge as the information which may be expressed in a response to the test task*
2. Selecting specific elements from language knowledge that will be used in a plan	These elements include syntax, pronunciation, vocabulary, cohesion, organization, function and socio linguistic features.
- *Drawing on grammatical knowledge that will be applied in the response*	Using relevant grammatical knowledge that one has
- *Selecting the grammatical features which will be used in the response*	
- Self-debate before selection	Debating between two or among several expressions before making a decision of which to be used
3. Formulating one or more plans	
▪ *Content plans*	
- *Planning the content*	
- *Elaborating the content*	Describing the content used earlier in a more detailed way
- Topic avoidance	Avoiding topics or concepts
- Message abandonment	Leaving a message unfinished
▪ *Language plans*	
- Message formulation	Putting a thought into English
- *Making more than one plan to express the intended meaning*	
- *Planning the utterance in chunks*	Planning the utterance in large units, not word by word
- *Planning the organization of the response*	
- *Planning the introduction of the*	

response	
- *Planning the order of the constituents in the utterances*	
- *Combining utterances*	
- Code switching	Switching from L2 to L1 or another foreign language
- Paraphrase	Using descriptions, circumlocutions or exemplification
- Approximation	Using a more general word for an unknown word e.g. ship for sailboat
- Word coinage	Construction of a new interlanguage word
- Directly translating Thai words into English	Using an English word that is direct translation of the Thai word. Although the English word exists, its meaning is different from the actual intended meaning.
- Using prefabricated patterns/ formulaic speech	Using memorized stock phrases
- Foreignizing	Using a L1 word by adjusting it o L2 phonology and/or morphology
- *Using vocabulary of the prompt*	Taking words or expressions directly from the task instructions to use in the response
- Formal avoidance	Avoiding using rules/items in phonology, morphology, syntax and vocabulary.
4. Selecting one plan	
Other test-taking behaviors	
- *Thinking in Thai and/or English*	
- *Reading/re reading the prompt*	
- *Writing the note in Thai/English*	

- *Reading out the note*	
- *Language forms coming to mind*	Language forms coming to mind of the test taker without preplanning or thinking about it

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix I
Transcription Notations

The transcription notations are as follows:

1. Unfilled pauses: pauses equal and longer than one second are timed and written in parentheses e.g. (1) = 1 second.

2. Transcription doubt: when the transcription is not clear, a question mark within parentheses (?) is used.

3. Thai words: when the speaker spoke Thai, per cent signs (% %) with Thai fonts in between are used, e.g. %วากเกอร์%

4. incomplete words: when words are pronounced with only some initial sound(s), the corresponding letter(s) is (are) written followed by three dots, e.g. parti… participated

Appendix J

Procedures for discourse analyses

Steps in doing the discourse analyses

Genre Analysis

Steps in doing genre analysis (adopted from Eggins and Slade, 1997):

1. Divide each test response into different stages, each of which consists of an utterance that has a particular function relating to the whole text.

2. Assign each stage a functional label. The labels which describe a genre found in the test responses are presented below.

3. Identify obligatory and optional stages. An obligatory stage is a key feature that defines the genre, while an optional stage is not and can occur across genres.

Stages in each genre:

1. Recount genre (Narrative task)

Stages	Functions
(Abstract)	Gives the point of the story.
(Orientation)	Orients the audience regarding time, place, actions and people in the story.
Record of events	Retells a sequence of events. In this stage, the speaker's evaluative comments of the events may be presented throughout the text.
(Coda)	Evaluates the whole story and returns to the present.

Note: An optional stage is written in parentheses.

2. Exposition (Martin, 1989): Opinion and Persuasive tasks

Stages	Functions
Thesis	Presents the speaker's judgment toward an issue.
Argument	Provides support to the judgment.

3. Description (Martin, 1989): Imaginary task

Stages	Functions
Description	Describes the characteristics of the topic given.

Functional analysis

   Steps used to do functional analysis are to divide the text into different moves and assign each move a speech function. As suggested by Eggins and Slade (1997), the following units are considered a single move:

1. a grammatically independent clause
2. two independent clauses produced as run ons (without any rhythmic or intonation break between the clauses)
3. a subordinate clause consisting of a dependent clause followed by its main clause
4. a subordinate clause consisting of a main clause followed by its dependent clause produced as run ons

  (Note: in the case of a subordinate clause in which the main clause preceding its dependent clause and both clauses are not produced as run ons, they are considered separate moves)

5. a clause with an embedded clause (e.g. post modifier or noun clause)
6. a clause followed by quoting or reporting clause (both direct and indirect)

Functions that may appear in all tasks

Function	Definition	Example (in bold)
Greeting		**Hello how are you**
Stating the purpose of the talk		**Today I would like to present my idea about an ideal world**
Emphasizing one's point		and Thailand's (2) uh you can travel here in affordable price **it's very important thing**
Elaborating	Clarifying, restating, exemplifying a prior move, adding further details or contrasting information to a prior move; qualifying or modifying a prior move by giving causal or conditional details	Thailand's very (1) interesting (1) **first (1) there are many tourist attractions in Thailand for example Wat Pra Kaew (2) which is ve very very beautiful and gorgeous (2)** • and (2) besides I I think there shouldn't be a gap between the poor and the rich they should share everything **because they are friend** • and (2) what is more is when you studying abroad you you will be on your own it's inpendent independent **so you have to be very responsible**
Expressing feelings		**this was a was a very happy time**
Asking a question		**how about you have you had anything great happen to you recently**

Expressing wish		if we live in a peaceful world **I hope that everyone (?)to be happy and**
Concluding		and you have to adapt yourself for the new social with many kind of person you will be how to live on your own **and (1) I think it's very (1) it going to be an experience for (1) real life than uh studying in our country**
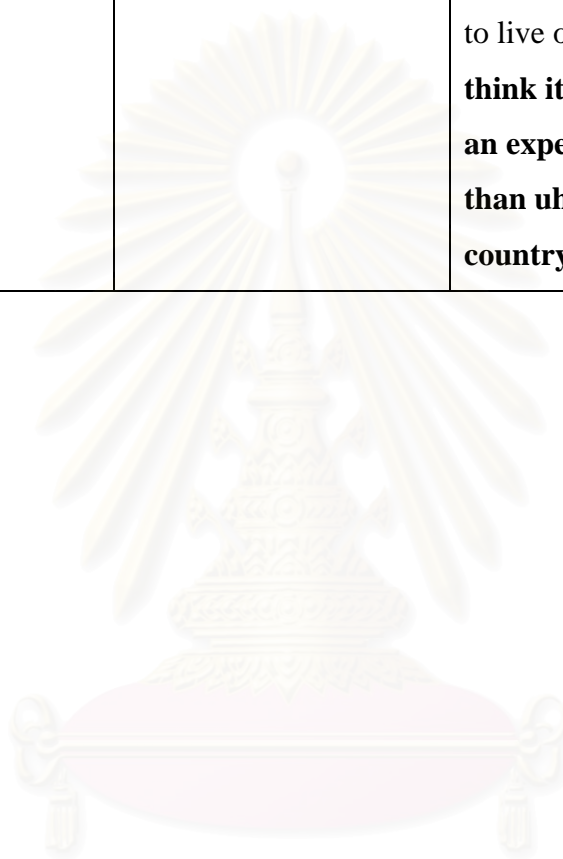
Functions in the narrative task

Function	Definition	Example
Setting temporal context	Describing when the story happened	the most enjoyable day on my campus is uh **freshy day and freshy night about (1) uh three months ago**
Setting physical context	Describing where the story happened	**um and it's um (2) was happen at my faculty of Law near Samyan you know**
Introducing characters		it's it's a party and activity **for the freshy**
Describing	• characters • the situation	• uh at first **we have to uh wear uniform of Chulalongkorn University** • **it's it's a party and activity for the freshy**
narrating	Telling the story	**and we have (1) have have been told to be many thing by the senior or the % ว้ากเกอร์ % like uh sing sing many songs in a very loud noise and**
Justifying		I think it's a good thing to have senior **because it's make me uh unity**

Functions in the imaginary task

Function	Definition	Example
Predicting		**In the future I think our world will have many environment and lacking of natural source**
Describing		**my ideal world's um the world with peace and happiness**
Recommending		**We should give to the others more than we right now**
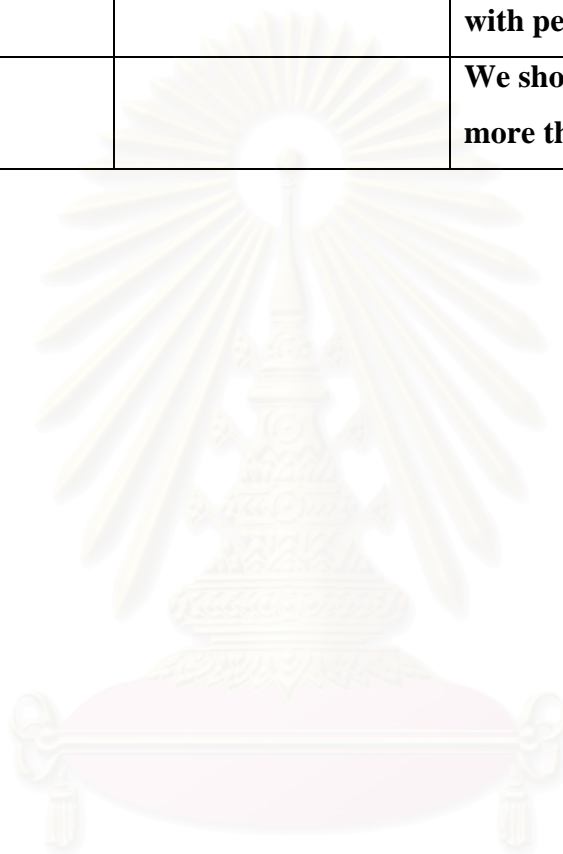
Functions in the opinion and persuasive tasks

Function	Definition	Example
Recommending		**If you are thinking about where to spend your vacation Thailand would be your answer**
Persuading		**If you come to Thailand you will see many different things in which your country don't have**
Justifying	Showing or proving something to be reasonable • Justifying a prior point stated explicitly • Justifying the point that may not be stated explicitly	• Please See Example "Justifying A" below. • Example "Justifying B"

Example

Justifying A	The speaker justifies their point that has been stated explicitly. For example, "if you are thinking about were to spend you vacation Thailand would be your answer/ **Thailand's very interesting**". The first move "if you…answer" is an explicit point the speaker would like to make. The second move, **"Thailand's very interesting"**, is a justification of the point.

Justifying B	The speaker may justify their point without stating it explicitly. For example, in the persuasive task a student began her answer by saying that "Thailand has many good place to visit", without saying explicitly that the travelers should visit the country. Since the test task requires the examinees to give reasons why foreigners should come to Thailand, this move then can be categorized as justification to the point "why foreigners should visit Thailand".

Analysis of grammatical elements

The grammatical elements found in the test responses which can be analyzed may include verb tenses, modals and conjunctions. Examples of these elements are:

Grammatical elements	Examples
Verb tenses	present tense, past tense, etc.
Modals	would, may, etc.
Conjunctions (Martin, 1992)	- additive: addition (e.g. and), alteration (or) - comparative: similarity (likewise), contrast (but) - temporal: simultaneous (while), successive (then, at first) - consequential: purpose (so that), condition (if), consequence (because), concession (although), manner (in this way)

Appendix K

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ProNar	.234	158	.000	.877	158	.000
ProOp	.231	158	.000	.877	158	.000
ProIm	.230	158	.000	.878	158	.000
ProPer	.226	158	.000	.877	158	.000
SynNar	.323	158	.000	.825	158	.000
SynOp	.297	158	.000	.839	158	.000
SynIm	.407	158	.000	.675	158	.000
SynPer	.320	158	.000	.822	158	.000
VocNar	.349	158	.000	.732	158	.000
VocOp	.312	158	.000	.777	158	.000
VocIm	.235	158	.000	.846	158	.000
VocPer	.343	158	.000	.780	158	.000
CohNar	.350	158	.000	.687	158	.000
CohOp	.325	158	.000	.768	158	.000
CohIm	.233	158	.000	.807	158	.000
CohPer	.339	158	.000	.746	158	.000
FuncNar	.282	158	.000	.847	158	.000
FuncOp	.364	158	.000	.695	158	.000
FuncIm	.367	158	.000	.767	158	.000
FuncPer	.344	158	.000	.819	158	.000

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix L

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
ProNar	1.662	3	136	.178
ProOp	1.639	3	136	.183
ProIm	1.234	3	136	.300
ProPer	.882	3	136	.452
SynNar	1.132	3	136	.338
SynOp	1.028	3	136	.382
SynIm	2.635	3	136	.052
SynPer	1.457	3	136	.229
VocNar	1.011	3	136	.390
VocOp	1.224	3	136	.304
VocIm	2.548	3	136	.058
VocPer	.430	3	136	.732
CohNar	3.604	3	136	**.015**
CohOp	.413	3	136	.744
CohIm	.565	3	136	.639
CohPer	.682	3	136	.564
FuncNar	1.471	3	136	.225
FuncOp	.948	3	136	.419
FuncIm	8.472	3	136	**.000**
FuncPer	.130	3	136	.942

Appendix M

Descriptive statistics for groups 1-4

Group 1 (N = 36)

Descriptive Statistics

	Mean	Std. Deviation
ProNar	3.3472	.50454
ProOp	3.3472	.50454
ProIm	3.3194	.49501
ProPer	3.3611	.47225
SynNar	3.1806	.38084
SynOp	3.2778	.40434
SynIm	3.0139	.25315
SynPer	3.2639	.43893
VocNar	3.6667	.41404
VocOp	3.6250	.45316
VocIm	3.6111	.46462
VocPer	3.7500	.42258
CohNar	3.6667	.41404
CohOp	3.7500	.36839
CohIm	3.5556	.45947
CohPer	3.6667	.39641
FuncNar	3.5556	.57044
FuncOp	4.0417	.40311
FuncIm	3.1528	.42794
FuncPer	3.8750	.56537

Group 2 (N = 33)

Descriptive Statistics

	Mean	Std. Deviation
ProNar	3.4242	.48608
ProOp	3.4091	.49140
ProIm	3.4242	.48608
ProPer	3.4242	.48608
SynNar	3.1970	.37374
SynOp	3.4394	.46364
SynIm	3.1061	.32494
SynPer	3.3182	.46466
VocNar	3.7273	.41629
VocOp	3.7273	.37689
VocIm	3.7424	.35622
VocPer	3.7727	.39708
CohNar	3.8485	.26472
CohOp	3.7879	.35422
CohIm	3.6364	.35953
CohPer	3.7576	.35622
FuncNar	3.7879	.55944
FuncOp	4.0303	.55817
FuncIm	3.1818	.52764
FuncPer	3.9394	.52675

Group 3 (N = 39)

Descriptive Statistics

	Mean	Std. Deviation
ProNar	3.4615	.41966
ProOp	3.4744	.42841
ProIm	3.4744	.42841
ProPer	3.4744	.42841
SynNar	3.1538	.38309
SynOp	3.2692	.47081
SynIm	3.1026	.36598
SynPer	3.1923	.43853
VocNar	3.7436	.41154
VocOp	3.7308	.39480
VocIm	3.6026	.38353
VocPer	3.7308	.42683
CohNar	3.8077	.31655
CohOp	3.7692	.32131
CohIm	3.5769	.38964
CohPer	3.7436	.34166
FuncNar	3.8205	.50637
FuncOp	4.1154	.45104
FuncIm	3.0385	.26618
FuncPer	3.8077	.46757

Group 4 (N = 32)

Descriptive Statistics

	Mean	Std. Deviation
ProNar	3.3906	.45320
ProOp	3.4063	.44789
ProIm	3.4063	.44789
ProPer	3.4063	.44789
SynNar	3.3438	.44789
SynOp	3.3125	.47093
SynIm	3.0313	.30946
SynPer	3.4219	.52532
VocNar	3.6563	.51490
VocOp	3.6406	.55698
VocIm	3.5469	.55879
VocPer	3.7500	.47519
CohNar	3.8125	.39656
CohOp	3.7656	.38067
CohIm	3.6719	.48542
CohPer	3.7656	.35886
FuncNar	3.8906	.53483
FuncOp	4.2188	.45680
FuncIm	3.0625	.37567
FuncPer	3.8750	.56796

Biography

Ms. Sutthirak Sapsirin received her B.A. in English from the Faculty of Arts, Chulalongkorn University in 1995 and M.A. in TESOL from University of Kansas, USA in 1999. She is currently an English language instructor at Chulalongkorn University Language Institute. Her research interests include oral language assessment and instruction.