Relationships Between Productive Vocabulary Knowledge and Speaking Performance of Japanese Learners of English at the Novice Level

A Dissertation

Submitted to the University of Tsukuba

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Linguistics

Rie KOIZUMI

Abstract

Vocabulary is one of the essential and fundamental components of communication (e.g., Levelt, 1993). However, few studies have focused specifically and systematically on associations between vocabulary and speaking. The current study examines relationships between productive vocabulary knowledge and speaking performance. Productive vocabulary knowledge is the knowledge to produce a word when one speaks and writes (e.g., Schmitt, 2000). Vocabulary knowledge is further divided into size and depth. Size refers to how many words are known with a primary meaning, whereas depth refers to the degree to which word aspects other than the primary meaning (e.g., associations and affixes) are known. This study examines how size and depth of productive vocabulary knowledge are related to speaking performance (i.e., fluency, accuracy, syntactic complexity, and lexical complexity). Exploring the importance of productive vocabulary knowledge on speaking performance is essential for both theoretical and practical purposes. The participants in this study are Japanese learners of English at the novice level (i.e., third-year junior and first-year to third-year senior high school students in Japan).

In order to examine the focus of this study, two studies were conducted. Study 1 (1C) was conducted as a pilot study of Study 2 (2E) and examines relationships between productive vocabulary knowledge and speaking performance with a specific focus on “size” of productive vocabulary knowledge. Study 2E was modified based on problems encountered in Study 1C. Study 2E investigates relationships between “size and depth” of productive vocabulary knowledge and speaking performance. In Study 2E, two tests were mainly used: (a) the Productive Vocabulary Knowledge Test and (b) the Speaking Test. The former consisted of two sections: the Size Section and the

Depth Section, which had three subsections: the Derivation Subsection, Antonym Subsection, and Collocation Subsection. The Speaking Test used a tape-mediated format, which included five tasks. Multiple regression analysis and the relative Pratt index were mainly used for analysis.

The current study mainly demonstrates the following two points. First, there is a moderate or strong tendency in some tasks (or when the content of the talk is already decided and when few formulaic expressions are used) for Japanese learners of English at the novice level who have larger and deeper productive vocabulary knowledge (than others of this level) to produce a greater number of tokens and types and better speaking performance mainly related to vocabulary when they produce monologic description and comparison without pre-task planning time. Speaking performance mainly related to vocabulary includes one element of fluency (speaking speed), accuracy in a task, one element of syntactic complexity, and one element of lexical complexity. Second, size of productive vocabulary knowledge is related more to accuracy in a task, and at least one aspect of depth (antonyms, or word connections) is related more to one element of fluency (speaking speed) when Japanese learners of English at the novice level produce monologic description and comparison without pre-task planning time. Some pedagogical and methodological implications for language teaching and testing are also addressed.

Acknowledgments

I am deeply indebted to many people who assisted me in pursuing my studies. First and foremost, I owe more than I could ever express to Akihiko Mochizuki for always caring deeply about my progress, for giving me weekly opportunities to make presentations, and for helping me rethink my research through his perceptive questions. He also spent his valuable time reading my entire thesis and giving me detailed and constructive ideas for improvement.

I would like to express my heartfelt gratitude to the following teachers:

Yuji Ushiro for his keen advice on my research design, analysis, and interpretation of the results.

Hirosada Iwasaki for his useful suggestions and encouragement while I was conducting the research and writing the thesis.

Akira Kubota for his critical comments, which helped me improve the quality of the dissertation.

Akiyo Hirai for her invaluable and thoughtful questions concerning interpretation of the results.

Chikako Nishigaki for her insightful suggestions regarding the overall writing quality.

I owe a debt of gratitude to many junior and senior high school teachers and their students who cooperated with me in this study, especially to Atsuko Kosuge, Tsunehisa Isaji, Masashi Kubono, Takao Hachimiya, Akemi Ueki, Ryoichi Yamamoto, Kenichi Sano, Tomohiro Sakurai, Koji Uenishi, Yoshiyuki Tsukahara, Tomoyuki Kawashima, Hitoshi Takiguchi, Kiyonori Shibayama, Chisato Saida, Itsumi Kino, Yoshinari Hirosawa, Fumiyasu Orihara, Akiko Nemoto, and Akiko Takeyama.

I am pleased to acknowledge the support of the Society for Testing English Proficiency (STEP) Inc. and the Japan Language Testing Association (JLTA) for giving me a generous grant. My profound appreciation goes out to Kenji Ohtomo for his helpful advice.

Many thanks are also due to my family and friends, who helped me to survive the demands of academic life, and especially to Yo In'nami, who constantly gave me warm encouragement and critical advice for improvement, and also my parents, who always believed in me. Moreover, I greatly appreciate the efforts of the many people who participated in the test-construction, interviewing, and rating sessions: Adam Jon Lebowitz, George Robert Maclean, William J. Bonk, J Lake, David Klapper, Yo In'nami, Maki Shimizu, Tsutomu Koga, Yuko Morimoto, Chikako Nakagawa, Yuko Hijikata, Aki Ohkubo, Masatoshi Nagahashi, Nahoko Inoue, Masashi Naya, Kozue Kikuchi, Nanako Kobayashi, Yoshiko Kitada, Takami Fukuzawa, Miki Hirose, Eriko Ishikawa, Misa Koide, Michiru Hanabusa, Asami Nishimagi, Tomoko Nagaoka, Rintaro Sato, and Akiko Okazaki.

Last but not least, my special gratitude is also extended to Hiroo Arikawa for offering me an enormous scholarship. His continuous financial and intellectual support enabled me to commit myself to pursuing my academic interest and forming it into the current dissertation.

For further inquiries concerning this dissertation, send an e-mail message to (Ms.) Rie Koizumi at rie-koizumi@mwa.biglobe.ne.jp or koizumirie@hotmail.com

Rie Koizumi

Tsukuba, Japan

December, 2005

Table of Contents

Part Ⅰ Background of the Study

Chapter 1 Introduction

1.1 Context of the Problem

An essential goal in English language teaching is to enhance communicative language ability, which is the ability that enables us to express ourselves and convey messages in English. The Course of Study (Ministry of Education, Science, and Culture, 1999a, 1999b), which is a national guideline for teachers, emphasizes the fostering of communicative language ability, including speaking ability. The Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2003) also proposed goals of teaching "English language abilities required for all Japanese nationals" in "an Action Plan to Cultivate 'Japanese with English Abilities.'" The goal for junior-high-school graduates is that they "can conduct basic communication with regard to areas such as greetings, responses, or topics relating to daily life," while the goal of senior-high-school graduates is that they "can conduct normal communication with regard to topics, for example, relating to daily life." Thus, enhancing the ability for successful communication is one of the main targets.

Vocabulary has long been regarded as a vital component and basics of (communicative) language ability[1.1] (e.g., Bachman & Palmer, 1996, p. 68; Carroll, 1968, pp. 54-55; Meara, 1996, p. 35, 37). It is conceived of as a good indicator of second language (L2) proficiency, which is supported by empirical investigation (Katagiri, 2001, e.g., p. 45; Meara & Buxton, 1987, p. 150; Read, 2000, p. 127). In addition, vocabulary is one of the essential and fundamental components of communication (e.g., Levelt, 1993; Nakamura, 2004, p. 123; Singleton, 1997, p. 213). Previous studies have demonstrated that it has connections with reading (e.g., Akase, 2005; Alderson, 2000; Alderson & Huhta, 2005; Eskey, 2005; Katagiri, 2001; Laufer,

1992; Noro, 2002; Qian, 1999, 2002; Shimamoto, 1988; Shiotsu, 2003), listening (e.g., Alderson & Huhta, 2005; Bonk, 2000; Katagiri, 2001; Rost, 2005; Yoshimura, Imai, Hiraiwa, Fuse, Edasawa, & Mine, 2005), writing (e.g., Arnaud, 1992; Baba, 2004, 2005; Ishikawa, 2004; Laufer, 1998; Laufer & Nation, 1995; Laufer & Paribakht, 1998; Linnarud, 1986; Schoonen et al., 2003; Yoda, 2004), and speaking (e.g., Adams, 1980; Adolphs & Schmitt, 2003, 2004; Durán, Malvern, Richards, & Chipere, 2004; Higgs & Clifford, 1982; Iwashita, 2005; Kaneko, 2004; Kimura, 2004; Koizumi & Kurizaki, 2002; Levelt, 1993; Read, 2005; Takiguchi, 2003; see section 2.4). Although there have been some studies examining relationships between vocabulary and skills, few studies have focused specifically and systematically on associations between vocabulary and speaking.

The reasons for the lack of a systematic examination of relationships between vocabulary and speaking may be the following. First, the prevalence of vocabulary studies is a relatively recent phenomenon. According to Meara (2002), studies on vocabulary have dramatically increased over the past 20 years. This change may be attributed to the re-recognition of the roles that vocabulary plays in communication, acquisition, and learning (Kadota, 2003, pp. 2-10). Second, speaking studies tend to face difficulties in collecting and analyzing spoken data (Fulcher, 1997, p. 81). Nevertheless, the fact that there have been studies that investigated relationships between vocabulary and reading, especially between vocabulary size and depth and reading (e.g., Akase, 2005; Noro, 2002; Qian, 1999, 2002), suggests that examination of relationships between vocabulary and speaking is an essential research target, despite its difficulty.

Speaking can be seen from two perspectives: performance and ability. Performance is defined as “actual instances of language use in real time” (McNamara,

1996, p. 54), which is something that can be observed. On the other hand, ability is an underlying entity that is relatively stable and can be inferred from performance. Although there are many models of language ability (e.g., Bachman, 1990; Bachman & Palmer, 1996; Canale, 1983; Canale & Swain, 1980; Davies, 1989; Hymes, 1972; Taylor, 1988), they all seem to agree with this view. Since speaking performance is affected by many factors other than speaking ability (e.g., affective factors and tasks[1.2]; Bachman & Palmer, 1996), performance and ability can be considered different entities. Although models of speaking performance have been proposed (e.g., Fulcher, 2003; Skehan, 1998; see section 2.1.1), relationships between speaking performance and speaking ability have not been elucidated, probably because, as Fulcher (1997) suggests, an empirical model of speaking ability itself has not yet been established (p. 82). Therefore, it seems rather difficult to investigate speaking ability and its relationships with vocabulary, although a final goal in researching, teaching and testing is to deal with speaking ability. For this reason, the current study does not deal with speaking ability but with speaking performance, which can be tackled more easily and more realistically. That is why this study investigates relationships between vocabulary and speaking performance.

It should be noted that in a theoretical framework of language ability, vocabulary knowledge is one component of the language ability (Bachman & Palmer, 1996, pp. 63-70; Carroll, 1968, pp. 54-55). Because of this, examination of relationships between vocabulary knowledge and speaking performance can be considered to be a study that deals with relationships between part of an aspect of speaking ability and speaking performance, which may eventually contribute to a model of relationships between the two. At this point, some may wonder if vocabulary knowledge can be examined as a research target since this too is an unobservable underlying entity.[1.3]

However, this study takes the stance that such an examination is possible because structures of vocabulary knowledge and its assessment methods have been examined to some degree (see section 2.2.1).

Thus, the current study examines relationships between vocabulary knowledge and speaking performance. Vocabulary knowledge can be further divided into receptive and productive vocabulary knowledge (e.g., Nation, 2001, pp. 24-26, 2005, pp. 584-585; Read, 2000, p. 154; Schmitt, 2000, p. 4). Receptive vocabulary knowledge is the knowledge to understand a word, which is often used in listening and reading, whereas productive vocabulary knowledge is the knowledge to produce a word[1.4] when one speaks and writes (Nation, 2005, p. 585; Schmitt, 2000, p. 4). Another classification of vocabulary knowledge is size and depth. Size refers to the number of words with a primary meaning that a learner has (Nakanishi & Shimamoto, 2003), whereas depth refers to the degree to which a learner knows a certain word in addition to knowing a primary meaning (e.g., associations and affixes; Nakanishi & Shimamoto, 2003). This study examines how size and depth of productive vocabulary knowledge were related to speaking performance (i.e., fluency, accuracy, syntactic complexity, and lexical complexity). Exploring the importance of productive vocabulary knowledge on speaking performance is essential for both theoretical and practical purposes. The participants in this study are Japanese learners of English at the novice level (i.e., third-year junior and first-year to third-year senior high school students in Japan) because studies to deal with novice learners have been scarce and previous studies (Adams, 1980; Higgs & Clifford, 1982) suggest that productive vocabulary knowledge plays a more important role in speaking performance among novice level learners than among intermediate or advanced learners.

1.2 Organization of the Thesis

The plan of this dissertation is as follows. There were four parts: Part Ⅰ (Background of the Study), Part Ⅱ (Study 1: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance), Part Ⅲ (Study 2: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance), and Part Ⅳ (Conclusion).

Part Ⅰ (Background of the Study) consists of two chapters: Chapter 1 (introduction; this chapter) and Chapter 2. Chapter 2 reviews previous studies on (a) factors that affect speaking performance, (b) productive vocabulary knowledge, (c) speaking performance, (d) relationships between vocabulary and speaking performance, and (e) validity and validation. The last section of Chapter 2 explains the purpose and a research question, significance and characteristics of the study, definition of key terms, and a summary of procedures regarding the analysis of the data used in this study.

Part Ⅱ (Study 1: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance) is made up of four chapters: Chapters 3 to 6. Of the four chapters, Chapter 6 is the main focus and explored relationships between "size" of productive vocabulary knowledge and speaking performance. Chapter 3 gives an overview of Study 1 and describes the method used in Chapters 4 to 6. Chapters 4 and 5 examines the validity of inferences based on scores of the Productive Vocabulary Knowledge Test (Productive VKT) and values of the Speaking Performance Measures, which are used in examining the main focus (Chapter 6). Study 1 is conducted as a pilot study of Study 2.

Part Ⅲ (Study 2: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance) is composed of six chapters:

Chapters 7 to 12. Chapter 12 represents the main focus and investigates relationships between "size and depth" of productive vocabulary knowledge and speaking performance. Chapter 7 describes the instruments and procedures utilized in Chapters 8 to 12. Chapter 8 examines appropriate scoring methods of the Productive VKT to be employed in Chapters 9 and 12. Chapter 9 reports the validity of inferences based on the scores of the Productive VKT including the Size and Depth Sections. Chapter 10 selects appropriate measures and criteria of the Speaking Performance Measures to be used in Chapters 11 and 12. Based on the findings in Chapters 8 to 11, Chapters 12 investigates the main focus in order to achieve the purpose of this study.

Part Ⅳ (Conclusion) had one chapter: Chapter 13. This summarizes the findings in the conclusion section, and discusses the pedagogical and methodological implications for language teaching and testing. Finally, limitations and areas for future research are discussed, and this chapter ends with concluding remarks.

Table 1.1 illustrates the organization of this dissertation. The first column shows the overall structures, whereas the third column demonstrates the focus of the current dissertation. The second column includes studies that examine qualities of instruments and scoring methods used in Chapters 6 and 12 in order to make the main analysis more meaningful. Figure 1.1 shows the structures of Studies 1 (Part Ⅱ) and 2 (Part Ⅲ) and where each chapter is located in the overall research. Chapters 3 to 5 provide the basis for Chapter 6, whereas Chapter 6 and Chapters 7 to 11 provide the basis for Chapter 12.

Table 1.1
Organization of the Dissertation

Part	Examination of qualities of instruments and scoring methods	Focus of the study
Ⅰ Background of the Study		Chapter 1 Introduction Chapter 2 Literature Review
Ⅱ Study 1: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance	Chapter 3 Overview of Study 1 Chapter 4 Study 1A: Validation of the Productive Vocabulary Knowledge Test Chapter 5 Study 1B: Validation of the Speaking Performance Measures	Chapter 6 Study 1C: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance
Ⅲ Study 2: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance	Chapter 7 Overview of Study 2 Chapter 8 Study 2A: Scoring Methods of the Productive Vocabulary Knowledge Test Chapter 9 Study 2B: Validation of the Productive Vocabulary Knowledge Test Chapter 10 Study 2C: Selecting the Speaking Performance Measures Chapter 11 Study 2D: Validation of the Speaking Performance Measures	Chapter 12 Study 2E: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance
Ⅳ Conclusion		Chapter 13 Conclusion

Note. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

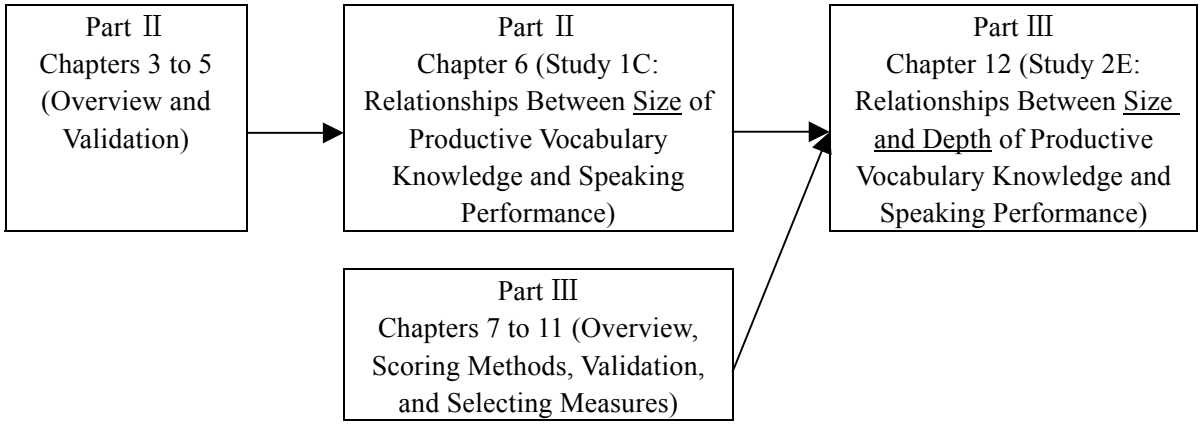


Figure 1.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

1.3 Summary of Chapter 1

Vocabulary has long been regarded as a vital component and basics of

language ability and communication. However, there are only a small number of studies examining relationships between vocabulary and speaking.

Speaking can be seen from two perspectives: performance and ability. This study focuses on performance and examined relationships between size and depth of productive vocabulary knowledge and speaking performance. The participants in this study are Japanese learners of English at the novice level (i.e., third-year junior and first-year to third-year senior high school students).

Chapter 2 Literature Review[2.1]

2.1 Factors That Affect Speaking Performance

Chapter 2 reviews previous studies on (a) factors that affect speaking performance, (b) productive vocabulary knowledge, (c) speaking performance, (d) relationships between vocabulary and speaking performance, and (e) validity and validation. This section describes performance models, which are used as a basis in this study, and language ability.

2.1.1 Performance Models

Performance refers to "actual instances of language use in real time" (McNamara, 1996, p. 54). Performance is affected by many factors and is produced through complex interactions of the factors (Bachman & Palmer, 1996). In order to locate the focus of this study among the various factors, a general model of language performance (Bachman & Palmer, 1996) is reviewed first.

Bachman and Palmer's (1996) model is a modified version of Bachman's (1990) model (see McNamara, 1996, for differences in the two models). According to Bachman and Palmer (1996), performance is affected by two major factors and is produced through interactions of the two. The factors are (a) language users or test takers, and (b) language use tasks and test tasks (p. 61). Between (a) and (b), and even among (a) themselves, there are "complex and multiple interactions" (p. 62; see Figure 2.1).

The first major factor (i.e., language users or test takers) consists of the following four individual characteristics: language ability, personal characteristics, topical knowledge, and affective schemata (p. 64). First, language ability has two components: language knowledge and strategic competence, which are reviewed in

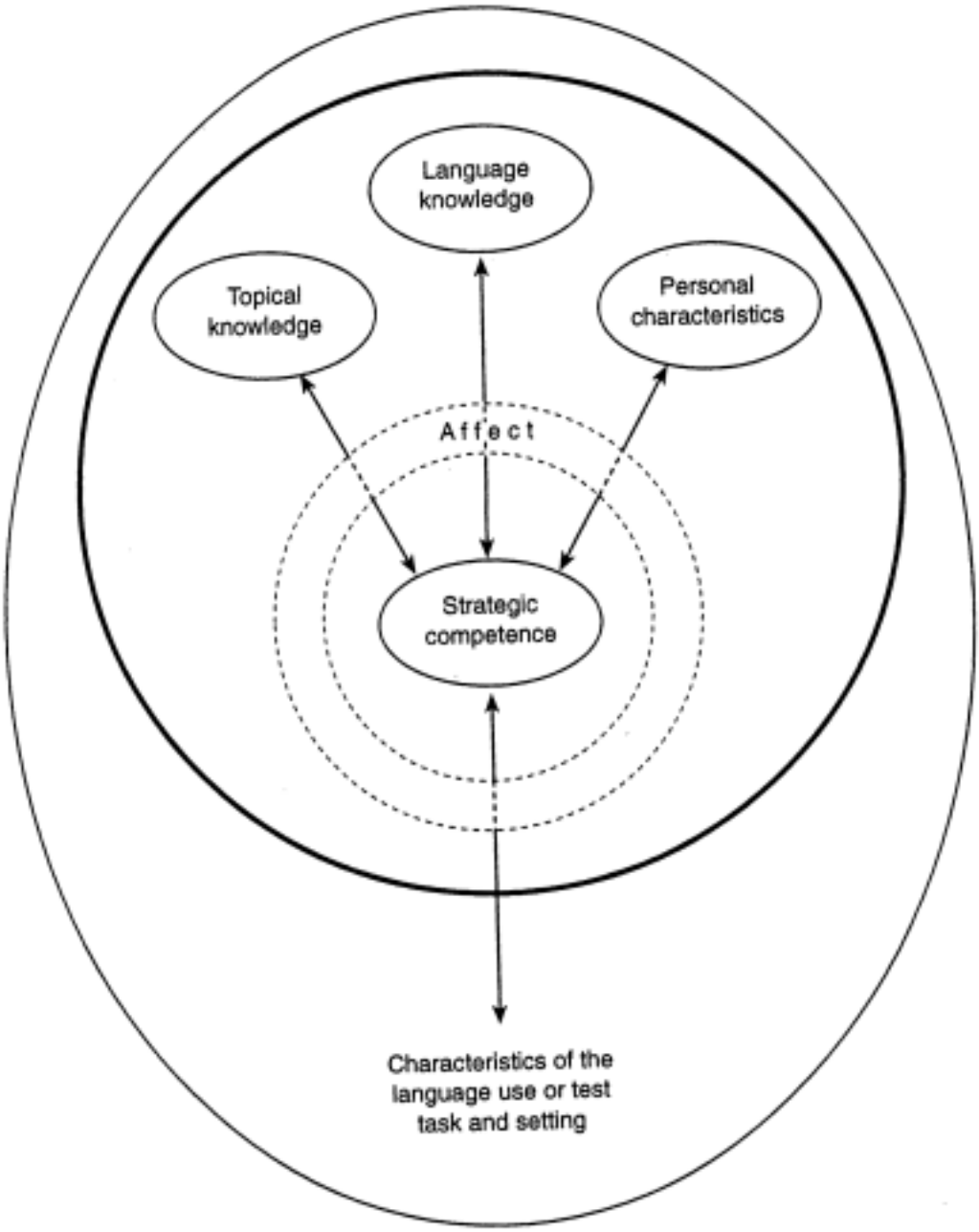


Figure 2.1. Some components of language use and language test performance. From *Language testing in practice*. by L. F. Bachman, & A. S. Palmer, 1996, p. 63.

section 2.1.2. Second, personal characteristics include age, sex, nationality, native

language, type and amount of preparation or prior experience with a given test, and others. Third, topical knowledge (i.e., knowledge schemata, or real-world knowledge) is "knowledge structure in long-term memory" (p. 65). Fourth, affective schemata are "affective or emotional correlates of topical knowledge" (p. 65). The second major factor (i.e., language use tasks and test tasks) is composed of five elements (p. 47): setting, test rubric, input, expected response, and relationship between input and response (see Table 2.1 for the summary).

Table 2.1
Summary of Bachman & Palmer's (1996) Factors That Affect Language Performance

Characteristics of language users or test takers	2) Characteristics of the language use or test task and setting: Example
(1) Language ability (1.1) Language knowledge (1.2) Strategic competence	(1) Setting (physical characteristics, participants, and time of task)
(2) Personal characteristics	(2) Test rubric (instructions, structure, time allotment, and scoring method)
(3) Topical knowledge	(3) Input (format and language of input)
(4) Affective schemata	(4) Expected response (format and language of expected response)
	(5) Relationship between input and response (reactivity, scope of relationship, and directness of relationship)

Note. From Bachman & Palmer (1996, pp. 49-50, 63-70). Language performance includes speaking performance.

Bachman & Palmer's (1996) performance model has been used in several studies (e.g., Douglas, 2000; Luoma, 2004; Phakiti, 2003). The current study is broadly based on their performance model because it seems to be the latest and most comprehensive, and accords well with previous studies.

Recently, models that focus on speaking performance (Fulcher, 2003; Luoma, 2001; McNamara, 1996; Reed & Cohen, 2001; Skehan, 1998, 2001; Upshur & Turner,

1999) have been proposed. Among them, the most detailed one is Fulcher. Yet, his proposed model is "programmatic" and "provisional upon future research" (p. 114). It is a comprehensive model that includes various factors that affect speaking performance, from stages of producing performance to stages of rating and interpreting speaking performance. Figure 2.2 shows how the factors interact with each other (see Table 2.2 for the summary). An example of the interaction of "individual variable" in test taker's factors and "topic" in task factors is that the eagerness with which test takers approach tasks may depend on task topics.

Table 2.2
Summary of Fulcher's (2003) Factors That Affect Speaking Performance

Factors that affect performance before rating	Factors that affect rating	Other interrelationships
(1) Test taker's factors (1.1) Task-specific knowledge or skills (1.2) Real-time processing (1.3) Abilities/capacities on constructs (1.4) Individual variable (e.g., personality) (1.5) Decisions and consequences (influenced by "score and inferences about the test taker") (2) Task factors (2.1) Orientation (2.2) Interactional relationship (2.3) Goals (2.4) Interlocutors (2.5) Topics (2.6) Situations (2.7) Additional task characteristics or conditions as required for specific contexts (3) Other factors (3.1) Local performance conditions (at the time of the test) (3.2) Interlocutor(s)	(4) Factors of rating scale/band descriptors (4.1) Construct definition (4.2) Orientation (4.3) Scoring philosophy (5) Rater(s)' factors (5.1) Characteristics (5.2) Training	(6) Construct definition (7) Score and inferences about the test taker (8) Abilities/capacities on constructs

Note. From Fulcher (2003, p. 115). Construct = "the trait or traits that a test is intended to measure" (Davies et al., 1999, p. 182).

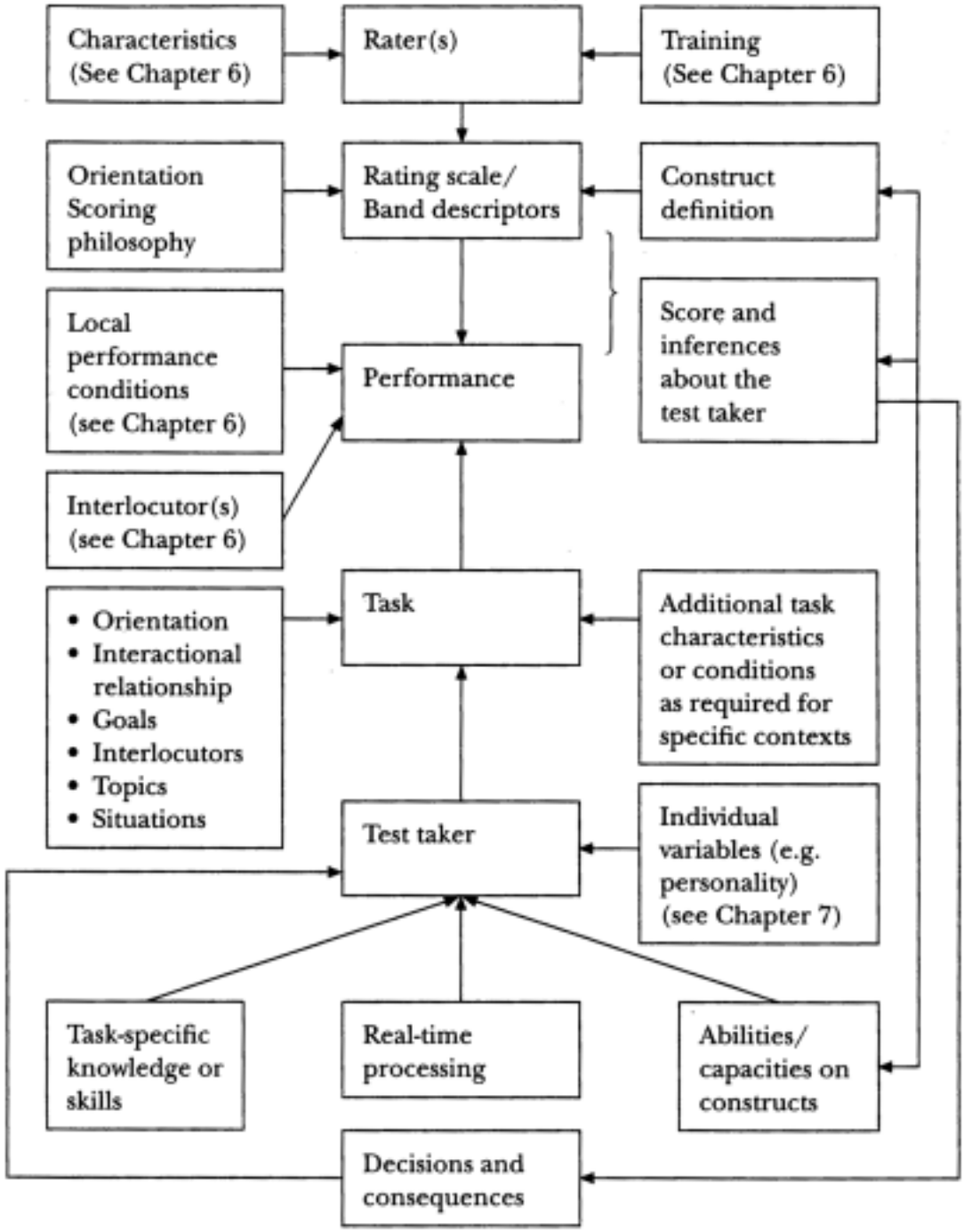


Figure 2.2. An expanded model of speaking test performance. From *Testing second language speaking.* by G. Fulcher, 2003, p. 115.

As Tables 2.1 and 2.2 indicate, Fulcher's (2003) model corresponds well with Bachman and Palmer's (1996). There are three common cores between Bachman and Palmer's and Fulcher's models. First, Bachman and Palmer's "characteristics of language users or test takers" corresponds to Fulcher's "test taker's factors." Second, Bachman and Palmer's "characteristics of language use tasks and test tasks" corresponds to Fulcher's "task factors." Third, there are complex interactions between several components. In summary, both models indicate that speaking performance is affected by three main factors: (a) test taker's internal factors, (b) external or contextual factors, and (c) interactions between (a) and (b). However, there are four main differences between the two models that are related to "test taker's factors" (Fulcher). First, "personal characteristics" and "affective schemata" (Bachman & Palmer) are combined into one test taker's factor, the "individual variable" (Fulcher). Second, some aspects of the "topical knowledge" (Bachman & Palmer) belong to "task-specific knowledge" and/or "individual variable" among test taker's factors (Fulcher). Third, "strategic competence" (Bachman & Palmer) corresponds to "task-specific knowledge or skills" and "real-time processing" among test taker's factors (Fulcher). Some may also belong to "abilities/capacities on constructs" (Fulcher). Fourth, "decisions and consequences" is added in test taker's factors (Fulcher).

Recently, research investigating effects of external factors (i.e., "characteristics of language use tasks and test tasks" in Bachman and Palmer, 1996, and "task factors," "other factors," "factors of rating scale/band descriptors," and "rater(s)' factors" in Fulcher, 2003) on speaking performance has been called for (e.g., Chalhoub-Deville, 2003; Swain, 2001). Some studies have been conducted to explore such effects by focusing on certain aspects (e.g., Bonk & Ockey, 2003; Brown, A.,

2003, 2004; Brown, A., & Lumley, 1997; Chalhoub-Deville, 1995, 1996; Kobayashi & Van Moere, 2005; Lazaraton, 1992, 1996; O'Loughlin, 2002; O'Sullivan, 2002, 2004; Pavlou, 1997). However, equally important is (a) to examine internal factors (i.e., "characteristics of language users or test takers" in Bachman and Palmer, and "test takers' factors" in Fulcher), and (b) to see relative effects and interactions of internal and external factors on speaking performance (Chalhoub-Deville, 2003). This study attempts to investigate the first issue, especially focusing on "language knowledge" in "language ability" in Bachman and Palmer's model and "abilities/capacities on constructs" in "test taker's factors" in Fulcher because it is more relevant to and has more direct impacts on teaching and testing.

Among components of "language knowledge," this study focuses on "vocabulary knowledge" because of its theoretical and empirical importance (see section 2.2.1). This study does not intend to examine other internal factors at the same time because focusing on one factor and examining it in detail using tests and measures with more validity seems a sounder and steadier approach than including many possible factors into a research design. In the next section, components of "language ability" are illustrated.

2.1.2 Language Ability

In this section, components of "language ability" are reviewed in order to locate the place of "vocabulary knowledge" in the framework. Although many researchers (Bachman, 1990; Bachman & Palmer, 1996; Canale, 1983; Canale & Swain, 1980; Davies, 1989; Hymes, 1972; Taylor, 1988) proposed their own categories of language ability, McNamara (1996) summarized their views and classified components of language ability into two categories, following Hymes (1972): language knowledge

and ability for use. In Bachman and Palmer's (1996) model, "language knowledge" consists of (a) organizational knowledge and (b) pragmatic knowledge. The first one is the knowledge to use a language in order to comprehend and produce grammatical utterances, sentences, and discourse, consisting of grammatical knowledge (i.e., knowledge of vocabulary, syntax, phonology, and graphology) and textual knowledge. The second one involves the knowledge to interpret and produce discourse by connecting utterances, sentences, and texts with meanings and intentions, and is separated into functional and sociolinguistic knowledge (pp. 67-69). Buck (2001) stated that "language knowledge" in Bachman and Palmer's model is composed of two types: declarative knowledge and procedural knowledge (p. 104). Declarative knowledge is "the knowledge of facts" (Buck, 1997, p. 67), "encyclopedic in nature" (Ellis, 2005a, p. 148), and static (Yamaoka, 2004, p. 27). In contrast, procedural knowledge is "knowledge as ability to do something" (Buck, 1997, p. 67) and dynamic knowledge that is related to language performance and that leads to performance directly (Yamaoka, 2004, p. 27).[2.2]

"Ability for use" is the ability to use one's knowledge of language. Bachman and Palmer (1996) incorporated it into a broader category of strategic competence, or "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). It includes the ability to (a) set goals, (b) assess, and (c) plan. Various aspects of a speaking activity (i.e., speaking performance) that is produced using language ability (especially speaking ability) are described below, based on Bachman and Palmer's terms.

Suppose that a female speaker feels hot and decides to ask her male friend if it is okay to open the window. This is a goal-setting activity of "identifying" and

"deciding" on the appropriate message. Next, she "assesses what is needed to complete" her task of conveying her message (assessment 1) and her language "knowledge components" (assessment 2). She then "selects elements from the areas of language knowledge," both from organizational knowledge, such as vocabulary and syntax, and from pragmatic knowledge (planning 1), and chooses one of the "plans formulated for implementing these elements" to express her message successfully (planning 2). Lastly, she "assesses the correctness or appropriateness" of the expressions formulated (assessment 3) and says, "Would you mind opening the window?" (Terms are from Bachman and Palmer, 1996, p. 71; Examples are made by the author).[2.3]

In Bachman and Palmer's (1996) models, vocabulary knowledge, which is the focus of this study, belongs to "organizational knowledge" in "language ability." The current study focuses on declarative knowledge rather than procedural knowledge of vocabulary (see Table 2.3 for the area where the focus of the current study is positioned). In the next section, literature on vocabulary knowledge is reviewed.

Table 2.3
Bachman & Palmer's (1996) Characteristics of Language Users or Test Takers

(1) Language ability
  (1.1) Language knowledge
    (1.1a) Organizational knowledge (including productive vocabulary knowledge, especially declarative aspect of productive vocabulary knowledge)
    (1.2b) Pragmatic knowledge
  (1.2) Strategic competence

(2) Personal characteristics

(3) Topical knowledge

(4) Affective schemata

Note. From Bachman & Palmer (1996, pp. 49-50, 63-70) with modification. Underlined = The focus of the present study.

2.2 Productive Vocabulary Knowledge

2.2.1 Aspects of Productive Vocabulary Knowledge

This section describes productive vocabulary knowledge from two perspectives: aspects of vocabulary knowledge (i.e., "receptive vs. productive" and "size vs. depth") and their assessment methods. Vocabulary can be conceptualized in many ways (e.g., Henricksen, 1999; Nation, 2001; Qian, 2002). For instance, Qian (2002) summarized previous categories and proposed a framework that consists of the following four aspects: (a) vocabulary size, (b) depth of vocabulary knowledge, (c) lexical organization (i.e., "the storage, connection, and representation of words in the mental lexicon"), and (d) automaticity of receptive-productive knowledge (i.e., "all fundamental processes through which access to word knowledge is achieved for both receptive and productive purposes," p. 516). Another example is Meara (1996), who described vocabulary from two dimensions: size and organization (i.e., "the 'network of associations between [a] word and other words in the language," p. 47).

The broadest framework can be seen in Chapelle (1994), who conceptualized vocabulary ability as having three components: the context of vocabulary use, vocabulary knowledge and fundamental processes, and metacognitive strategies for vocabulary use. The first aspect, the context of vocabulary use, includes not only the linguistic context (e.g., sentences with the target word) but also the pragmatic context (e.g., "differences across generations [teenagers vs. adults] and between colloquial and more formal uses of words"; Read, 2000, p. 31). The second constituent, vocabulary knowledge and fundamental processes, has four dimensions: (1) vocabulary size, (2) knowledge of word characteristics, (3) lexicon organization, and (4) fundamental vocabulary processes. The first two are described below. Lexicon organization is a way in which words are related to one another, whereas fundamental processes involve the

automaticity in accessing and utilizing each word. The third component of vocabulary ability categorized by Chapelle (1994) is metacognitive strategies for vocabulary use. It is the same as Bachman's (1990) strategic competence, or "the mental capacity for implementing the components of language competence in contextualized communicative language use" (Bachman, p. 84).

While there are many types of classification, this study mainly analyzes vocabulary knowledge from two viewpoints: (a) receptive vs. productive, and (b) size vs. depth. As for the distinction between receptive and productive (e.g., Nation, 2001, pp. 24-26, 2005, pp. 584-585; Read, 2000, p. 154; Schmitt, 2000, p. 4), receptive vocabulary knowledge is the knowledge to understand a word in listening and reading, whereas productive vocabulary knowledge is the knowledge to produce a word[1.4] when one speaks and writes (Nation, 2005, p. 585; Schmitt, 2000, p. 4). This study focuses on productive vocabulary knowledge by dealing with its relationships with speaking performance because productive vocabulary knowledge seems more relevant to speaking performance by its definition.

As for relationships between the two types of vocabulary, there are two ways to look at them. First, receptive vocabulary and productive vocabulary are considered to be located at opposite ends of the continuum of vocabulary development (Melka, 1997). In Melka's model, it is overall considered that receptive vocabulary gradually changes into productive vocabulary, and learners are gradually able to use productive vocabulary in speaking and writing, but there is overlap between receptive and productive vocabulary. Second, Meara (1990) regarded receptive and productive vocabulary as different entities that are activated by different stimuli (i.e., external stimuli, such as sound and spelling, vs. association with other words), not something that belongs to the continuum. The current study uses Melka's view because it is more

widely accepted (e.g., Meara, p. 150; see Aizawa, Ochiai, & Osaki, 2003, for the use of Melka's framework).

Although the general concept of receptive and productive vocabulary is consistent, previous studies have used definitions of the terms "receptive (sometimes called passive)" and "productive (sometimes called active)" ambiguously or differently, which leads to difficulty in interpreting previous results (Read, 2000).

There have been two attempts to address these problems. First, Read (2000) proposed a four-cell matrix of differentiating receptive and productive vocabulary (Read, 2000, pp. 154-157), as seen in Table 2.4. He divided receptive vocabulary into two types: recognition and comprehension. He also separated productive vocabulary into two types: recall and use, depending on whether there is a context. The context here "includes whole texts and, more generally, discourse" (Read, 2000, p. 11). "Recall" refers to retrieval of vocabulary from memory in response to a stimulus word, whereas "use" refers to production of vocabulary in speaking and writing (Read, pp. 155-156).

Table 2.4
Four Types of Vocabulary

	Receptive	Productive
Out of context	Recognition	Recall
In context	Comprehension	Use

Note. From Read (2000, pp. 154-157).

Second, Laufer, Elder, Hill, and Congdon (2004) classified vocabulary knowledge into four types using two perspectives. The first perspective was what to retrieve using vocabulary knowledge (i.e., retrieve meaning from form vs. retrieve form from meaning). The second perspective was how to retrieve either meaning or form (i.e., recognition vs. recall). For example, when a person says "I like English,"

meaning or concept comes first and the form "English" is recalled from it. This belongs to the upper cell on the left and what to retrieve is "form" and the retrieval method is "recall" (see Table 2.5).

Table 2.5
Types of Vocabulary Knowledge (Four Degrees of Knowledge of Meaning)

	Recall	Recognition
Active (Productive) (Retrieval of form)	Active recall e.g.,) Turn into water *m*_____ (Answer: melt)	Active recognition e.g.,) *Turn into water* a. elect b. blame c. melt* d. threaten
Passive (Receptive) (Retrieval of meaning)	Passive recall e.g.,) When something *melts* it turns into _____ (Answer: water)	Passive recognition e.g.,) *Melt* a. choose b. accuse c. make threats d. turn into water*

Note. From Laufer et al. (2004, pp. 206-208). * = Answer.

While Read (2000) categorized productive vocabulary according to the existence or absence of context, Laufer et al. (2004) further subdivided Read's "recall" into two types according to what is to be retrieved (i.e., form or meaning). Their classifications are essential and highly valuable, and they are used as a basis in this study. To be specific, in the current study, productive vocabulary knowledge is defined as "knowledge to produce vocabulary forms," whereas differences between knowledge and performance are specified as having the context or not. It should be noted that the definition of productive vocabulary knowledge is different from Laufer et al.'s in that their definition of productive vocabulary knowledge included both "active recall" and "active recognition" (see Table 2.5).

Another distinction of vocabulary knowledge is that it can be classified into two types: size and depth (Anderson & Freebody, 1981; Nakanishi & Shimamoto, 2003). Vocabulary size is also called breadth, which is a quantitative dimension of vocabulary knowledge. It refers to the number of words with a primary meaning that a

learner has (Nakanishi & Shimamoto, 2003, p. 32), or "the number of words for which the person knows at least some of the significant aspects of meaning" (Anderson & Freebody, 1981, p. 93). Vocabulary depth is a qualitative dimension of vocabulary knowledge and refers to the degree to which a learner knows a certain word in addition to knowing a primary meaning (Nakanishi & Shimamoto, 2003, p. 32).

When size is focused on, there are four ways of counting words: tokens, types, lemmas, and word families (e.g., Mochizuki, Aizawa, & Tono, 2003; Nation, 2001; Nakanishi & Shimamoto, 2003, p. 36). In counting tokens, all words that exist in a text are counted as one separate word, even when the same words are repeatedly used. In counting types, the same words are counted as only one word when repeatedly used, and a base form (i.e., headword) and inflected forms (e.g., the base form of *lives* is *live* and one inflected form of *live* is *lives*) are considered to be different types. In counting lemmas, a base form, inflected forms, and reduced forms (e.g., *n't*) are considered to be the same lemma and are only counted as one. In counting word families, a base form, inflected forms, reduced forms, and derived forms are considered to be the same word family and only counted as one. Therefore, the number is largest for tokens, followed by types, and word families have the smallest number (see Table 2.6). As for relationships between lemmas and word families, the following formula has been used: the number of lemmas = the number of word families*1.6 (Laufer, 1992, p.130, 1997, pp. 23-24).

Table 2.6
Four Ways to Count Words
Example: *I have a friend. She is very friendly. She has a dog.*

	I	have	a	friend.	She	is	friendly.	She	has	a	dog.
Tokens	1	2	3	4	5	6	7	8	9	10	11
Types	1	2	3	4	5	6	7		8		9
Lemmas	1	2	3	4	5	6	7				8
Word families	1	2	3	4	5	6					7

Note. The number in the example: 11 tokens, 9 types, 8 lemmas, and 7 word families.

Depth includes word frequency, association, affix knowledge, syntactic characteristics, and other aspects (Nation, 2001). The more comprehensive framework of aspects of vocabulary knowledge can be seen in Table 2.7 (Nation, 2001, p. 27). One characteristic of this framework is that each aspect of vocabulary knowledge has a receptive and a productive dimension, and Nation's framework belongs to "knowledge of word characteristics" in Chapelle's (1994) vocabulary ability framework (Read, 2000, p. 32). While size seems to consist of spoken and written form ((a) to (d)) and form and meaning ((g) and (h)), depth is made up of the other aspects. Read (2004) presented another classification and divided depth into three aspects: precision of meaning, comprehensive word knowledge, and network knowledge. First, precision

Table 2.7
Aspects of Vocabulary Knowledge

Form	Spoken	Rec	(a) What does the word sound like?
		Pro	(b) How is the word pronounced?
	Written	Rec	(c) What does the word look like?
		Pro	(d) How is the word written and spelled?
	Word parts	Rec	(e) What parts are recognizable in this word?
		Pro	(f) What word parts are needed to express the meaning?
Meaning	Form and meaning	Rec	(g) What meaning does this word form signal?
		Pro	(h) What word form can be used to express this meaning?
	Concept and referents	Rec	(i) What is included in the concept?
		Pro	(j) What items can the concept refer to?
	Associations	Rec	(k) What other words does this make us think of?
		Pro	(l) What other words could we use instead of this one?
Use	Grammatical functions	Rec	(m) In what patterns does the word occur?
		Pro	(n) In what patterns must we use this word?
	Collocations	Rec	(o) What words or types of words occur with this one?
		Pro	(p) What words or types of words must we use with this one?
	Constraints on use (register, frequency …)	Rec	(q) Where, when, and how often would we expect to meet this word?
		Pro	(r) Where, when, and how often can we use this word?

Note. From Nation (2001, p. 27) with minor modification in the title and spellings. Rec = Receptive knowledge; Pro = Productive knowledge.

of meaning includes “having a limited, vague idea of what a word means and having much more elaborated and specific knowledge of its meaning” (p. 211). Second, comprehensive word knowledge refers to “knowledge of a word which includes not only its semantic features but also its orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics” (p. 211). Third, network knowledge is defined as “the incorporation of the word into a lexical network in the mental lexicon together with the ability to link it to -- and distinguish it from -- related words” (p. 212). This study only deals with the second and third aspects because the first seems closely associated with meaning and the size aspect.

As for the relative importance of size and depth, size, which is essentially knowledge of a primary meaning and a form of a word, has been considered the most important aspect of vocabulary knowledge (Laufer et al., 2004; Laufer & Goldstein, 2004; Webb, 2005). Meara (1996) argued that size is the only important dimension when learners know a small number of words (i.e., 5,000 to 6,000 words [which probably equals lemmas] in English, p. 45)[2.4] and that “once this critical threshold is reached, vocabulary size per se seems to become less important” (p. 45). In addition, vocabulary studies have mainly focused on size (Nakanishi & Shimamoto, 2003, p. 45), and depth has not been investigated until 1990 (pp. 49-50) because “they [tests] can give a more representative picture of the overall state of the learners’ vocabulary than an in-depth probe of a limited number of words” (Read, 2000, p. 115). However, by examining depth as well as size, it is possible to examine vocabulary knowledge more comprehensively and systematically and enjoy the advantages of both size and depth (Nakanishi & Shimamoto, 2003, pp. 45-46).

2.2.2 Tests to Assess Productive Vocabulary Knowledge

In order to examine vocabulary knowledge empirically, tests that can be interpreted more consistently and meaningfully are needed. Although some tests have been made, they have not gone through a validation process and the constructs ("the trait or traits that a test is intended to measure," Davies et al., 1999, p. 182) have been rather vague (Read, 2000). Thus, different results using different tests have been difficult to interpret (Shimamoto, 2005).

Read (2000) presented a framework containing "three dimensions of vocabulary assessment" to describe constructs more precisely (pp. 8-13; see Table 2.8). The first dimension examines whether vocabulary is the whole construct (discrete) or part of the

Table 2.8
Dimensions of Vocabulary Assessment

(1) Relationship between vocabulary and the construct	Discrete when a whole test is intended to measure vocabulary knowledge (e.g.,) Vocabulary Levels Test, Lexical Frequency Profile, Vocabulary Knowledge Scale	Embedded when vocabulary is assessed in a writing test; when reading tests include vocabulary items (e.g.,) ESL Composition Profile, TOEFL vocabulary items, Lexical Density Index
(2) The range of vocabulary to be included in the assessment	Selective when target words are underlined and test takers respond to items (e.g.,) Vocabulary Levels Test, TOEFL vocabulary items, Vocabulary Knowledge Scale	Comprehensive when all the words uttered in a speaking test are evaluated in rating "vocabulary range" (e.g.,) Lexical Frequency Profile, ESL Composition Profile, Lexical Density Index
(3) The role of context: to what extent the test takers are being assessed on the basis of their ability to engage with the context provided in the text	Context-independent when vocabulary items can be solved without looking at the context in reading tests (e.g.,) Vocabulary Levels Test, Vocabulary Knowledge Scale	Context-dependent when vocabulary items cannot be solved without looking at the context; when aspects of vocabulary in a speaking test are assessed (e.g.,) Lexical Frequency Profile, ESL Composition Profile, Lexical Density Index, (TOEFL vocabulary items: "variably context dependent," Read & Chapelle, p. 6)

Note. From Read (2000, p. 9) and Read & Chapelle (2001, p. 6). See section 2.2.2, for Lexical Frequency Profile and Vocabulary Knowledge Scale. The Lexical Density Index is "derived by calculating the percentage of lexical items in the test-takers' responses to each format for each task" (Read & Chapelle, p. 31).

construct (embedded). The second dimension examines whether the target of the assessment is certain selected vocabulary items (selective) or the whole text or responses (comprehensive). The third dimension examines whether it is necessary to use the context (i.e., "whole texts and, more generally discourse," p. 11) when learners take a test. If it is not needed, it is context-independent, and otherwise, it is context-dependent. It should be noted that in this framework whether or not vocabulary items are placed in the context is not a matter of concern.

In this section, nine existent tests were analyzed using Read's (2000) framework. A summary was made only for tests in which test takers are required to produce forms (in the upper cell on the left in Laufer et al.'s [2004] classification; see Table 2.5). Thus, receptive vocabulary tests, such as the Eurocentres Vocabulary Size Test (Meara & Buxton, 1987), a Vocabulary Levels Test (Nation, 2001), a Vocabulary Size Test for Japanese EFL Learners (Aizawa, 1998), and a Vocabulary Size Test for Japanese Learners of English (Mochizuki, 1998), and receptive vocabulary depth tests, such as Read (1993, 1998), Shimamoto (2000), and Mochizuki (2002a) were not included.

There are nine main tests of productive vocabulary knowledge, and these assess different aspects of vocabulary. Firstly, in the Lexical Frequency Profile (LFP; Laufer & Nation, 1995), test takers write an essay of more than 200 words (p. 314). Then "the percentage of words a learner uses at different vocabulary frequency levels" (p. 311) is computed using computer software, and the ratio obtained is interpreted as "free productive ability" (Laufer & Nation, 1999, p. 37). Originally, the frequency levels used were "the first 1,000 most frequent words, the second 1,000, the University Word List, and the 'not-in-the-lists' word list" (p. 315), but Laufer (1995) suggested that a ratio of words "beyond 2000" is also a good measure of productive vocabulary (p. 267).

Two main problems have been pointed out with the LFP by Meara and Fitzpatrick (2000): the effects of topics on writing samples and use of vocabulary, and a lack of practicality. They doubted whether the topics for essays are appropriate for eliciting varied sample of vocabulary and whether obtaining long essays is feasible from the learners, especially when they have limited time (p. 21). Additionally, Meara (2005) conducted a simulation study and stated that the LEP "works best when the groups being compared have very disparate vocabulary sizes, and is probably not sensitive enough to pick up modest changes in vocabulary size" (p. 32).

Secondly, a vocabulary-size test of controlled productive ability (VTCPA; Laufer & Nation, 1999) requires test takers to fill in blanks of the underlined target words (see Table 2.9), and their answers are subsequently scored dichotomously (i.e., scored either right or wrong).

Table 2.9
Examples of the Vocabulary-Size Test of Controlled Productive Ability

Complete the underlined words. The example has been done for you.

He was riding a bic*ycle*.

I'm glad we had this opp____ to talk.	[Answer: opportunity]
The thieves threw ac___ in his face and made him blind.	[acid]

Note. From Laufer & Nation (1999, p. 46).

The main problem with the VTCPA is that the aspect of vocabulary ability that the test measures is rather equivocal (Aizawa et al., 2003, p. 158) and may not be limited to productive vocabulary ability for some students (Mochizuki & Aizawa, 2001, pp. 93-99; Read, 2000, p. 125). For example, Mochizuki and Aizawa showed that the VTCPA scores were affected by background knowledge and the understanding of stems and contexts (pp. 94-96), concluding "the VTCPA may be unsuitable for beginning to low-intermediate learners" (p. 99).

Thirdly, the Lex30 (Meara & Fitzpatrick, 2000) measures context-limited productive vocabulary by eliciting word associations (p. 22). Test takers are expected to write, within 30 seconds, as many words as possible that they come up with when looking at a stimulus word, as demonstrated by the next example in Table 2.10.

Table 2.10
Examples of the Lex30

(Explanation was directly cited from p. 23)
The testees were asked to write a series of response words (at least three if possible) for each stimulus word, using free association (an example was worked through with each class before the test). Stimulus words were presented one at a time, and testees had 30 s to respond to each cue, after which the administrator called the next number; the entire test therefore took 15 min to complete.

attack (stimulus): (example of a test taker's answer) war, castle, guns, armour
spell: grammar, test, bell

Note. From Meara & Fitzpatrick (2000, p. 28).

The words generated are classified according to their frequency levels, and the number of words is counted as scores, except for the "high frequency structure words, proper names and numbers" and "the 1,000 most frequent content words" (p. 23). As for the problems with the Lex30, Mochizuki (2002b) reported almost zero correlation (.11) between Lex30 scores and translation scores (regarded as an external criterion of productive vocabulary) among beginners, suggesting that one reason may be that many loanwords are produced in response to a stimulus word (p. 462). These loanwords are often low frequency words, which lead to higher Lex30 scores and, by extension, higher estimates of productive vocabulary. Therefore, using the Lex30 with beginners could be considered a problem.

Fourthly, Laufer et al. (2004) constructed a test intended to measures four types of vocabulary knowledge (computer adaptive test of size and strength [CATSS]; see Table 2.5). Among them, a format that asks test takers to produce forms can be seen in

Table 2.11.

Table 2.11
Example of the Computer Adaptive Test of Size and Strength

(Explanation was directly cited from p. 206)
The task is to supply the L2 target word (*melt*). The first letter of the word is provided to avoid nontarget words that have the same meaning (e.g., *condense*).

Turn into water (L2 meaning) *m*_____ (Answer: melt)

Note. From Laufer et al. (2004, pp. 206-208).

Laufer and Goldstein (2004) used a similar translation format but it is different from Laufer et al. (2004) in that Laufer and Goldstein presented a meaning of first language (L1) to elicit an equivalent L2 word. Nation (2001, p. 351) recommended translation as a useful way of assessing vocabulary and stated that "the use of the first language to convey and test word meaning is very efficient" (p. 351). Moreover, recent studies have shown that in the vocabulary representation of beginners, L1 vocabulary meaning is mediated between L2 meaning and the concept (Chen & Leung, 1989; Jiang, 2000; Kawakami, 1994; Matsumi, 1993), which can be the rationale for using L1. In fact, Mochizuki and Aizawa (2001) as well as Mochizuki (2002b) used translation for validation of other productive vocabulary tests. Their assumption seems to be that translation is a good method for measuring productive vocabulary. The current study uses this format and examines its quality (see Chapters 4 and 9).

Fifthly, the Vocabulary Knowledge Scale was created by Wesche and Paribakht (1993) in order to measure "certain states in the initial development of core knowledge of given words" (p. 29). As seen in Table 2.12, the format includes both self-assessment and some production items (see also Table 2.13 for the meaning of scores).

Table 2.12
Vocabulary Knowledge Scale

(Explanation was directly cited from p. 29)
Students were asked to indicate their degree of knowledge of each target word using the scale.

Self-report categories
Ⅰ. I don't remember having seen this word before. [Possible score: 1]
Ⅱ. I have seen this word before, but I don't know what it means. [Possible score: 2]
Ⅲ. I have seen this word before, and I *think* it means _____. (synonym or translation) [Possible scores: 2 or 3]
Ⅳ. I *know* this word. It means _____. (synonym or translation) [Possible scores: 2, 3, or 4]
Ⅴ. I can use this word in a sentence:_____. (If you do this section, please also do Section Ⅳ.) [Possible scores: 2, 3, 4, or 5]

Note. From Wesche & Paribakht (1993, p. 30). Scores vary according to a quality of responses (e.g., If the synonym or translation is wrong, those who choose Ⅲ gain a score of 2.).

Table 2.13
Meaning of Scores of the Vocabulary Knowledge Scale

Score	Meaning of scores
1	The word is not familiar at all.
2	The word is familiar but is [sic] meaning is not known.
3	A correct synonym or translation is given.
4	The word is used with semantic appropriateness in a sentence.
5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Note. From Wesche & Paribakht (1993, pp. 30-31) with modification.

There are three main problems with this test. First, it takes time to conduct and score the test (Mochizuki et al., 2003, p. 197). Second, it is not clear whether a score difference reflects the degree of vocabulary knowledge (Mochizuki et al., p. 197; Read, 1997, p. 102) and whether using five stages of vocabulary development is appropriate (Nakanishi & Shimamoto, 2003, p. 50; Read, 1997, p. 102). Third, according to Nakanishi and Shimamoto, the aspects of vocabulary knowledge assessed in this test are limited (e.g., multiple meanings are not assessed, p. 50). From the point of view of selecting appropriate measures for research, the second point seems to be the largest problem because the ambiguity of test scores may lead to difficulty in interpreting results. The third problem is crucial, but almost any test seems to have this limitation because it is impossible to cover all the content, and selection of relevant and representative content is always necessary in test development, although the degree of

content relevance and representativeness is an issue.

Sixthly, Schmitt and Meara (1997) used a format that elicited knowledge of suffixes and word associations (see Table 2.14). With this, test takers were presented with a stimulus word and required to write (a) all suffixes that they thought could be added to the stimulus word, and (b) three word associations prompted from the stimulus. The responses were scored based on a "native speaker baseline master key" or a judgment by native speakers (p. 23). One possible problem with this format is that test takers can write as many suffixes as possible, so test scores may reflect individual characteristics of test takers to some degree. For instance, those who tend to write more than others, even when they are not sure of their accuracy may or may not be rewarded depending on scoring methods. Despite the importance of the scoring methods, Schmitt and Meara did not describe this element in their paper. Another problem is that it may be a little difficult to compare the results of suffixes and word associations (which Schmitt and Meara did in the paper) due to differences in eliciting responses (i.e., test takers can write as many suffixes as possible, whereas they can write a limited number of [i.e., three] associations). An individual interview format

Table 2.14
Examples of Schmitt & Meara's (1997) Format

(Explanation was directly cited from p. 22)
The productive section listed the 20 prompt words, followed by a blank on which to write any suffixes the subjects thought were allowable and then three blanks on which the subjects were to give three word associations. The instructions made clear that all of the prompt words were verbs.

use _____ _____ _____ _____
(Suffix answers: -able, -age, -ed, -er/or, -ing, -s)
(Association example answers: computer, employ, and tool)

halt _____ _____ _____ _____
(Suffix answers: -able, -ed, -er/or, -ing, -s)
(Association example answers: famous, say, and person)

Note. From Schmitt & Meara (1997, p. 36).

was used by Schmitt (1998, 1999), which elicited productive vocabulary knowledge in relation to a stimulus word (e.g., knowledge of associations, word classes, and multiple meanings), but the two problems mentioned above also seem to be related here.

Seventhly, Schmitt and Zimmerman (2002) developed a format to measure knowledge of derivatives as shown in Table 2.15. While what each item measures seems clear and there is a comparability of items, this format is not appropriate for the purpose of the current research because a sentence with a gap to be filled is considered to be a context of language use, and knowledge of derivatives measured with this format does not fit the definition of "knowledge" in this study.

Table 2.15
Examples of Schmitt & Zimmerman's (2002) Format

Directions: Most words can be changed to different parts of speech. For example, the word STIMULATE is a verb but can be changed to a noun form (STIMULATION) or an adjective form (STIMULATING). In the case of STIMULATE, there is no adverb form, but many other words do have an adverb form.

In this section, look at each word and write the correct form in each sentence. If there is more than one possibility (e.g., more than one adjective form) you only need to write one. If there is no form, put an "X" in the blank on the left. Sometimes the form will not need changing, as it is already correct (such as STIMULATE in the example below).

Example:
stimulate

stimulation	Noun	A massage is good _____.
stimulate	Verb	Massages can _____ tired muscles.
stimulating	Adjective	A massage has a _____ effect.
X	Adverb	He massaged_____.

1. assume

Noun	He made an _____ that she likes meat.	(Answer: assumption)
Verb	He can _____ that she likes meat.	(Answer: assume)
Adjective	He had an _____ idea that she likes meat.	(Answer: assumed or X)
Adverb	He decided _____ that she likes meat.	(Answer: X)

Note. From Schmitt & Zimmerman's (2002, pp. 168-169). See Note 2.5 for how scoring keys were made.

Eighthly, Webb (2005) used a format that assessed (a) meaning and form, (b) grammatical functions, (c) syntax (i.e., syntagmatic associates[2.6]), (d) associations (i.e.,

paradigmatic associates[2.7]), including coordinates, superordinates, subordinates, antonyms, and synonyms, and (e) orthography. Example items can be seen in Table 2.16. The first format to measure (a) meaning and form is the same as the one used in

Table 2.16
Examples of Webb's (2005) Format

(Explanation was directly cited from pp. 39-41)
(a) Productive knowledge of meaning and form
The learners were given the L1 meanings and asked to write the words that the meanings had been paired with in the treatment.

機関車 _____ (Answer: masco [nonword taught in class])

(b) Productive knowledge of grammatical functions
Learners were cued with the target words and had to write each one in a sentence. It was made clear in the instructions to the participants that the only determining factor for a correct response was using the target words with grammatical accuracy.

masco _____
(An example answer: The masco left the station early. It is a masco.)
(An example incorrect response: The girl mascoed to school.)

(c) Productive knowledge of syntax
The learners had to produce a L2 syntagmatic associate beside the cues, which were the target words.

masco _____
(An example answer: station, tracks, left, and arrived)
(An example incorrect response: clock, ate, and hard; Paradigmatic associates were marked as incorrect. This was carefully explained in the instructions to the participants.)

(d) Productive knowledge of association
Learners were presented with the target words and asked to write an associate beside each item.

masco _____
(An example answer [coordinates, superordinates, subordinates, antonyms, and synonyms]: train, airplane, and vehicle)
(Syntagmatic associates were marked as incorrect, as was carefully explained in the instructions.)

(e) Productive knowledge of orthography
The learners heard each target nonsense word pronounced twice and then had 10 seconds to write it correctly.

Note. From Webb (2005, pp. 39-41).

Laufer and Goldstein (2004) except that first letter of the word was not given. The second format (b) seems to have the same problem that Schmitt and Zimmerman (2002) have for the purpose of this research. The third and fourth formats (c and d) seem promising because the number of responses is specified as one and what the format assesses seems clearer, although empirical examination is needed to examine this.

Ninthly, Shimamoto (2005) used test formats to measure two types of productive vocabulary knowledge: paradigmatic and syntagmatic knowledge[2.8] (see Table 2.17 for the formats). One problem with Shimamoto (2005) seems to be the same as the first one in Schmitt and Meara (1997): the possible effects of individual characteristics and scoring methods on test scores. However, these formats may be useful if the instruction is modified from "eliciting as many words as possible" to "eliciting a word," which can be seen in Webb (2005).

Table 2.17
Examples of Shimamoto's (2005) Formats

Paradigmatic knowledge test items Write as many words related to the word in terms of meanings (e.g., synonyms and antonyms) as possible. disease (noun) _____ (An example answer: illness, sickness, and health) clean (adjective) _____ (An example answer: neat, white, and clear) perform (verb) _____ (An example answer: do, make, and carry out)
Syntagmatic knowledge test items (1) Write as many words (adjective and noun) that fit ( ) as possible. _____ disease (noun) (An example answer: rare, social, and fatal) (2) Write as many words (noun) that fits ( ) as possible. clean (adjective) _____ (An example answer: room, air, and water) (3) Write as many words (noun) that fits ( ) as possible. to perform (verb) (a/an/the) _____ (An example answer: operation, experiment, and role)

Note. The exact formats can be seen in a handout in Shimamoto & Kadota (2004).

So far, nine tests or formats (formats, hereafter) to assess productive vocabulary knowledge have been examined. It seems that each format assesses different aspects of

productive vocabulary knowledge and that aspects assessed are limited (Aizawa, 2005). These assessed aspects and characteristics of the formats are summarized in Table 2.18.

Table 2.18
Summary of Tests or Formats to Assess Productive Vocabulary Knowledge

	Aspects assessed	Characteristics
Lexical Frequency Profile (Laufer & Nation, 1995)	Size	Embedded; comprehensive; context-dependent
vocabulary-size test of controlled productive ability (Laufer & Nation, 1999)	Size	Discrete; selective; context-dependent
Lex30 (Meara & Fitzpatrick, 2000)	Size and depth (word associations)	Discrete; selective; context-independent
Laufer et al. (2004; Computer Adaptive Test of Size and Strength), Laufer & Goldstein (2004)	Size	Discrete; selective; context-independent
Vocabulary Knowledge Scale (Wesche & Paribakht, 1993)	Size and depth (synonym, grammatical functions)	Discrete; selective; context-independent and dependent
Schmitt & Meara (1997)	Depth (suffixes and word associations)	Discrete; selective; context-independent
Schmitt & Zimmerman (2002)	Depth (derivatives)	Discrete; selective; context-dependent
Webb (2005)	Size and depth (meaning and form, grammatical functions, syntax, associations, and orthography)	Discrete; selective; context-independent and dependent
Shimamoto (2005)	Depth (paradigmatic and syntagmatic knowledge)	Discrete; selective; context-independent
Productive Vocabulary Knowledge Test (the present study)	Size and depth (derivations, antonyms, and collocations)	Discrete; selective; context-independent
Speaking Test (the present study)		Embedded; comprehensive; context-dependent

These nine productive vocabulary test formats, which assess different aspects of productive vocabulary, certainly have their own values. However, in the context of selecting more appropriate formats of productive vocabulary knowledge for novice

learners, some formats tend to excel over others. This study decides to use the formats that seems able to measure the knowledge they are intended to measure (i.e., the construct) as a basis, that is, to utilize Laufer and Goldstein (2004) to assess vocabulary size, and Shimamoto (2005) to measure vocabulary depth. Then, their formats are modified and the validity of inferences and uses based on their scores are examined. In addition, by using Read's (2000) classification, this study can be conceptualized as a study that compares the two types of vocabulary that are assessed using two tests: (a) a discrete, selective, and context-independent test for productive vocabulary knowledge, and (b) an embedded, comprehensive, and context-dependent test for speaking performance.

2.3 Speaking Performance

Speaking refers to "a process of *oral language production*" (Tarone, 2005, p. 485), and speaking performance is defined as actual instances of producing oral language in real time (adapted from McNamara, 1996, p. 54). Speaking performance, or oral production, is a popular research target and has been assessed in many fields. For example, with the advent of task-based learning and teaching, a substantial amount of research has been conducted into task effects on speaking performance in second language acquisition (SLA) field (e.g., Robinson, 2001; Skehan & Foster, 2001). Speaking performance has been the focus not only in SLA research, but also in language testing (e.g., O'Loughlin, 2001; Wigglesworth, 1998) and in other research areas dealing with language use (Ortega, 2003). This section reviews previous literature from two perspectives: aspects of speaking performance and its assessment.

2.3.1 Aspects of Speaking Performance

In this section, Levelt's (1993) model of speaking process is first described, followed by various aspects of speaking performance from a few perspectives. One of the most influential models of the speaking process from a psycholinguistic perspective is Levelt's (1993), which is highly valued by many researchers (e.g., de Bot, 1992, p. 2; Ellis, 2005b, p.11; Katagiri, Komuro, Takayama, Takeda, & Takei, 1997, p. 76), and has been used as a basis for recent research (Bygate, 2001; de Bot, 1992, 1996; de Bot, Pribakht, & Wesche, 1997; Dörnyei & Kormos, 1998; Doughty, 2001; Towell, Hawkins, & Bazergui, 1996). Levelt (1989) summarized previous findings related to speaking and proposed a model from a psycholinguistic perspective. He then revised it in 1993 and included both speaking and listening processes. Since the current research focuses on speaking, only the relevant parts are described.

In Levelt's (1993) model, there are three stages of speech production: (a) conceptualization, (b) formulation, and (c) articulation, as seen in Figure 2.3. People form messages in the conceptualizer, and put them in a form of language in the formulator, which are pronounced and expressed in the articulator. In the second stage, the lexicon plays a crucial part. The lexicon contains all the information related to vocabulary and is divided into two constituents: lemmas[2.9] and lexemes.[2.10] The lemmas consist of information on meanings and syntax, while lexemes constitute information on morphology and phonology. After messages are formed in the conceptualizer, lemmas are searched for words, and grammatical structures are derived accordingly, which results in surface structures. Next comes phonological encoding, in which morphological and phonological information is incorporated by the use of lexemes. The resulting product is called a phonetic plan, articulatory plan, or internal speech, which is articulated later as overt speech. During the different processes of speaking, outcomes are monitored. Formed messages are checked according to their

appropriateness, and the internal and overt speeches are also monitored. Levelt stated that these processes are conducted in parallel and done automatically.

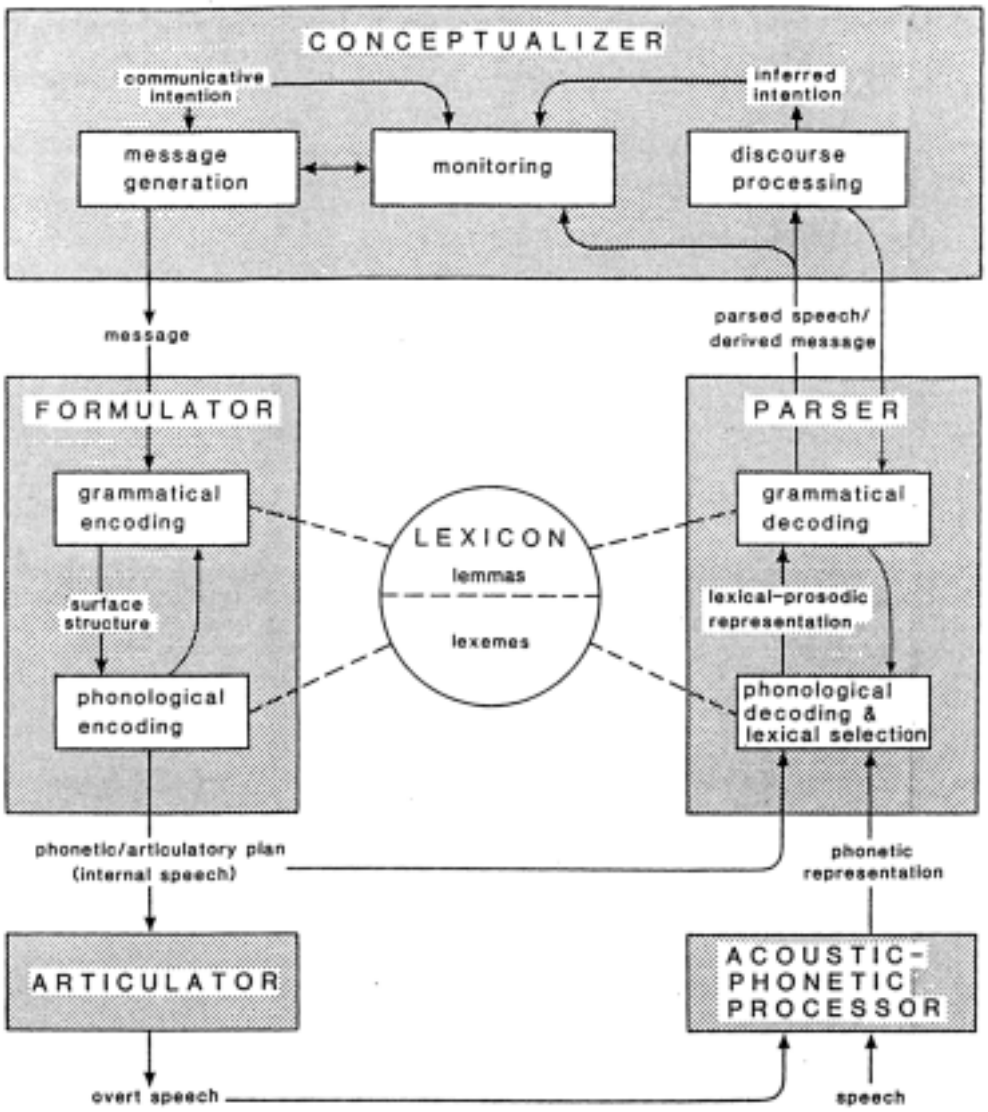


Figure 2.3. Schematic representation of the processing components involved in spoken language use. From "The architecture of normal spoken language use," by W. J. M. Levelt, 1993, In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook*, p. 2.

Levelt's (1993) model was developed in order to explain monolingual L1 speakers' speaking mechanisms without planning time. It has been used for L2 learners with some modification, and the following three problems seem very important. One obvious problem with Levelt is a lack of automaticity, especially in searching for vocabulary (de Bot, 1992, pp. 14-15). Another is in the use of monitoring. In addition to monitoring the internal and overt speech as well as formed messages, it seems that messages that are not encoded well in the formulator can be monitored and then reformulated (Dörnyei & Kormos, 1998, p. 538). Last, all the information on language knowledge, such as syntax and morphology, appears to exist in the lexicon in Levelt's model, but some aspects of language knowledge that can be abstracted as grammar may be located outside the lexicon, as Dörnyei and Kormos (p. 378) point out. Although this separation of grammar and vocabulary must be tested using clear definitions, it seems reasonable to assume that abstract rules exist which can be used for formulating grammatical utterances (Dörnyei & Kormos, p. 378).

While Levelt's (1993) model explicates the overall process of speaking performance, speaking can be divided into several types from four perspectives: (a) function, (b) interaction, (c) planning time, and (d) linguistic aspects. First, according to Tarone (2005), there are three types of speaking function: interactional, transactional, and ludic (p. 486). The purpose of interactional speaking is to "establish or maintain social relationships," whereas that of transactional speaking is to "convey information" (p. 486). The purpose of ludic speaking is to "amuse and entertain oneself or others" (p. 490). The second function (i.e., transactional) is the main focus in this study because the first type seems to be particularly related to the learner's personality and social skills in addition to language ability, and the third is "less common than" the other two

(p. 490). The second function was further divided into eight kinds: description, narration, instruction, comparison, explanation, justification, prediction, decision (Luoma, 2004, p. 31-32, which was based on Bygate, 1987; see Table 2.19).

Table 2.19
Functions of Speaking Performance

Function	Further division
Interactional	
Transactional	Factually oriented talk: description, narration, instruction, comparison Evaluative talk: explanation, justification, prediction, decision
Ludic	

Note. From Tarone (2005), Bygate (1987), and Luoma (2004, pp. 31-32). The distinction of interactional and transactional functions was originally proposed by Brown, T. and Yule (1983). They originally divided the transactional function into four: description, instruction, storytelling, and opinion-expressing (p. 107). Bygate (1987) expanded them into eight. Luoma (2004) labeled the eight as "factually oriented talk" and "evaluative talk."
Factually oriented talk is also called fact-oriented talk.

Second, speaking performance is divided into monologues and dialogues depending on whether there is interaction between the speaker and the listener (see Skehan, 2001).

Third, speaking performance is also separated into utterances with planning time (i.e., prepared speech) and without planning time (i.e., spontaneous speech). Strictly speaking, there are two types of planning: planning conducted before starting speaking (i.e., pre-task planning) and planning conducted during speaking (i.e., on-line or within-task planning; see Ellis, 2005c; Kawauchi, 2005).[2.11]

The main focus in this study is "description and comparison" in the "transactional function" and "monologues without pre-task planning time."[2.12] Description and comparison are selected among other types (see Table 2.19) for the following three reasons.

●First, "factually oriented talk" is more basic than "evaluative talk" because the latter

"involves the drawing of conclusions, usually requiring the expression of reasoning." (Bygate, 1987, p. 24). Since this study focuses on novice learners, the basic one (i.e., factually oriented talk) is used. In addition, "evaluative talk" assumes that there is some kind of content in the test taker's mind, which may not hold true for some test takers in this study. Thus, this study focuses on "factually oriented talk" and used tasks that presents pictures and gave content to talk about or asked test takers to speak about something very familiar (i.e., self-introduction). It should be noted that picture description is part of speaking based on the definition of speaking (i.e., a process of *oral language production*; see section 2.3).

●Second, description is defined as saying things "not in a temporal structure, but in terms of … physical attributes" (Brown, T. & Yule, 1983, p. 46; e.g., "describing other people, other people's houses, meals, clothes, books, films, household gadgets" p. 46). In contrast, narration is defined as storytelling with "a temporal structure" (Brown, T. & Yule, p. 46), which involves a "sequence of events," a setting, and "a particular time" (p. 39; e.g., "news reports on radio and television," p. 38). According to Brown, T. and Yule, a rather general guide to level of difficulty is that "straight description is easier, in some sense," than narration (p. 107), which was supported by the current pilot study. Thus, narration is excluded and description is selected. Although including "a temporal structure," which is a feature of narration, is too difficult, the function of connecting something with something else, or the discourse with cohesion and/or coherence seemed to be needed to be elicited in order to cover a wider area of speaking performance. Thus, comparison is selected, which is defined as saying things "in terms of … physical attributes" (p. 46), while connecting things with other things (e.g., comparing two similar pictures).

●Third, according to Brown, T. and Yule (1983), description "has a lot in common"

with instruction “from the point of view of the language used” (p. 46). Instruction is defined as “telling someone else how to do something” (Brown, T. & Yule, p. 46; e.g., “where to find a shop, how to get from the hotel to the railway station” p. 46). Instruction is excluded because the kind of instruction task that can be conducted by beginners seems to be limited to a task of giving street directions, which often occurs as an interaction, not as a monologue.

Monologues are selected because dialogues tend to be more affected by external factors (e.g., interviewer’s reactions, Brown, A., 2003, 2004). Since this study examines relationships between vocabulary knowledge (i.e., one of the internal factors) and speaking performance, restricting factors outside the investigation is considered appropriate.

Speaking performance without pre-task planning is selected for the following reason. This study examines relationships between productive vocabulary knowledge and speaking performance, and if there was planning time, it was expected that the process of producing words in the Productive VKT would be similar to that of producing words in speaking performance, which could lead to stronger relationships between the two. This is an empirical question that needs examination, but it is not the focus of this study. In order to investigate relationships between productive vocabulary knowledge and speaking performance more independently, planning time is not included. One thing to be noted is that extended monologues without planning time can be considered authentic because they can happen in everyday conversation when a person keeps on talking without much intervention from listeners.

The last perspective of categorizing speaking performance is linguistic aspects, which can be used for assessment categories. Speaking performance can be seen from

several linguistic aspects: fluency, accuracy, syntactic complexity, lexical complexity, quality and quantity of contents, cohesion and coherence, sociolinguistic appropriateness, use of communication strategies, pronunciation, and other aspects. Of the many aspects of speaking performance, fluency, accuracy, syntactic complexity, and lexical complexity are targeted in this study because of their recent extensive use and their importance as learning objectives (Skehan, 1996, 1998). Fluency and accuracy have been regarded as essential aspects of speaking performance and as having complementary roles (Brumfit, 1984). Skehan (1996, 1998) proposed that complexity should be investigated in addition to fluency and accuracy because the focus of the three is different. According to Skehan and Foster (2001), fluency is associated more with meaning and "getting a task done" (p. 190). In contrast, accuracy and complexity are more related to form, but accuracy is associated with use of "an interlanguage system of a particular level to produce correct, but possibly limited, language" (Skehan, 1998, p. 5), and complexity is more related to "hypothesis testing with recently acquired structures" and "language learning process of restructuring (Skehan & Foster, 2001, pp. 190-191). Complexity is often subdivided into syntactic complexity and lexical complexity (e.g., Wolfe-Quintero, Inagaki, & Kim, 1998; see section 2.3.2 for the definitions).

2.3.2 Assessment of Speaking Performance

Speaking performance is elicited using speaking tasks in a speaking test. Tasks refers to "what a test taker is required to do during a test or part of a test" and "more specifically, a type of test item involving complex performance in a test of productive skills" (Davies et al., 1999, p, 196). A speaking test is defined as a test in which oral utterances are elicited and rated (based on Davies et al., p. 182). They

include direct and semi-direct formats, either where a test taker talks to an interviewer or their peer(s), or where he or she speaks to and records his or her voice on a recorder (e.g., a tape recorder). It should be noted that this definition expels indirect paper-and-pencil tests.

In assessing speaking performance, two methods are often used: rating scales and speaking performance measures. The latter is used in the current study. Ellis (2003) summarized the three main methods of assessing speaking performance: (a) direct assessment of task outcomes, (b) discourse analytic measures, and (c) external ratings (p. 296). He maintains that in the first method (i.e., direct assessment of task outcomes), tasks are scored as either successful or unsuccessful based on "the outcome of the task" (Ellis, 2003, p. 296). As an example, he gives Robinson and Ross's (1996) task, in which test takers have to find a journal article of a particular topic in a library. In this task, if test takers find the article successfully, they are given a point, while if they do not find it or ask for help, they are not given a point. This method is similar to a rating scale of task achievement and can be included in (c) external ratings, which is usually called "rating scales." A rating scale is defined as "a scale for the description of language proficiency consisting of a series of constructed levels against which a language learner's performance is judged" (Davies et al., 1999, p. 153). There are two main types of rating scales: holistic and analytic rating scales. The holistic scale produces only one score that covers overall aspects of performance, whereas each analytic scale is constructed for each aspect (e.g., fluency), and after speaking performance is evaluated, more than one score is derived. A discourse analytic measure is defined as a measure that "provides counts of specific linguistic features occurring in the discourse that results from performing the task" (Ellis, 2003, p. 298). It is derived from quantifying target aspects in utterances and computing values that reflect

a certain dimension of language use (e.g., “the number of error-free clauses divided by the number of clauses” for accuracy).

The rating scales and discourse analytic measures have different characteristics (based on Ellis, 2003, pp. 296-302; Ellis & Barkhuizen, 2005, pp.139-164). First, if rating scales are used, rating can be easier and quicker, while if discourse analytic measures are used, transcripts have to be made, a process which involves more time and laborious work. Second, in using rating scales, the importance of the rater’s judgment on the speaking product is greater than in using discourse analytic measures, and use of rating scales requires more practice or training in order to learn the scoring criteria. In this sense, discourse analytic measures can be scored more objectively than rating scales. Third, another difference is the aspects of performance measured. What rating scales assess can be more like what people broadly feel in everyday life than the aspects that discourse analytic measures assess. In contrast, the discourse analytic measures may represent something distant from “how we judge communicative behaviour in the real world” (Ellis, 2003, p. 298), but the rather narrow but focused aspects of speaking performance can be assessed and small variations that rating scales cannot detect may be reflected in differences in values of the measures (Ellis, 2005b, p. 26). Fourth, constructing rating scales with higher validity, and examination of what rating scales assess (i.e., validation of rating scales) are difficult tasks. In the case of discourse analytic measures, the target aspects of performance are narrow and the meaning of the results is easy to understand.

While using both rating scales and discourse analytic measures are more beneficial, this study uses only discourse analytic measures for the following two reasons. First, since this study deals with only novice learners, the range of performance is narrower than when learners of a broad range of proficiency level are

used. Therefore, constructing rating scales and training raters can be much more difficult, and differentiating differences of participants' performance and getting varied results may be harder when rating scales are used.

Second, discourse analytic measures have been substantially utilized in assessing fluency, accuracy, and complexity in SLA research (e.g., Skehan & Foster, 2001) and more measures have been proposed. The discourse analytic measures are called "speaking performance measures" or "measures" in this study.

There are three types of computation method for speaking performance measures, as there are for writing ones (Wolfe-Quintero et al., 1998): (a) a frequency measure, (b) a ratio measure, and (c) "an index based on a formula that yields a numerical score" (p. 10). A frequency measure is "a simple frequency count of a particular feature, structure, or unit" (Wolfe-Quintero et al., p. 9; e.g., "the number of repetitions" in Foster, 1996). A ratio measure is one in which "the presence of one type of unit is expressed as a percentage of another type of unit, or one type of unit is divided by the total number of comparable units" (Wolfe-Quintero et al., p. 10; e.g., "the number of repetitions divided by total amount of speech" in Iwashita, McNamara, & Elder, 2001). An index measure is computed by counting the number of target features and putting them into a special formula (e.g., lexical variation index, or "the number of lexical word types minus number of lexical errors divided by the number of lexical words [i.e., open-class words, such as nouns and adjectives]," Engber, 1995, pp. 145-147). Since it is difficult to use the frequency measure when results across different tasks and studies are compared (Ortega, 1999; Wolfe-Quintero et al.) and the index measure has not often been utilized in the previous literature on speaking, the current study only deals with the ratio measure.

Fluency is defined as how fast and how much a learner speaks without

dysfluency markers (e.g., functionless repetitions, self-corrections, and false starts) in "coping with real time processing" (Wolfe-Quintero et al., 1998, p. 14). Fluency measures can be divided into two categories: "temporal" measures and measures of "a degree of freedom from various dysfluency markers" (Lennon, 1990b, p. 403). It should be noted that the term "fluency" does not represent the overall proficiency but only one aspect of speaking performance.

Accuracy refers to how much a learner speaks without errors in real-time communication (adapted from Wolfe-Quintero et al., 1998) and there are two types of associated measure: general measures and specific measures (e.g., Ortega, 1999; Skehan & Foster, 2001).

Complexity is defined as the degree to which a learner uses varied and sophisticated structures and vocabulary in speaking (adapted from Wolfe-Quintero et al., 1998) and is divided into syntactic complexity (also called grammatical complexity, syntactic maturity, and linguistic complexity) and lexical complexity. The latter is also called lexical richness and can be separated into three types: lexical diversity or lexical variation (as measured, for example, by the type token ratio, the Guiraud index, the mean segmental type-token ratio, and the Index of lexical diversity), lexical density (as measured, for example, by the number of lexical words per word), and lexical sophistication (as measured, for example, by the number of sophisticated word types per word).

The measures of fluency, accuracy, syntactic complexity, and lexical complexity were summarized in Table 2.20. Because there are a variety of types of measure, the measures that met all the following criteria were included in Table 2.20: ratio measures used in two or more empirical L2 speaking studies published since 1990, and measures that used the category of "fluency," "accuracy," or "complexity" except for "lexical

complexity," which tends to have several labels (e.g., lexical variation, lexical density, lexical sophistication, lexical richness). Even when the formula of measures is different but the meaning is the same, they were treated as the same measure. As seen in Table 2.20, various types of measure exist that are used to assess the same type of performance or the same construct (e.g., Ellis, 2005b; Ellis & Barkhuizen, 2005; Fujimori, 2004; Ortega, 1999). There are 14 measures of fluency, six measures of accuracy, four measures of syntactic complexity, and seven measures of lexical complexity. It seems that each measure assesses its own dimension of intended construct.

There are many studies that used speaking performance measures, but the quality of the measures has not been examined in much detail. There are at least two areas to be investigated, and these are dealt with in this study. First, as seen in Table 2.20, the number of words per unit (i.e., the unit length) has been interpreted as having two meanings: fluency (see F2 in Table 2.20) and syntactic complexity (see SC3 in Table 2.20). The unit length is derived from the formula "the number of words" divided by "the number of units." There are several units used in calculating the unit length, and these can be categorized into two types: a unit mainly related to pauses or dysfluency markers (e.g., unit defined by dysfluent pauses or hesitations, pausally defined unit, filler-free unit, dysfluency-marker-free unit, and utterance) and a unit mainly related to syntactic structures (e.g., T-unit, c-unit, AS-unit, and sentence). Table 2.21 shows studies that used the unit length, and they were categorized according to the two unit types. Broadly speaking, the number of words per unit mainly related to pauses or dysfluency markers is interpreted as fluency, whereas the number of words per unit mainly related to syntactic structures is interpreted as syntactic complexity.

Table 2.20
Summary of Speaking Performance Measures Used in Previous Studies

Measure	Recent source
Fluency	
No. of words per minute [F1]	Freed, Segalowitz, & Dewey (2004)
No. of words per unit [F2: Unit length]	Freed et al. (2004)[ABC]
No. of clauses per minute [F3]	Takiguchi (2004)
No. of syllables per minute	Kormos & Dénes (2004)
No. of syllables per unit	Kormos & Dénes (2004)[A]
No. of pauses per minute	Kormos & Dénes (2004)
No. of pauses per unit	Bygate (2001)[D]
No. of pauses per word	Freed (2000)[b]
Length of pauses / the total No. of pauses	Kormos & Dénes (2004)
Length of time spent speaking / length of total delivery time (including pause time)	Kormos & Dénes (2004)
No. of dysfluency markers per minute [F4]	Kormos & Dénes (2004)
No. of dysfluency markers per word [F5]	Freed (2000)
No. of dysfluency markers per unit [F6]	van Gelderen (1994)[E]
No. of clause containing dysfluency markers per clause	Wigglesworth (1998)
Accuracy	
No. of error-free clauses per clause [A1]	Kormos & Dénes (2004)
No. of error-free units per unit [A2]	Robinson (2001)[F]
No. of errors per word [A3]	Takiguchi (2004)
No. of errors per unit [A4]	Bygate (2001)[D]
Proportion of correct target features	Yuan & Ellis (2003, verbs)
No. of definite articles / No. of definite and indefinite articles	Wigglesworth (1998)
Syntactic Complexity	
No. of clauses per unit [SC1]	Yuan & Ellis (2003)[D]
No. of subordinate clauses per unit [SC2]	Mehnert (1998)[D]
No. of S-nodes per unit	Mehnert (1998)[D]
No. of words per unit [SC3: Unit length]	Ortega (1999)[H]
Lexical Complexity	
Type token ratio (TTR): No. of different word types / No. of words	Daller, van Hout, & Treffers-Daller (2003)
Guiraud index: No. of types / the square root of No. of words [LC1]	Daller et al. (2003)
Mean segmental type-token ratio (MSTTR)	Yuan & Ellis (2003)
Index of lexical diversity (D)	Kormos & Dénes (2004)
Lexical density: No. of lexical words per word [LC2]	Vermeer (2000)[c]
Weighted lexical density: ([No. of sophisticated lexical words] + [No. of basic lexical words]*0.5) / No. of words [LC3]	Mehnert (1998)
No. of sophisticated word types per word [LC4]	Daller et al. (2003)

Note. [ ] = Measures used in the current study; / = divided by; [a]Divided by total amount of speech; [b]Per 100 words; [c]Function words per word. The units used in the relevant study: [A]Pausally defined unit; [B]Filler-free unit; [C]Dysfluency-marker-free unit; [D]T-unit; [E]Tone group; [F]c-unit; [G]Clause; [H]Utterance. F2 and SC3 are the same measures. See Notes 2.13 and 2.14 for more information on the Guiraud index and the Index of lexical diversity (D). See Appendix 2.1 for a comprehensive summary.

Table 2.21
Summary of Previous Studies That Used the Unit Length (No. of Tokens per Unit [F2 or SC3])

Interpretation	Unit mainly related to pauses or dysfluency markers	Unit mainly related to syntactic structures
Fluency [F2]	Freed (2000),[a] Freed, Segalowitz, & Dewey (2004),[bcd] Lennon (1990b),[b] Robinson (1995),[b] Segalowitz & Freed (2004),[bcd] Shimada (2004),[b] Robinson, Ting, & Urwin (1995)[e]	Ano (2002),[i] Kimura (1995),[g] Niwa (2000, as cited in Robinson, 2005),[f] Robinson (2001)[g]
Syntactic Complexity [SC3]	Ortega (1999)[e]	Bygate (2001),[f] Egusa & Yokoyama (2004a),[g] Fujimori (2004),[h] Kawauchi (1998, 2005),[f] Kamimoto & Kawauchi (2000),[f] Kawashima (2004),[f] Kawauchi & Kamimoto (2000),[f] Koizumi & Kurizaki (2002),[f] Mehnert (1998),[g] Takiguchi (2003, 2004),[h] Van den Branden (1997)[f]

Note. Units mainly related to pauses or dysfluency markers: [a]Unit defined by dysfluent pauses or hesitations; [b]Pausally defined unit; [c]Filler-free unit; [d]Dysfluency-marker-free unit; [e]Utterance (unit defined by intonation and pauses, Ortega, 1999). Units mainly related to syntactic structures: [f]T-unit; [g]c-unit; [h]AS-unit, [i]Sentence.

However, there are five exceptions: Ano (2002), Kimura (1995), Niwa (2000, as cited in Robinson, 2005), Robinson (2001), and Ortega (1999). This study focuses on the unit length using a unit mainly related to syntactic structures, especially, a T-unit, a c-unit, and an AS-unit because it has been used widely and its meaning has been controversial.

A T-unit and a c-unit are very similar to an AS-unit (Analysis of Speech Unit), which is proposed for analysis of spoken texts. Specifically, an AS-unit is a modified version of a T-unit or a c-unit, which both had some ambiguity in their definitions (Foster, Tonkyn, & Wigglesworth, 2000). An AS-unit refers to "a single speaker's utterance consisting of *an independent clause, or sub-clausal unit,* together with any *subordinate clause*(s) associated with either" (Foster et al., p. 365). The following sentences are all one AS-unit.

|Hello.| |Open the door.| |I like English.|

|I like tennis and play it every day.| |I study English because I like it.|

(The latter two examples may be considered to contain two AS-units when they have particular patterns of intonation and/or pauses. See Foster et al., 2000, for details)

In relation to the unit length using a T-unit, a c-unit, and an AS-unit, some argue that the unit length means fluency because the measure reflects the "speed of lexical and syntactic access" (Bygate, 2001, p. 34). By contrast, others consider the same measure (i.e., the unit length) a syntactic complexity measure because "longer production units are assumed to indicate the presence of more complex structures" (Wolfe-Quintero et al., 1998, p. 14) and the measure represents "the extent to which lexical accessing can be managed according to basic syntactic parameters" (Bygate, pp. 34-35). The problem of a single measure having two meanings is very serious because the interpretation varies depending on which meaning a user believes in. Two empirical studies (Mehnert, 1998; Takiguchi, 2003) examined this problem using principal component analysis and factor analysis and they supported the view of syntactic complexity. However, their findings were based on a small number of participants for the use of multivariate analyses (N = 31 in Mehnert, using c-units; N = 17 in Takiguchi, using AS-units)[2.15] and thus need to be further examined. This interpretation of the AS-unit length being a syntactic complexity measure is posed as a hypothesis in this study (see sections 5.1 and 11.1). The present research used an AS-unit as a basic unit of analysis and examined the meaning of "the AS-unit length."

Second, speaking performance and thus results of speaking performance measures can vary according to a great variety of factors, such as tasks, a test taker's proficiency, real-time processing, and other individual variables (Fulcher, 2003). However, it seems that there are no studies that have examined the magnitude and the relative effect of each factor on speaking performance measures. The current research focused on effects of tasks, or task generalizability. Generalizability is defined as the

degree to which interpretations based on test performance can be generalized across target groups, tasks, and other settings and is an essential aspect of validity (e.g., Messick, 1996; see section 2.5). In other words, to what extent do tasks affect speaking performance? If they influence speaking performance to a large degree and the results of speaking performance measures change substantially across tasks, this means that the measures lack generalizability across tasks. This is a serious problem if users of measures want to generalize their results beyond the tasks they used or if test takers who take different tasks are compared. This perspective of generalizability of measures is one aspect of validity (e.g., Messick, 1989, 1996). Although there have been few previous studies, two hypotheses were made that examined relationships of the same measure between tasks: "The speaking performance measures have strong correlations with the same measures of different tasks" in Study 1 (1B) and "Each measure across tasks loads on the same component." in Study (2D). The statements mean the same, but use different phrases due to the different analytical methods used.

2.4 Relationships Between Vocabulary and Speaking Performance

This section reviews studies that dealt with relationships between vocabulary and speaking. Four points need to be made explicit before the review of previous studies. First, the studies reviewed were the ones that analyzed language that was produced orally first. Thus, the following studies were not included: studies that analyzed language that was produced first in a written text, which was then used as a spoken script (e.g., Hidai, 1999; Nomura, 2005). Second, despite the fact that each study interpreted speaking performance in its own way (e.g., speaking ability or BICS [Basic interpersonal communication skills] type speaking ability), all are reviewed as "speaking performance" because this interpretation had the fewest assumptions. Third,

as for types of vocabulary, the term “productive” was only used when test takers were required to produce a form. Otherwise, vocabulary targeted was considered to be receptive. Fourth, although vocabulary can be seen from a multi-word level (e.g., Ota, Kanatani, Kosuge, & Hidai, 2003, Chapter 3; Read, 2000; Schmitt, 2004; Wray, 2002) as well as an individual level, only relationships between vocabulary at the latter level and speaking are reviewed.

There is a theoretical background to the relationship between vocabulary and speaking performance. In the theoretical model of L1 and L2 speaking (Levelt, 1989, 1993), as seen in section 2.3.1, vocabulary has a central position in forming an utterance with appropriate meanings and with syntactic, morphological, and phonological structures. Levelt’s model suggests two points. First, vocabulary is always required in the formulation stage. In other words, no speech can be produced without vocabulary, and vocabulary is indispensable to speaking performance. Second, the lexicon consisting of lemmas and lexemes includes not only vocabulary size (i.e., primary meaning and form [phonology]) but also depth (i.e., syntax and morphology), which suggests that both size and depth are related to speaking performance. This study attempts to examine the degree of relationships between vocabulary and speaking performance.

In addition to the theoretical importance, several empirical studies have been conducted that investigated relationships between vocabulary and speaking. They can be classified into two types: studies that dealt with (a) vocabulary used in speaking performance, and (b) relationships between vocabulary knowledge and speaking.

2.4.1 Vocabulary Used in Speaking Performance

In this section, studies that examined vocabulary used in speaking performance

are described. Some studies examined characteristics of vocabulary in the corpus of spoken discourse. For example, Adolphs and Schmitt (2003) reported that approximately 95% of words in the utterances belonged to 2,000 word families or 3,000 individual words (i.e., lemmas) in the CANCODE (Cambridge and Nottingham Corpus of Discourse in English) and the British National Corpus (BNC). Adolphs and Schmitt (2004) further demonstrated that 2,000 word families covered more than 90% of spoken words across contexts (i.e., pedagogic, intimate, socio-cultural, professional, and transactional) although there were some small variations between the contexts. Their studies suggest that a relatively small number of words are needed for speaking performance.

Other studies examined relationships between some learner characteristics and vocabulary in the spoken corpus. Durán, Malvern, Richards, and Chipere (2004) examined relationships between ages (from 18 months to 60 months) and lexical diversity (n = 15 to 30). They found a consistent rise in lexical diversity as children (L1 learners) got older. In relation to language proficiency, Read (2005) and Iwashita (2005) reported a trend of test takers with higher proficiency producing better vocabulary use in speaking performance (e.g., lexical diversity and the number of tokens). Kimura (2004) analyzed spoken texts in the NICT JLE Corpus (National Institute of Information and Communications Technology Japanese Learner English Corpus; Izumi, Uchimoto, & Isahara, 2004). She demonstrated that the qualities of vocabulary used (i.e., what types of vocabulary were used) were different across oral proficiency levels. For instance, the proportions of nouns and lexical (i.e., content) words used were greater than those of verbs and function words at lower levels, whereas at intermediate levels, the trend of the proportions became inverted, followed by an increase of lexical variety at advanced levels (pp. 140-143). Kaneko (2004) also

used the NICT JLE Corpus and found that test takers at intermediate and advanced levels tended to use the most frequent 100 words in the JACET List of 8000 Basic Words (JACET8000, hereafter; JACET Basic Words Revision Committee, 2003) more frequently (51.51 to 56.61%) than test takers at novice levels (39.35 to 46.78%). In other words, the proportions of sophisticated words (i.e., words not in the list of the 100 most frequent words) decreased as proficiency got higher. These studies suggest that as learners gain more proficiency, they are likely to increase the qualities of their vocabulary use, but not sophisticated word use.

The roles of vocabulary used in speaking performance as compared to other qualities (e.g., accuracy) were also examined in two main studies. First, Adams (1980) examined speaking factors that separated neighboring level groups using discriminant analysis among workers in foreign affairs and their families. It was found that out of five factors (i.e., accent, comprehension, fluency, grammar, and vocabulary), vocabulary was the only one that affected level difference between learners at novice and intermediate levels. For learners at intermediate levels and above, several of the factors affected level differences (see Table 2.22 for the summary of crucial factors affecting level differences). Although Adam's study has often been cited (e.g., Fulcher, 2003, p. 183), the number of participants was small at the lower levels (n = 7 for Novice Low to Novice High, p. 2), so precautions should be taken when one interprets his results.

Table 2.22
Discriminating Factors That Affected Level Differences in Adams (1980)

FSI Level difference	Factors in descending order of significance
0+ to 1 (Novice High to Intermediate Low & Mid)	Vocabulary
1 to 1+ (Intermediate Low & Mid to Intermediate High)	Fluency, Comprehension, Grammar, Vocabulary
1+ to 2 (Intermediate High to Advanced)	Comprehension, Grammar, Accent, Fluency
2 to 2+ (Advanced to Advanced Plus)	Fluency, Comprehension, Accent, Vocabulary
2+ to 3 (Advanced Plus to Superior)	Grammar, Accent, Vocabulary, Comprehension
3 to 3+ (Superior)	Comprehension, Fluency, Grammar
3+ to 4 (Superior)	Vocabulary, Accent, Grammar
4 to 4+ (Superior)	(Grammar, Vocabulary: Both nonsignificant)

Note. Adams (1980, p. 4) with minor modification. FSI = Foreign Service Institute; ( ) = corresponding ACTFL (American Council for the Teaching of Foreign Languages) Levels (Fulcher, 2003, p. 15).

Second, based on Adams (1980), Higgs and Clifford (1982) made a model of the relative importance of various elements (i.e., vocabulary, grammar, pronunciation, fluency, and sociolinguistics) to overall ability. In order to examine the model empirically, they asked 50 teachers for their opinions on how each element affects speaking proficiency at each proficiency level. Higgs and Clifford showed that their model was similar to the teachers' perceptions and that vocabulary contributed the most to speaking proficiency at a novice level (see Table 2.23). One problem concerning their study is that they only examined the teachers' perceptions instead of

Table 2.23
Factors That Affected Speaking Proficiency at Each Proficiency Level in Case of All Languages in Higgs & Clifford (1982)

FSI Level	Factors in descending order of significance
0 (Novice)	Vocabulary, Pronunciation, Grammar, Fluency, Sociolinguistic
1 (Intermediate)	Vocabulary, Pronunciation, Grammar, Fluency, Sociolinguistic
2 (Advanced)	Grammar, Vocabulary, Pronunciation, Fluency, Sociolinguistic
3 (Superior)	Grammar, Vocabulary, Sociolinguistic, Fluency, Pronunciation
4 (Superior)	Grammar, Vocabulary, Sociolinguistic, Fluency, Pronunciation
5 (Superior)	Fluency = Sociolinguistic, Grammar = Vocabulary, Pronunciation,

Note. From Higgs & Clifford, 1982, p. 71) with simplification and modification. FSI = Foreign Service Institute; ( ) = corresponding ACTFL (American Council for the Teaching of Foreign Languages) Levels (Fulcher, 2003, p. 15).

having teachers rate learners' utterances from interviews (Magnan, 1988, p. 274). De Jong (2004) replicated Higgs and Clifford's (1982) model and obtained favorable results overall.

The results of Adams (1980) and Higgs and Clifford (1982) seem to indicate close relationships, and closer ones between vocabulary as part of overall speaking performance and overall speaking performance at low levels than at intermediate and advance levels. The strong associations between vocabulary as part of overall speaking performance and overall speaking performance at low levels were also supported by the following three studies that target Japanese learners of English at the novice level. First, Koizumi and Kurizaki (2002) found strong associations (r_s = .69 to .80, n = 76, p. 24) between speaking test scores of Japanese junior high school students and the number of words that they uttered on the speaking test in self-introduction and picture description tasks of monologue. Second, Takiguchi (2003) conducted principal components analysis using holistic speaking scores and speaking performance measures of Japanese junior high school students in interactive tasks. As a result, the holistic scores and the number of words uttered for one minute loaded on the same factor (p. 50), which suggests a strong association between speaking performance and vocabulary. Third, Usuda (2002) found a strong correlation (r = .79, n = 16, p. 50) between the number of tokens uttered by Japanese senior high school students and their overall speaking test scores in picture description tasks of monologue. Although the three studies consistently indicate the strong relationships between "vocabulary use in speaking performance" and "overall speaking performance" at the novice level, there is a question of what consists of "overall speaking" and whether their "overall speaking" can be generalized. These concerns seem to point to the necessity of further theoretical and empirical studies.

2.4.2 Relationships Between Vocabulary Knowledge and Speaking Performance

In this section, studies are reviewed that deal with relationships between vocabulary knowledge and speaking performance and that assess vocabulary knowledge and speaking using different tests. There are nine studies that have investigated this topic.

First, in Vermeer (2000), there were almost zero to moderate relationships of lexical aspects of speaking performance (e.g., the number of types and the Guiraud index) with receptive vocabulary (r = -.19 to .53, n = 146, p. 73). Vermeer's participants were Dutch native (L1) and ethnic minority (L2) children at kindergarten, and they took two receptive vocabulary tests: (a) a receptive vocabulary test of selecting one picture out of four in response to orally presented stimulus words and (b) a definition test of "explaining or describing the meaning of a given word" (p. 71). The children were also interviewed. During the interview, they told a story from a picture book and their utterances were analyzed.

Second, Ukrainetz and Blomquist (2002) found that among American children (L1; 3 to 6 years old) there were some relationships of the number of lemmas and tokens and the mean length of utterance with receptive vocabulary test scores (r = .12 to .61, n = 28, p. 69) and with productive vocabulary scores (r = .25 to .48, n = 28, p. 69). In the receptive vocabulary test, children were required to select the corresponding pictures from a choice of four when the examiner said a word. In the productive vocabulary test, test takers needed to provide a name, synonym, or category name when the examiner "pointed to a picture silently or with a word label" (pp. 65-66). They also provided "conversational discourse" (e.g., talking about animals in a toy farm) mainly (p. 66), which was used for the analysis.

Third, Durán et al. (2004) examined relationships between receptive vocabulary scores and lexical diversity (the Index of lexical diversity [D]; see Table 2.20) at 39 months for children in England (L1 learners), and showed a weak correlation (r_s = .218, n = 26, p. 233). The test used was called the English Picture Vocabulary Tests (Brimer & Dunn, 1963, as cited in Durán et al., p. 233) but no information regarding item types was presented in the article. The oral sample analyzed was derived through recordings of natural talk at home (p. 227).

Fourth, in Ishizuka (2000), there was a moderate correlation (r = .43, n = 26, pp. 15-18) between (a) receptive vocabulary depth test scores and (b) interview test scores among Japanese senior high school students. The receptive vocabulary depth test was constructed using Read's (1993) format, which was intended to assess synonyms, collocations, analytic relationships (e.g., *edit* and *publishing*). For the interview tests, the Society for Testing English Proficiency Test in Practical English Proficiency (STEP Test; STEP, 2004) was employed, and scores from rating scales were combined (p. 16) although the details were not described in the paper.

Fifth, Ota et al. (2003, Chapter 4) showed that there were relationships to some degree between receptive vocabulary knowledge and changes in speaking performance (n = 4) in a longitudinal study. Their study suggests that changes in speaking performance of Japanese junior high school students are affected not only by receptive vocabulary knowledge but also by other individual characteristics (e.g., a tendency to focus on studying at cram school intensively rather than at school, and a lack of motivation). In their study, receptive vocabulary knowledge was assessed using a Vocabulary Size Test for Japanese Learners of English (Mochizuki, 1998), which utilizes a multiple-choice format (see section 3.2.2.1 for more information). The speaking performance targeted was interactive interview with a foreign teacher, and

analytic rating was conducted using scales consisting of Fluency, Sentence (Grammatical accuracy), and Pronunciation, (sometimes plus Content; p. 101).

Sixth, Nomura (2004) found that among Japanese senior high school students (n = 38 or 39; calculated based on the reported percentages), there was a relationship between receptive vocabulary test scores and speaking test levels (p. 64). The receptive vocabulary test was Nation's (1990, as cited in Nomura) Vocabulary Levels Test, whereas topics in the speaking tests varied for each period of time (e.g., school and international marriage), but the format involved only talking with a foreign teacher (pp. 64-65). Overall speaking performance was assessed using holistic rating scale scores (p. 63).

Seventh, Yoshikawa (2005) examined relationships between "non-linguistic variables, including affective, cognitive, meta-cognitive, and linguistic variables and two types of speaking" (p. 53). She found that vocabulary test scores were not significantly retained in regression analysis and that variables affecting speaking performance were two affective (i.e., intrinsic motivation and willingness to communicate), two cognitive (i.e., self-evaluation of speaking ability and expectation), two meta-cognitive (i.e., two types of meta-cognition), and four linguistic (i.e., grammar, coherence, reading comprehension, and organization) variables. This result suggests that vocabulary is not related much to speaking performance (n = 22, p. 54). The two speaking tasks were intended to elicit narrative and argumentative discourse and composite scores were used of analytic rating scales consisting of Content, Communicability, Intelligibility, Educated level, and Segmental/super-segmental or discourse organization (p. 55).

Eighth, Uenishi (2005) conducted a series of studies for Japanese senior high school and university students that investigate factors affecting speaking performance.

His results were all consistent in that receptive vocabulary was not significantly related to speaking performance (β = -.055 to .093, n = 36 to 70) in the following factors: listening, (receptive) vocabulary, grammar, comprehensive English ability, and non-linguistic factors derived from factor analysis (sometimes plus motivation and learning anxiety). The receptive vocabulary was assessed using the STEP vocabulary items for senior high school students and Nation's (2001) Vocabulary Levels Test for university students. As for the speaking performance, the students were required to narrate a series of pictures and "talk freely about their hobbies or what they like in their daily life" (p. 81). Raters evaluated Content, Cohesion, Utterance, pattern, Fluency, Correctness, Understandability, Pronunciation, and Number of delivery words, and the composite scores from analytic rating scale scores were analyzed (p. 34)

Ninth, Koizumi (2005d) reported strong relationships of productive vocabulary knowledge with overall speaking performance (r = .77, n = 138, p. 20) and almost zero to strong correlations with lexical aspects of speaking performance (e.g., the number of types uttered, the Guiraud index; r = -.02 to .80, n = 138, p. 20). The analytic rating scales for assessing speaking performance consisted of Task fulfillment, Vocabulary volume, Accuracy, and Fluency (p. 5), and the overall speaking logit scores were derived from Rasch analysis (see section 2.6.4.2 for the analysis). Her study used participants overlapping with the present study, and the research target of the two studies was not the same. Some (n = 73) of the participants in Koizumi were the same, but the participants in the two studies were different in that Koizumi included those who did not utter clauses as well (see sections 3.2.2.1 and 3.2.2.4 for the tests used in Koizumi).

Table 2.24 provides a summary of types of participants, target vocabulary and speaking performance, and the results. What has been found so far seems to be that

Table 2.24
Characteristics of the Previous Studies and the Present Study

	Participants	Vocabulary targeted	Aspects of speaking performance targeted	Result
Vermeer (2000) [cross-sectional]	L1 and L2 learners (*n* = 146)	Receptive	Lexical aspects (e.g., the number of types and the Guiraud index)	*r* = -.19 to .53
Ukrainetz & Blomquist (2002) [cross-sectional]	L1 learners; children (*n* = 28)	Receptive and Productive	The number of types and tokens, the mean length of utterance	Receptive: *r* = .12 to .61; Productive: *r* = .25 to .48
Durán et al. (2004) [cross-sectional]	L1 learners; children (*n* = 26)	Receptive	Lexical diversity	r_s = .218
Ishizuka (2000) [cross-sectional]	L2 learners; SHS (*n* = 26)	Receptive	Overall [composite scores of analytic rating scales]	*r* = .43
Ota et al. (2003) [longitudinal]	L2 learners; JHS (*n* = 4)	Receptive	Overall [composite scores of analytic rating scales consisting of Fluency, Sentence, and Pronunciation, (sometimes plus Content)]	relationship to some degree
Nomura (2004) [longitudinal]	L2 learners; SHS (*n* = 38 or 39)	Receptive	Overall [holistic rating scale scores]	relationship to some degree
Yoshikawa (2005) [cross-sectional]	L2 learners; university (*n* = 22)	(not described)	Overall [composite scores of analytic rating scales consisting of Content, Communicability, Intelligibility, Educated level, and Segmental/super-segmental or discourse organization]	non-significant
Uenishi (2005) [cross-sectional]	L2 learners; SHS and Univeristy (*n* = 36 to 70)	Receptive	Overall [composite scores of analytic rating scales consisting of Content, Cohesion, Utterance, pattern, Fluency, Correctness, Understandability, Pronunciation, and No. of delivery words]	β = -.055 to .093
Koizumi (2005d) [cross-sectional]	L2 learners; JHS and SHS (*n* = 138)	Productive	Overall [composite (logit) scores of analytic rating scales consisting of Task fulfillment, Vocabulary volume, Accuracy, and Fluency] and lexical aspects (e.g., lexical diversity)	*r* = .77 *r* = -.02 to .80
The present study [cross-sectional]	L2 learners; JHS and SHS (*n* =73 to 225)	Productive	Fluency, accuracy, syntactic complexity, and lexical complexity	

Note. JHS = junior high school students; SHS = senior high school students; university = university students. [ ] = Research design.

some relationships tend to exist, for example, between vocabulary and overall speaking performance (e.g., Ishizuka, 2000; Koizumi, 2005d) and between vocabulary and lexical aspects (e.g., Durán et al., 2004; Vermeer, 2000), and that the results of these relationships vary substantially according to the aspects of speaking performance examined and measures used. For example, Ukrainetz and Blomquist's (2002) results demonstrate that speaking performance is not always more strongly related to productive vocabulary (r = .25 to .48) but is sometimes more related to receptive vocabulary (r = .12 to .61). Thus, it is difficult to interpret the results and make a hypothesis for the current study.

In addition, each study seems to suffer from the following four problems, and the unstable results of the previous studies may be due to some of these problems.

●First, there have been no studies that systematically examined relationships between both size and depth of productive vocabulary knowledge and speaking performance. Linguistic aspects of speaking performance examined so far seems to be limited to overall speaking performance or the lexical aspects. One advantage of including other aspects in the investigation is that in addition to the possibility of grasping a wider picture of relationships, associations are revealed with lexical aspects in comparison to other non-lexical aspects in a relative sense.

●Second, a question remains of whether their speaking test scores reflected overall speaking performance and whether their "overall speaking" can be generalized. This is because while aspects included in their speaking test scores are clear, each included rather different aspects of speaking performance in their scores, and results may have

been influenced by the aspects included (i.e., vocabulary volume and fluency). For instance, Uenishi (2005) included a wide range of speaking performance aspects (i.e., Content, Cohesion, Utterance, pattern, Fluency, Correctness, Understandability, Pronunciation, and Number of delivery words), whereas Ota et al. (2003) assessed a relatively narrow range of aspects (i.e., Fluency, Sentence, and Pronunciation). However, due to the absence of an empirical model of overall speaking performance, it seems rather difficult to deal with the notion of "overall speaking performance" at present. This is always the problem when overall speaking performance is assessed (see also section 2.4.1), but this can be overcome if target aspects are focused on a limited range of aspects, such as fluency and accuracy.

●Third, some studies lack sufficient information on construct(s) of vocabulary or speaking tests, on types of item or task used, and on validity (including reliability) of tests or measures. This is crucial information for interpreting results because, for example, if reliability of tests is low, weak correlations may have occurred due to the low reliability rather than the absence of relationships. Strictly speaking, even when some tests have gone through a validation process, when the target learners are not the same, some degree of examination of validity needs to be made. This problem applies to all the areas in SLA, and according to Ellis (2005a, p. 168) and Douglas (2001), SLA researchers have not attempted to show validity and reliability of their tests.

●Fourth, the number of participants in the previous studies was rather limited as a representative sample from the population, and some studies (e.g., Nomura, 2004; Ota et al., 2003) analyzed only students from one school. In addition, some studies (i.e., Uenishi, 2005; Yoshikawa, 2005) used multivariate analyses for a small sample (n = 36 to 70 using principal component analysis (PCA), multiple regression analysis (MRA), and structural equation modeling (SEM) in Uenishi; n = 22 using MRA and SEM in

Yoshikawa) when many variables are included.[2.16] Thus, precautions should be taken when one interprets the results.

Due to the varying extent of results and the problematic nature of the previous studies, further studies are needed with a modification of design. In order to obtain more interpretable results, it is necessary to examine speaking performance from multiple perspectives (response to the first problem), to utilize multiple tests or measures with detailed descriptions and with more validity (response to the third problem), and to collect more participants that represent the population more precisely (response to the fourth problem). The present study attempts to include these aspects. This study does not deal with overall speaking performance because of the difficulty of dealing with this matter at the moment. Instead, the current study examines relationships between productive vocabulary knowledge and speaking performance, and the specific focus is given to (a) size and depth aspects of vocabulary because of their importance in vocabulary studies and (b) fluency, accuracy, syntactic complexity, and lexical complexity of speaking performance (response to the first and second problem).

2.5 Validity and Validation

Validity is considered the most important aspect (e.g., AERA/APA/NCME, 1999, p. 9; Alderson, Clapham, & Wall, 1995; Bachman, 1990, p. 236) and it should always be examined in test development and use (Chapelle, 1999). Traditionally, validity has been defined as the degree to which a test measures the trait it intends to measure (e.g., Henning, 1987). It has been subdivided into three main types: content validity, criterion-related validity, and construct validity (e.g., Kane, 2001; Messick, 1989;

Shepard, 1993, for the history of validity).

Recently, validity has been considered a unitary concept (Messick, 1989, p. 14) with several approaches to the validity argument (Messick, 1995, 1996; Chapelle, 1999), rather than several types of validity. With the concept of unitary validity in mind, validity is defined as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (Messick, 1989, p. 13). It is also defined as "the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (Messick, 1989, p. 13).

Validation refers to "the process of establishing the validity" (Davies et al., 1999, p. 220). The current view of validity is that the target of validation is not the tests but the interpretation and uses based on test scores (e.g., Chapelle, 1999, p. 258; Kane, 2001; Luoma, 2001; Messick, 1996, p. 245). Many researchers have presented methods to conduct validation or the validation framework (e.g., Bachman, 2005; Bachman & Palmer, 1996; Borsboom, Mellenbergh, & van Heerden, 2004; Chapelle, 1999; Chapelle, Enright, & Jamieson, 2004; Hasselgren, 1998; Kane, 2001; Messick, 1989, 1996; Weir, 2005; Zumbo, 2005b). The present study uses Messick (1996) for two reasons. First, it is based on Messick (1989), which describes validity in detail, and by using Messick (1996), the theory of validity as a unitary concept and the practice of validation become conceptually coherent. Second, the validity procedures are clear thanks to Chapelle (1999) and Messick (1989, 1996). His framework has been used in several studies (e.g., Chapelle, 1994; Guerrero, 2000; Hasselgren, 1998; In'nami, 2004; Miller & Linn, 2000).

According to Messick (1989, 1996), there are six aspects of validity. Although

validity is a unitary concept, “unified validity” was separated “into several distinct aspects to underscore issues and nuances that might otherwise be downplayed or overlooked” (Messick, 1996, p. 248). Table 2.25 explicates the target elements to be examined and a sample of questions to be asked for each aspect as well as methods that can be used for validation of each aspect. Methods used in this study are summarized in Table 2.26, where some possible methods were added by the author.

Table 2.25
Validation Framework Using Six Aspects of Validity in Messick (1996)

Aspect	■Element to be examined □Examples of questions to be asked	･Methods to be used
The content aspect [Content analysis]	■The content of the test in relation to the content of the domain of reference □Are the boundaries of the construct domain to be assessed specified? □Is there evidence for the hypothesized match between test items or tasks and the construct that the test is intended to measure? (Chapelle, 1999, p. 260) □Is there evidence of content relevance and representativeness as well as of technical quality (e.g., appropriate reading level, unambiguous phrasing and correct keying)?	･Job analysis, task analysis, curriculum analysis, and especially domain theory ･ Expert professional judgement ･ Use of analytical frameworks (Banerjee & Luoma, 1997, pp. 281-282) ･Discourse analysis (Fulcher, 2003, p. 195) (see also Alderson, Clapham, & Wall, 1995, pp. 173-176, 183, 193 for other methods)
The substantive aspect [Empirical item or task analysis]	■Ways in which individuals respond to the items or tasks □Are there theoretical rationales for the observed performance regularities and item correlations, including process models of task performance along with empirical evidence that the theoretical processes are actually used by respondents in the assessment tasks? □Do tasks provide appropriate sampling of domain processes? □Is there empirical evidence that the ostensibly sampled processes are actually used by respondents in task performance? □Is there empirical evidence of response consistencies or performance regularities reflective of domain processes? □To what degree do relevant factors affect item difficulty and discrimination? (Chapelle, 1999, p. 261; e.g., Anderson, Bachman, Perkins, & Cohen, 1991) □What does a test appear to measure to the untrained eye?	･Quantitative analyses (e.g., verbal report, observation, questionnaires, & interviews; Banerjee & Luoma, 1997, pp. 227-280) ･Documenting the strategies and language that learners use as they complete test tasks (Chapelle, 1999, p. 261) ･ Discourse analysis by focusing on the language produced by the test taker (Chapelle, 1999, p. 261; see also Banerjee & Luoma, 1997, pp. 280-281) (see also Alderson et al., 1995, pp. 176-177, 193 for other methods)

(Table 2.25 continues)

(Table 2.25 continued)

Aspect	■Element to be examined □Examples of questions to be asked	･Methods to be used
The structural aspect [Dimensionality analysis]	■Relationships among responses to the tasks, items, or parts of the test, that is, internal structure of test responses □Does the theory of the construct domain guide the rational development of construct-based scoring criteria and rubrics? □Is the internal structure of the assessment (e.g., score scales, scoring models) consistent with what is known about the internal structure of the construct domain? □To what extent is the observed dimensionality of response data consistent with the hypothesized dimensionality of a construct? Does the test response data fit to a psychometric model? (Chapelle, 1999, p. 261)	･ Classical true-score reliability methods and item response theory (IRT) methods (Chapelle, 1999, p. 261) ･Factor analysis, & structural equation modeling (SEM; Bachman & Eignor, 1997, pp. 233-237) (see also Alderson et al., 1995, pp. 183-185, 194 for other methods)
The generalizability aspect [Differences in test performance]	■Differences in these test processes and structures over time, across groups and settings, and in response to experimental interventions (e.g., instructional or therapeutic treatment and manipulation of content, task requirements, or motivational conditions) □To what extent do score properties and interpretations generalize to and across population groups, settings and tasks, including generalizability of test-criterion relationships across settings and time periods? □Does the assessment provide representative coverage of the content and processes of the construct domain? □Is there evidence that the score interpretation is not limited to the sample of assessed tasks but is generalizable to the construct domain more broadly?	･Reliability, generalizability theory as well as construct theory in showing consistency of performance across tasks, occasions, and raters of a particular assessment ･ Differential item functioning (DIF; e.g., Bachman, Davidson, Ryan, & Choi, 1995; Brown, A., & Iwashita, 1998; Clapham, 1996) ･Multifaceted Rasch analysis (Fulcher, 2003, p. 196) (see also Alderson et al., 1995, pp. 185, 194 for other methods)
The external aspect [Relationships of test scores with other tests and behaviors]	■Relationships of the test scores with other measures and background variables (i.e., the test's external structure) □Is there convergent and discriminant evidence from multitrait-multimethod (MTMM) comparisons, as well as evidence of criterion relevance and applied utility? □Do the constructs represented in the assessment rationally account for the external pattern of correlations? □Are empirical relationships with criterion measures, (or quantifiable performances; Chapelle, 1999, p. 262), or the lack thereof, consistent with the meaning of the scores?	･Correlations (e.g., Davies, 1984), MTMM (Chapelle, 1999, p. 262), factor analysis, & structural equation modeling (SEM; Bachman & Eignor, 1997, pp. 233-237; Henning, 1987; e.g., Bachman & Palmer, 1982) (see also Alderson et al., 1995, pp. 177-183, 185-186, 193-194 for other methods)

(Table 2.25 continues)

(Table 2.25 continued)

Aspect	■Element to be examined □Examples of questions to be asked	･Methods to be used
The consequential aspect [Testing consequences]	■Social consequences of interpreting and using the test scores in particular ways, scrutinizing not only the intended outcomes but also unintended side effects □Are there value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness and distributive justice, as well as to washback □Are there rationales and evidence for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term (especially the intended and unintended consequences associated with bias in scoring and interpretation), with fairness in test use, and with positive or negative washback effects on teaching and learning?	･Qualitative or ethnographic methodology, including observations, interviews, or questionnaires (Watanabe, 2004, pp. 22-23) ･Seeking the following three sources to make hypotheses for the research: (a) public opinions as reflected in various media, such as newspapers, magazines, TV programs, (b) interviews, and (c) the description of the target exams (Watanabe, 2004, p. 25) ･ Comparisons between pretest and posttest scores (Hayes & Read, 2004) ･ Analysis of teaching materials and documents (Shohamy, 2001, p. 52, 62, 69, 79) ･Using the descriptors by Foucault (1979, as cited in Shohamy, 2001, p. 54)

Note. The terms in the first column was derived from Messick (1996) and Chapelle (1999; expressed in [ ]). Chapelle was added in the table because she "offers slightly more transparent terminology" than Messick; Luoma, 2001, p. 86). The explanation in the second and third column was from Messick (1989, 1996) unless it was cited. See also Lumley & Brown (2005) for other methods to be used.

One point to be noted is that "reliability," or the consistency of tests, test items, tasks, occasions, and raters, is included in Messick's (1996) framework and is treated as one source of evidence for validity. This view of reliability has been accepted by many language testers (Chapelle, 1999; Kunnan, 2005, p. 781). To be specific, reliability belongs to two aspects in the framework: (a) internal consistency in the structural aspect, and (b) intra-rater reliability, inter-rater reliability, and test-retest reliability in the generalizability aspect.

Chapelle (1999) argued that in order to validate interpretation and uses based on test scores, three steps are needed: (a) constructing hypotheses about test results using

Table 2.26
Validity Aspects Examined in the Present Study

Aspect	Study 1A	Study 1B	Study 2B	Study 2D
The content aspect [Content analysis]	Expert professional judgment	Expert professional judgment	Expert professional judgment	Expert professional judgment
The substantive aspect [Empirical item or task analysis]			·Questionnaire · Discourse analysis by focusing on the language produced by a test taker	Questionnaire
The structural aspect [Dimensionality analysis]	Classical true-score reliability methods and item response theory (IRT) methods		· Classical true-score reliability methods and item response theory (IRT) methods ·Factor analysis, & structural equation modeling (SEM)	Principal components analysis (PCA)
The generalizability aspect [Differences in test performance]	Reliability in showing consistency of performance across raters	Reliability and correlations in showing consistency of performance across tasks and raters	Correlations in showing consistency of performance across modes	· Reliability in showing consistency of performance across raters ·PCA
The external aspect [Relationships of test scores with other tests and behaviors]	Correlations		Correlations ANOVA	Correlations
The consequential aspect [Testing consequences]			Questionnaire	Questionnaire

several different approaches for validation, (b) reporting the results in relation to the hypotheses, as evidence for validity, and (c) making a validity argument by integrating "evidence and rationales" (pp. 258-263). It is worth noting that evidence includes positive or negative types, and one of these types of evidence is derived when hypotheses are confirmed or not supported. This study intends to follow these three

steps of validation when possible. In addition, every attempt was made to cover as many aspects of validity as possible and to employ as many methods for one aspect as possible, based on Messick (1996).

There is one controversial issue to be discussed here: how to deal with face validity in the framework. Face validity refers to "the degree to which a test appears to measure the knowledge or abilities it claims to measure, as judged by an untrained observer" (Davies et al., 1999, p. 59). It is one of the important aspects in language testing because "tests that do not appear to be valid to users may not be taken seriously for their given purpose" and because "if test takers consider a test to be face valid, … they are more likely to perform to the best of their ability on that test and to respond appropriately to items" (Alderson et al., 1995, p. 173). However, whether face validity should be included in validity and, if it is included, in what aspect it should be included is a matter of debate (Nevo, 1985, p. 287). The current study puts it in the substantive aspect because face validity is closely related to response validity (Henning, 1987, p. 92)or "the extent to which examinees responded in the manner expected by the test developers" (Henning, p. 96) and response validity belongs to the substantive aspect. This is based on Alderson et al.'s (p. 173) statement cited above and Fulcher's (1999) argument that "if students do not take tests seriously their responses to test tasks are not likely to be adequate samples of their ability, which in turn threatens score meaning, and hence validity" (p. 233).

Nevo (1985) presented four dimensions of face validity (pp. 288-290): (a) the type of rater [a testee, a nonprofessional user (e.g., employers), or an interested individual (e.g., parents of testees)], (b) what is rated [a test item, a test, or a battery of tests], (c) what technique is used [an absolute (e.g., on the 5-point scale) or a relative (by comparing tests) technique], and (d) on what aspects raters give responses [from

"very suitable" to "unsuitable" or from "very relevant" to "irrelevant"]. In the present study, face validity was assessed as follows: (a) A testee rated (b) a test by employing (c) an absolute technique (i.e., on the 5-point scale) (d) on a scale from "relevant" to "irrelevant" for its intended use. It was done "immediately after the test has been administered while the testees' impressions are still fresh" (p. 290). The responses were not anonymous.

2.6 Present Study

2.6.1 Purpose, a Hypothesis, and a Research Question

The main purpose of this study is to examine relationships between size and depth of productive vocabulary knowledge and speaking performance among Japanese learners of English at the novice level. In order to achieve this purpose, one Hypothesis and one Research Question are posed.

Hypothesis: Size of productive vocabulary knowledge is related more to speaking performance than depth of productive vocabulary knowledge is.

Research Question: What aspects of speaking performance are related to productive vocabulary knowledge substantially (i.e., to a moderate or strong degree)?

The Hypothesis was set based on Meara (1996; see section 2.2.1). Since there were no studies into relationships between productive vocabulary knowledge and several aspects of speaking performance, one Research Question is posed.

Aspects of speaking performance investigated in this study are fluency, accuracy, syntactic complexity, and lexical complexity. In order to examine the focus of this study, two studies are conducted. Study 1 (1C) was conducted as a pilot study of Study

2 (2E) and Study 1C examined relationships between productive vocabulary knowledge and speaking performance with a specific focus on "size" of productive vocabulary knowledge. Study 2E was modified based on problems encountered in Study 1C. Study 2E investigated relationships between "size and depth" of productive vocabulary knowledge and speaking performance. The Hypothesis was tested in Study 2 (Hypothesis 4; see section 12.1), whereas the Research Question was investigated first in Study 1C (see section 6.1), whose results were put into three hypotheses (Hypotheses 1 to 3; see section 12.1) in Study 2E.

2.6.2 Significance and Characteristics

The present study is essential from theoretical and pedagogical perspectives. For a theoretical perspective, this study contributes to the following two aspects. First, examination of relationships between vocabulary knowledge and speaking performance has been a missing but necessary element of vocabulary studies. In particular, this study includes not only size but also depth of vocabulary knowledge, which has not been included in the previous studies despite its importance.

Second, this study provides one step toward building an empirical model of speaking performance (e.g., Bachman and Palmer, 1996) by examining effects of factors on speaking performance (e.g., to what degree speaking performance is affected by vocabulary knowledge). Model building is important because it "leads to a clearer understanding of both concepts and will allow testers to define more completely the constructs upon which tests of performance are based" (O'Sullivan, 2002, p. 292).

For pedagogical perspectives, this study provides evidence and rationales for teachers having long-term perspectives of teaching and testing vocabulary and

speaking.

In addition, this study has seven characteristics that differentiate it from previous studies. First, vocabulary and speaking are assessed using different tests. Recently, there have been studies that have used a learner corpus (e.g., Izumi et al., 2004; Kimura, 2005), and examination of vocabulary use in speaking performance can be done in this way (see section 2.4.1). However, in the existing corpora, there does not seem to be relevant data containing information on productive vocabulary knowledge and speaking performance, which is the focus of this study. Thus, data was collected by the author to examine this aspect.

Second, participants were novice learners, including junior and senior high school students in Japan, and the number of participants was large (e.g., n = 225 in Study 2E; see Table 2.24 for the number of participants in the previous studies).

Third, this study deals with speaking performance in relation to productive vocabulary knowledge. Speaking performance was conceptualized as not being holistic, but having multiple aspects (i.e., fluency, accuracy, syntactic complexity, and lexical complexity). These aspects were assessed using speaking performance measures, not speaking test scores using rating scales (see section 2.3.2 for the details).

Fourth, this study uses effect sizes for the interpretation rather than significant tests. Use of effect sizes has been advocated by the American Psychological Association (2001) and Wilkinson and Task Force on Statistical Inference (1999), and they have been used as an essential tool to interpret results (see section 2.6.4.1).

Fifth, this study uses multiple tests (to be specific, test sections), measures, or tasks in assessing productive vocabulary knowledge and speaking performance. Each instrument may have some degree of construct irrelevant variances, which the instrument is not intended to assess but which cannot be excluded completely in any

instrument. Using more tests, measures, and tasks provides a possibility that more unbiased assessment can be possible.

Sixth, in order to assess the knowledge intended to be measured (i.e., the construct) this study develops tests. Furthermore, tests and measures used in this study go through validation that examines the validity of inferences based on Messick (1996). This validation is important in order to present evidence that tests or measures tend to measure what is intended to be measured to a large degree, and that they can be interpreted as intended. One thing to be noted is that it is not always easy to obtain only positive evidence and that it is a difficult task to make a reasonable validity argument based on conflicting results (Davies & Elder, 2005). This study takes a stance of presenting both positive and negative evidence so that readers can get more information on qualities of tests and measures used.

Seventh, this study combines methodologies used in language testing and second language acquisition, and thus is a study that belongs to an interface between language testing and second language acquisition, which has been advocated (Bachman & Cohen, 1998) but not been spread widely.

2.6.3 Definitions of Key Terms

The following definitions are used in this study.

(a) productive vocabulary knowledge: knowledge to produce vocabulary a form without the context of language use (Adapted from Laufer et al., 2004, and Read, 2000; see section 2.2.1)

(b) vocabulary size (breadth): the number of words whose written forms a person can write in response to a primary meaning (Adapted from Nakanishi & Shimamoto,

2003; see section 2.2.1)

(c) vocabulary depth: the degree to which a person can write written forms of derivatives, antonyms, and collocates of stimulus words (Adapted from Nakanishi & Shimamoto, 2003; see section 2.2.1).

(d) derivations: "the formation of new words by adding affixes to other words or morphemes" (Richards & Schmidt, 2002, p. 151) and by removing "an affix from an existing word" (p. 45; see section 7.2.2.1.2.1).

(e) antonyms: "a word which is opposite in meaning to another word" (Richards & Schmidt, 2002, p. 27; see section 7.2.2.1.2.2)

(f) collocations: "the tendency of two or more words to co-occur in discourse" (Schmitt, 2000, p. 76). This study considers two adjacent words as collocations if they occurred together (see section 7.2.2.1.2.3).

In Read's (2004) framework of depth consisting of three aspects (i.e., precision of meaning, comprehensive word knowledge, and network knowledge; see section 2.2.1), derivations, antonyms, and collocations belonged to both the second and third aspects.

(g) speaking performance: actual instances of producing oral language in real time (Adapted from McNamara, 1996, p. 54, and Tarone, 2005, p. 485; see section 2.3.1). Speaking performance is regarded as including aspects of fluency, accuracy, syntactic complexity, and lexical complexity. It should be noted that this study considers knowledge as different from performance. Knowledge is inferred from the activity of producing a word without the context of language use, whereas performance is observed from the activity of producing words in the context (i.e., discourse; Read, 2000, p. 11) of language use. Furthermore,

speaking performance is different from speaking ability, which underlies speaking performance (see section 1.1). Ability and knowledge are similar and both involve using (i.e., recognizing and producing) a language, but the term "knowledge" is used in this study only when there are no contexts (i.e., discourse; Read, 2000, p. 11; see section 2.2.1), whereas the term "ability" is used when there are contexts.

(h) fluency: how fast and how much a learner speaks without dysfluency markers (i.e., functionless repetitions, self-corrections, and false starts) in "coping with real time processing" (Wolfe-Quintero et al., 1998, p. 14; see section 2.3.2)

(i) accuracy: how much a learner speaks without errors in real-time communication (Adapted from Wolfe-Quintero et al., 1998; see section 2.3.2)

(j) complexity: the degree to which a learner uses varied and sophisticated structures and vocabulary in speaking (adapted from Wolfe-Quintero et al., 1998). Complexity is divided into syntactic complexity and lexical complexity (see section 2.3.2).

(k) description: saying things "not in a temporal structure, but in terms of … physical attributes" (Brown, T. & Yule, 1983, p. 46; e.g., "describing other people, other people's houses, meals, clothes, books, films, household gadgets" p. 46)

(l) comparison: saying things "in terms of … physical attributes" (Brown, T. & Yule, 1983, p. 46), while connecting things with each other (e.g., comparing two similar pictures)

(m) a speaking performance measure (i.e., a discourse analytic measure): a measure that "provides counts of specific linguistic features occurring in the discourse that results from performing the task" (Ellis, 2003, p. 298; see section 2.3.2).

(n) learners at the novice level: third-year junior high school students and first-year to third-year senior high school students in Japan. First-year and second-year junior high school students were considered to be learners at the starting level, which is lower than the novice level.

Some definitions are adapted from the original sources in a way that specifically fits the focus of this study. For example, the definition of vocabulary size is normally "the number of words with a primary meaning that a learner has" (Nakanishi & Shimamoto, 2003). However, this study uses the definition of "the number of words whose written forms a person can write in response to a primary meaning" because the study elicits written forms by writing words.

2.6.4 Analysis Procedures

This section provides a summary of procedures regarding the analysis of the data used in this study. After methods related to effect sizes are introduced, other methods are presented. Analyses were conducted using SPSS (2003), Amos (Arbuckle, 1999), and FACETS (Linacre, 2003).

2.6.4.1 Methods Related to Effect Sizes

As pointed out in section 2.6.2, the current study used effect sizes for the interpretation. Recently, the idea has been spread that the degree to which an independent variable (IV) affects a dependent variable (DV), which is called an effect size, should be examined rather than significant tests (Kline, 2004). The latter method only examines whether there is a statistically significant effect of an IV on a DV and is

problematic because test statistics vary according to not only "the effect size" but also "the sample size" (Kline, 2004, p. 41). In other words, when there are a large number of participants, results may always be significant despite a small or almost zero effect, whereas when there are a small number of participants, results may be non-significant despite a large effect (Kline, 2004, p. 41). Thus, results using significance tests were presented in tables or appendixes for reference, but they were not mentioned in this dissertation. As for the criteria to judge the importance of variables, this study used Cohen (1988) and treated medium (i.e., moderate) or large (i.e., strong) effects as substantial and important for the interpretation.[2.17] Although Cohen's criteria was not developed empirically (Cohen, p. 12), employing his criteria is advantageous in that results of several analyses (e.g., simple correlation, multiple regression, and analysis of variance [ANOVA]) can be interpreted consistently. Table 2.27 gives a summary of effect size indexes for each statistical method and the criteria of small, medium, and large effects. For example, the interpretation of η^2 was as follows: Lower than small $< .0099$; $.0099 \leq$ small $< .0588$; $.0588 \leq$ medium $< .1379$; $.1379 \leq$ large.

Table 2.27
Criteria of Interpreting Effect Sizes in Cohen (1988)

Effect size	Method	Small	Medium	Large	Pages in Cohen (1988)
$\|r\|$	Correlation coefficient	.10	.30	.50	pp. 79-80
q ($=\|z_1 - z_2\|$)	Difference between correlation coefficients	.10	.30	.50	p. 110, 115
η^2 ($= \sigma_m^2/\sigma_t^2$)	F test on means in the one-way ANOVA	.0099	.0588	.1379	p. 281, 283
Adjusted R^2	Multiple regression	.0196	.13	.26	pp. 413-414

Note. z = Fisher's z. In case of one-way ANOVA, $\eta^2 = \eta_G^2$ (generalized eta squared; Olejnik & Algina, 2003). $\eta^2 = \sigma_m^2/\sigma_t^2 = SS_{effect}/SS_{T.}$ = "the sum of squares for the effect" divided by "the total sum of squares" (Kline, 2004, p. 180).

In this section, three methods were explained: simple correlations, simple regression analysis, and multiple regression analysis. First, in order to examine a

relationship between two variables, a simple correlation (Pearson's product-moment correlation coefficients) was used because the number of participants was large enough (more than 14; Shiba & Watabe, 1984, pp. 109-112) although some variables were not normally distributed.

Second, simple regression analysis was utilized when there was a single IV that predicts a DV. When there were multiple DVs, simple regression analysis was repeated.

Third, multiple regression analysis (MRA) was used when there were multiple IVs that predicted a DV. When there were multiple DVs, MRA was repeated. Some may wonder if repeating simple or multiple regression analyses can be a problem due to lack of control of Type 1 errors when significance testing is used, but this study used effect sizes for the interpretation, and repeating the analyses was not a problem. It should be noted that R^2 ("the proportion of variation in the DV that is predictable from the best linear combination of the IVs," Tabachnick & Fidell, 2001, p. 124) has zero or positive values, whereas adjusted R^2, with which the overestimation of R^2 is adjusted (p. 147), may have negative values. Table 2.28 demonstrates the method of computing adjusted R^2 and an example of a negative adjusted R^2 value. Six assumptions to be checked for MRA are described in section 12.3.1.

Table 2.28
Formula of Adjusted R^2 and Example of a Negative Adjusted R^2 Value

(Adjusted R^2) = $1 - (1 - R^2)*(N - 1)/(N - k - 1)$

R^2 = squared multiple correlation
N = sample size
k = number of independent variables (IVs) (Tabachnick & Fidell, 2001, p. 147)

Example (see Appendix 12.1 for the data)

Variable	R^2	N	k	Adjusted R^2
PF1	.24	225	4	$1 - (1 - .24)*(225 - 1)/(225 - 4 - 1) = .23$
PF2	.01	225	4	$1 - (1 - .01)*(225 - 1)/(225 - 4 - 1) = -.01$

According to Tabachnick and Fidell (2001, p. 131), there are three main types of multiple regression: standard, sequential (i.e., hierarchical), and stepwise (i.e., statistical). Results of R^2 and adjusted R^2 vary depending on what type to use, so careful selection of a method is necessary. In standard multiple regression, all IVs are "entered into the regression at once" (p. 131), whereas in the sequential multiple regression, IVs are entered into the regression "in an order specified by the researcher" (p. 131). The order of sequential multiple regression is determined by "logical or theoretical considerations" (p. 132). In stepwise multiple regression, IVs are entered in an order "based solely on statistics computed from the particular sample drawn" (p. 133). Bieber (1988) argued that results derived from stepwise multiple regression "may be difficult, if not impossible, to reproduce with an independent data set" (p. 9). The current study mainly used standard regression because the order of entry of IVs was difficult to determine and because stepwise regression may give a misleading picture by depending only on statistical procedures. Sequential regression was also used in a complementary way.

In order to examine which IV is more important in MRA, there are four methods: (a) using *t* tests for regression coefficients, (b) standardized regression coefficients (β), (c) partial or semi-partial correlations, and (d) the relative Pratt index. Of the four methods, Thomas, Hughes, and Zumbo (1998) and Zumbo (2005a) recommended the relative Pratt index (*d*) because the other methods cannot deal with suppressor variables (i.e., IVs that are "not individually related to" the DV, "but do make a significant contribution to" the prediction of the DV; Zumbo, 2005a, p. 11) and because it has a mathematical foundation (Zumbo, personal communication, July 18, 2005). The current study mainly used the relative Pratt index, but standardized

regression coefficients were also presented in order to check the consistency of results regarding the relative importance of IVs. Another reason for reporting standardized regression coefficients was to examine the direction of relationships (i.e., positive or negative).

The relative Pratt index (*d*) is calculated using the formula: $d = (\beta * r)/R^2$ (Zumbo, 2005a, p. 9; β = standardized regression coefficient; r = simple Pearson correlation between the IV and DV). This index can be used in descriptive discriminant analysis and logistic regression as well as in MRA (Zumbo, 2005a). Table 2.29 illustrates an example and how to compute *d*.

Table 2.29
*Examples of Using the Relative Pratt Index (*d*)*

MRA was conducted using "size, derivations, antonyms, and collocations" as the IVs and "PF1 (speaking speed)" as the DV.

Result: R^2 = .24 (*Note*. This is not adjusted R^2.) (see Appendix 12.1)

The next question is to what degree each IV contributes to the total R^2.

	β	r	d
Size	.15	.45	(.15*.45)/.24 = 0.28
Derivations	.04	.39	(.04*.39)/.24 = 0.07
Antonyms	.29	.48	(.29*.48)/.24 = 0.58
Collocations	.05	.36	(.05*.36)/.24 = 0.07

All *d* values sum up to 1.00: .28 + .07 + .58 + .07 = 1.00 (see Table 12.13)

The criterion of detecting the unimportant IV(s) (Thomas, 1992, p. 347): less than 1/2*p* (Zumbo, 2005a, p. 11; *p* = the number of IVs)
In this case, *d* of less than 0.13 (i.e., 1/(2*4) = 1/8) was considered unimportant.
Therefore, derivations and collocations were considered unimportant, and size and antonyms were interpreted. Higher *d* indicates more importance on the DV. Thus, *d*s suggests the most important IV was antonyms.

Note. *d* changes depending on what IVs are entered. *d* can have negative values and go beyond +1.

2.6.4.2 Other Methods

This section describes three other methods used in this study: (a) factor analysis (FA) and principal components analysis (PCA), (b) structural equation modeling

(SEM), and (c) item response theory (IRT).

First, in order to clarify relationships between variables, FA and PCA were used. When there were highly correlated measures (.90 or above, Tabachnick & Fidell, 2001, p. 84, 589), PCA was selected because FA cannot include variables with high correlations (p. 589), but for the research purposes, inclusion of all variables was necessary. Six assumptions to be checked for PCA are described in section 11.3.1. Factors were extracted based upon the eigenvalues of 1.0 or above through principal components method, and then rotated using a method of oblique promax with Kaiser Normalization. Items with factor loadings of .30 or above were selected in order to be consistent with the criterion of medium effect sizes of *r* (.30 or above) and because this study wanted to use more variables for interpretation. Decision of the number of components and interpretation was done based on the scree plot and on the policy that more measures were adopted. Components were interpreted based on high loading variables.

Second, SEM was used in order to examine the relationships between observed and unobserved (i.e., latent) variables with errors removed from the model, and to see how well the model fits the data using fit statistics (e.g., Kano & Miura, 2002; Tabachnick & Fidell, 2001; Toyoda, 2003). SEM has an advantage of enabling researchers to construct a model in a flexible way that reflects existing knowledge (Yamamoto, 1999; see section 9.3.2 for assumptions to be checked for SEM).

Third, IRT is described in detail because the concepts behind this theory are rather complicated to understand. It is “a modern measurement theory, as opposed to classical test theory, based on the probability of a test taker with a certain underlying ability getting a particular item right or wrong” (Richards & Schmidt, 2002, p. 277). Henning (1987, pp. 108-109) and Ohtomo (1996, pp. 17-20) state that IRT has three

main advantages over classical test theory. First, item difficulty can be estimated independently of the sample and the ability of the sample group (sample-free item difficulty estimation). Second, person ability can be estimated independently of the difficulty of the items (test-free person ability measurement). Third, the reliability and standard error can be obtained "for estimates of both person ability and item difficulty" (Henning, p. 109; multiple reliability estimation). The scores derived from IRT analysis are expressed on a logit scale. IRT can be used to show one type of evidence for validity (Henning, p. 115).

There are four assumptions in using IRT (Henning, 1987, pp. 182-183): (a) unidimentionality of a test, (b) the use of a non-speeded test, (c) local independence of test items, and (d) sample invariance. First, for a test to be unidimensional, it needs to assess only one type of ability or one content. Second, the test should provide enough time to solve each problem. In other words, the test should not be influenced by the speed at which test takers solve problems. Third, local independence is met when the possibility of answering one item does not affect the possibility of answering another. Fourth and last, sample invariance means that the quality of the sample during a first administration does not vary much when the items are conducted during a second administration. In the present research, focus lies on unidimensionality because the other three can be considered to be fulfilled.

IRT can be classified into three types: a one-parameter logistic model, which estimates examinee ability and item difficulty, a two-parameter logistic model, which also contains a discrimination factor, and a three-parameter logistic model, which is the same as the second, except that it contains an additional guessing factor. The two- and three-parameter models contain more information than the one-parameter model, while the first two models require more test takers than the one-parameter model. The

one-parameter model is also referred to as the Rasch model, which is implemented by FACETS (Linacre, 2003) and used in the present study.

As seen above, one of the assumptions in using IRT is unidimensionality. There are two ways to check for unidimensionality. One is to do factor analysis before using IRT, and the other is to use the information on fit statistics (McNamara, 1996, p. 279). Since McNamara recommended the latter method (p. 279), this study used the latter in order to sustain unidimensionality.

"Fit" refers to "the degree of match … between the expectations of the model and the actual data for that candidate on each item" (McNamara, 1996, p. 137). If the data fits the model, unidimensionality is considered to be satisfied. There are two types of fit statistics: misfit and overfit. Misfitting items or persons are different from common response patterns (McNamara, pp. 170-172), whereas overfitting ones "conform to the predictions of … an IRT model" more than expectations (Davies et al., 1999, pp. 138-139) and are "redundant items" (McNamara, 1996, p. 176). This study considered only misfit as problematic to the model because fitting the model more than expected was not considered a problem. In this study, the criterion of misfit was set as an infit mean square statistic being "a mean plus two standard deviations ($M + 2SD$)" or above (when $n = 30$ or more; McNamara, 1996). If misfitting items were less than 10% of the total items (based on Stansfield & Kenyon, 1995, p. 132), it was judged that the data on the test fit the model and that a hypothesis for validity was supported (see sections 4.3.2 and 9.3.2). The acceptable criterion of having misfitting students (McNamara, 1996) was set as less than 2% (McNamara, 1996).

The analysis using IRT was basically performed as follows. First, the students' ability and item difficulty (and rater severity if necessary) were calibrated by including all items and persons (and raters) available. After the initial calibration and check of fit

statistics, the misfitting items were excluded in order to sustain unidimensionality. After the exclusion of misfitting items in the first run, the second analysis was conducted. The percentage of misfitting items was computed. If misfitting items were less than 10% of the total items in the second analysis, they were retained, and the second set of results was used in the subsequent analysis. In this procedure, a few misfitting items in the second run were basically retained because unidimensionality is a relative concept (McNamara, 1996, p. 278; North, 2000, p. 165, 233). As misfitting or overfitting items are excluded, other items may start to misfit or overfit (North, 2000, pp. 262-263), and it can turn into an endless task.

There are three main characteristics of FACETS (Linacre, 2003). First, it can handle polytomous data as well as dichotomous data. Another characteristic is that it can analyze data with many facets included. Facets are "aspects of the setting," such as candidates, items, raters, gender, time of day, physical settings, and interlocutors (McNamara, 1996, p. 128). The ability of students and the difficulty of items (and the characteristics [severity] of raters, if necessary) were included in the present study. Thanks to the inclusion of many facets, "FACETS allows … the appropriate adjustments or allowances to be made automatically" (McNamara, p. 128) and the results can be presented "in a single frame of reference" (p. 134). Third, FACETS can be implemented with missing data if some elements are shared in each facet (McNamara, p. 133), but a more accurate estimation is possible with less missing data (McNamara, p. 133).

This study did not use the results of test takers with zero or full scores after IRT analysis because their ability estimates were larger than the other test takers.

2.7 Summary of Chapter 2

Chapter 2 reviews previous studies on (a) factors that affect speaking performance, (b) productive vocabulary knowledge, (c) speaking performance, (d) relationships between vocabulary and speaking performance, and (e) validity and validation. The current study is broadly based on the performance model (Bachman & Palmer, 1996; Fulcher, 2003). In Bachman and Palmer's (1996) models, vocabulary knowledge, which is the focus of this study, belongs to "organizational knowledge" in "language ability." The current study focuses on declarative knowledge rather than procedural knowledge of vocabulary.

While there are many types of classification of vocabulary, this study mainly analyzes vocabulary knowledge from two points: (a) receptive vs. productive and (b) size vs. depth, and investigates size and depth of productive vocabulary knowledge. After nine existent tests are analyzed using Read's (2000) framework, this study decides to use Laufer and Goldstein (2004) to assess vocabulary size, and Shimamoto (2005) to measure vocabulary depth.

Speaking can be divided into several types. The main focus in this study is description and comparison in transactional function and monologues without pre-task planning time. In addition, speaking performance is seen from four linguistic aspects: fluency, accuracy, syntactic complexity, and lexical complexity. The four aspects are assessed using speaking performance measures.

There is a theoretical background to the relationship between vocabulary and speaking performance. Additionally, several empirical studies have been conducted. These studies that investigated relationships between vocabulary and speaking can be classified into two types: studies that dealt with (a) vocabulary used in speaking performance, and (b) relationships between vocabulary knowledge and speaking.

Previous studies suggest that the degree of relationships vary substantially according to aspects of speaking performance examined and measures used. Furthermore, each study seems to suffer from problems. In order to overcome the problems, the current study attempts to examine speaking performance from multiple perspectives and utilized multiple test sections or measures with detailed description and with more validity, with more participants that represent the population.

Validity is considered the most important aspect in test development and use. The present study uses Messick's (1996) validation framework.

In the last section, a purpose is described, and then significance and characteristics, definitions of key terms, and procedures regarding the analysis of the data used in this study are presented.

Part Ⅱ Study 1: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance

Chapter 3 Overview of Study 1 [3.1]

3.1 Purpose

The main purpose of Chapters 3 to 6 is to examine relationships between size of productive vocabulary knowledge and speaking performance, with a specific focus on Japanese beginning learners of English. In order to achieve this purpose, Chapter 3 (this chapter) describes the method used in the subsequent chapters. Chapters 4 and 5 examine the validity of inferences based on scores of the Productive Vocabulary Knowledge Test and values of the Speaking Performance Measures. Then relationships between size of productive vocabulary knowledge and speaking performance are investigated in Chapter 6. Figure 3.1 shows where the present chapter (i.e., Chapter 3 Overview of Study 1) is located in the overall research.

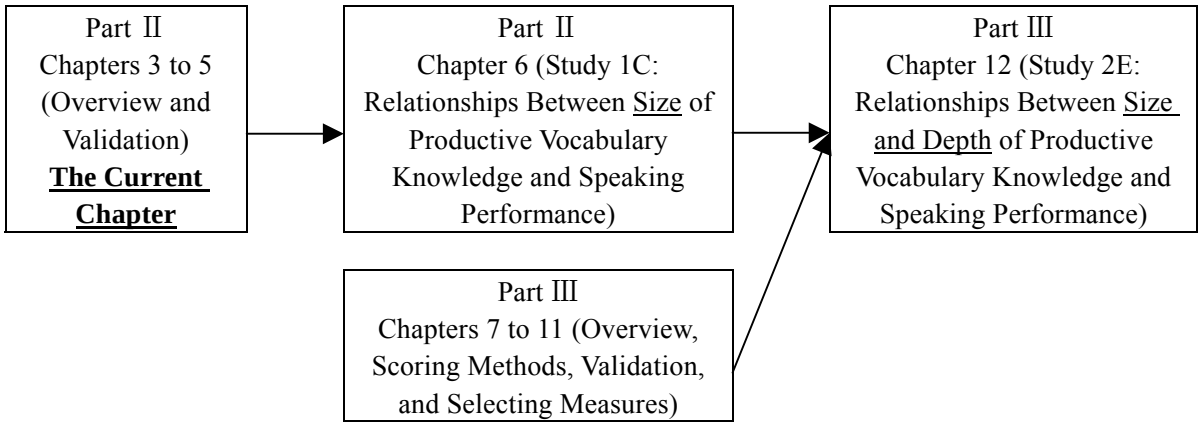


Figure 3.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

3.2 Method

3.2.1 Participants

Participants were described in Chapters 4 and 5.

3.2.2 Instruments

Four tests were used: (a) the Productive Vocabulary Knowledge Test (Productive VKT), (b) the Receptive Vocabulary Knowledge Test (Receptive VKT), (c) the Grammar Test, and (d) the Speaking Test. The first three tests were paper-and-pencil tests. The Productive VKT and the Speaking Test were the main focus in this study, whereas the other two tests were used for validation of the Productive VKT.

3.2.2.1 Productive Vocabulary Knowledge Test

The Productive VKT was designed to assess productive vocabulary knowledge for beginning learners of English. Productive vocabulary knowledge consists of (1) vocabulary size and (2) the knowledge of word characteristics in Chapelle's (1994) second dimension. To be more specific, productive vocabulary knowledge is defined as having knowledge of (d) a written form and (h) form and meaning in Nation's (2001) vocabulary knowledge framework (see section 2.2.1 Table 2.7), and as having knowledge of producing (recalling) a form related to a primary meaning in Laufer et al.'s (2004) classification. The detailed description of test purpose and design was based on Read and Chapelle's (2001) vocabulary assessment framework (see Appendix 3.1).

As the method to assess productive vocabulary knowledge, a translation format was selected, in which a Japanese (L1) meaning was presented to elicit an equivalent English (L2) word (see section 2.2.2). The first letter of the word was provided as a clue to avoid a word that had the same meaning but was different from the answer (see Table 3.1). This clue was provided to "restrict the number of possible responses to one" (Read, 2000, p. 174) and to make items easier to answer and score. This is the

same method as that used by Laufer and Goldstein (2004). It should be noted that providing the first letter can have both facilitating and debilitating effects on lexical retrieval. This is because the first letter can be a clue to solving a question, and it can also restrict possible answers and enhance the difficulty of an item.[3.2] Laufer and Nation (1995, 1999) and Meara and Fitzpatrick (2000) were not used in this study because there are problems in using their test formats with novice learners (see section 2.2.2). Test takers were not allowed to write katakana words in Japanese (these can express word pronunciation) as answers when they did not have enough confidence in their answers mainly because it is rather difficult to make scoring criteria of what is accepted as correct.

Table 3.1
Example of the Productive Vocabulary Knowledge Test

Write the English word that best corresponds to the Japanese meaning on your answer sheet. The first letter of the English word is already given. Write only part of the word even when you are very unsure about the exact answer. Example: 町 (t_____) The answer is *town*.

Note. All was originally written in Japanese. See Appendixes 3.2 to 3.4.

Although Study 1 (1C) examined relationships between size of productive vocabulary knowledge and speaking performance and may need to assess the oral aspect of productive vocabulary knowledge, a written (paper-and-pencil) format was used because the oral version lacked practicality in administration. However, in order to reduce possible effects of mode differences (i.e., oral vs. written) on results, scoring methods to take the oral aspect of productive vocabulary knowledge into account were devised and their validity is examined in Chapter 4.

The test developed, the Productive VKT, was adapted from a Vocabulary Size Test for Japanese Learners of English (Mochizuki Test, hereafter; Mochizuki, 1998),

which was designed to assess receptive vocabulary size in a multiple-choice format. The present study used mainly the 1,000 to 3,000 word frequency sections in the second version, with some modifications. One reason for using the Mochizuki Test was that it was developed for Japanese learners of English and overcame problems that previous tests had (e.g., the biased nature of a wordlist used for constructing a test; Mochizuki, pp. 36-41).

The validity of inferences based on the scores of the first version was examined (Mochizuki, 1998), but the second version was used in the current study because the first version had been modified into a proficiency test (Katagiri, 2001), and it appeared that students were more likely to have encountered the first version. In order to avoid data contamination, the second version seemed to be more appropriate for use in the current research.

Although the Mochizuki Test is self-contained, the following three modifications were needed for this study. First, the Mochizuki Test is based on the corpus using written language from the Hokkaido University English Vocabulary List (derived primarily from *Time* and Science Journals from the United States Environmental Protection Agency; Mochizuki, 1998, p. 38). There seems to be a possibility that a bias toward written language might affect the results (see McCarthy & Carter, 1997, for differences between spoken and written vocabulary). In order to take elements of spoken language into consideration, every word in the Mochizuki Test was checked using the *Longman Dictionary of Contemporary English* (3rd ed.; Summers et al., 1995), which contains spoken word marks (S1, S2, S3), indicating either the most frequent 1,000, 2,000, and 3,000 spoken words in English (see Appendixes 3.2 and 3.4). This dictionary was consulted because its frequency is based on British and American English, and on both casual and formal words (Kilgarriff, 1997, p. 138), and

the frequency seemed less biased than that of other dictionaries. After checking the spoken words against the word frequency marks, only words whose frequency was 3,000 or below, in terms of written and spoken language, were selected for the present study.

The second step used to modify the Mochizuki Test was to change the formats into the Productive VKT because the Mochizuki Test assesses only receptive vocabulary knowledge by having test takers select an appropriate English word corresponding to its Japanese meaning. The Productive VKT required learners to translate a Japanese word into an English one but allowed them to see the first letter of the target word as a clue, as described above (see Table 3.1).

Third, some words that belonged to the 1,000 word frequency section in the first version of the Mochizuki Test were also used because the number of vocabulary items successfully adapted from the second version turned out to be lower than 40, and because the items in the modified version seemed rather difficult. Additional items from the first version were checked utilizing the same method, as stated in the first step.

After making the necessary modifications, the Productive VKT was constructed. Words whose written and spoken frequencies were lower than 3,000 and which were not used in the Receptive VKT, were used in the Productive VKT. The Productive VKT included 40 items (see Appendix 3.2 for the whole test).

Preliminary content analysis of the Productive VKT showed that one item (Item 13) was not below the 3,000 spoken word frequency count. Therefore, it was excluded from analysis.

3.2.2.2 Receptive Vocabulary Knowledge Test

The Receptive VKT was designed to assess receptive vocabulary knowledge for beginning learners of English. Receptive vocabulary knowledge consists of (1) vocabulary size and (2) the knowledge of word characteristics in Chapelle's (1994) second dimension. To be more specific, receptive vocabulary knowledge is defined as having knowledge of (c) a written form and (g) form and meaning in Nation's (2001) vocabulary knowledge framework (see section 2.2.1 Table 2.7), and as having knowledge of recognizing a form related to a primary meaning in Laufer et al.'s (2004) classification. The detailed description of test purpose and design was based on Read and Chapelle's (2001) vocabulary assessment framework (see Appendix 3.1).

As the method to assess receptive vocabulary knowledge, a multiple-choice format was selected. The Receptive VKT, was adapted from the Mochizuki Test (Mochizuki, 1998) after a similar procedure to the modification described above (i.e., the first and third steps). In addition, although the Mochizuki Test had six options per item, the number of options was reduced to four because teachers commented that more than four options would be quite a burden on the students (see Table 3.2).

Table 3.2
Example of the Receptive Vocabulary Knowledge Test

Read the four possible answers and choose the English word that best corresponds to the Japanese meaning. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter of the answer that you have chosen.
Example: 町　(a) place　(b) scene　(c) square　(d) town*

Note. All was originally written in Japanese. See Appendixes 3.2 to 3.4.

After making the necessary modifications, the Receptive VKT was reconstructed, based on its original structure from the Mochizuki Test in the following procedure. Japanese meanings in the Receptive VKT were derived primarily from the Mochizuki

Test. When new words were used in answers, meanings were used from the first definition given in the *Longman Dictionary of Contemporary English* (3rd ed.; Summers et al., 1995). Then, three dictionaries (Konishi, 1996; Konishi, Yasui, & Kunihiro, 1987; Shimaoka, 2002) were consulted to find a translation that is easy to understand. In the end, the Receptive VKT included 40 items. The Receptive VKT was in a multiple-choice format with four options (see also Appendix 3.2 for the whole test).

3.2.2.3 Grammar Test

The Grammar Test was designed to assess grammatical knowledge, which refers to the knowledge of systematic rules (Newby, 2000, p. 248) that are not restricted to target words, phrases, sentences, or discourse. Fifteen items on the Grammar Test were derived from previous versions of the pre-second to fifth grade levels of the STEP (Society for Testing English Proficiency) Test (STEP, 2004; see Appendix 3.2). The items were selected so that a variety of grammatical descriptions in items from the Course of Study (Ministry of Education, Science & Culture, 1989) were covered (see Appendix 3.5). They were all multiple-choice questions, as seen in Table 3.3.

Table 3.3
Example of the Grammar Test

Read the four possible answers and choose the word that best completes the sentence. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter of the answer that you have chosen.
1. ( ) play the guitar late at night, Bob.
(a) Isn't (b) Doesn't (c) Don't* (d) Not

Note. All was originally written in Japanese. See Appendix 3.2.

Preliminary content analysis revealed that one item in the Grammar Test (Item 5; see Appendix 3.2) did not assess grammatical knowledge, but rather knowledge related

to the target word *spend*. Therefore, Item 5 was excluded, and 14 out of 15 items were analyzed further.[3.3]

3.2.2.4 Speaking Test

The speaking test was intended to assess speaking performance, specifically fluency, accuracy, syntactic complexity, and lexical complexity. Aspects and types of speaking performance in investigation were "description, comparison, and explanation with and without planning time."

The speaking test was a face-to-face oral interview composed of five tasks that elicited both dialogue (Tasks 2 and 3, which were both scripted role plays) and monologue (Tasks 1, 4, and 5). Tasks were developed based on North (2000) and the Course of Study (Ministry of Education, Science & Culture, 1989, 1999). Out of the five tasks, only three monologic tasks (Tasks 1, 4, and 5) designed to elicit long utterances were analyzed in this study because the other two tasks only elicited one-sentence-level utterances (e.g., *Good bye.*, *What time is it now?*; see Table 3.4, for all the elicited responses) and it was not common to use speaking performance measures for such short utterances. Task 1 was a self-introduction task, which had test takers speak about their school life, family, friends or other relevant topics. Task 4 involved describing their favorite things and people (i.e., singers, TV programs, or animals), reasons they liked them, and how popular these favorites were. Task 5 involved explaining differences between two pictures. Tasks 1 and 5 had no planning time, while Task 4 had one minute of planning time. No verbal feedback from an interviewer was given during the monologic tasks (see Appendixes 3.6 to 3.11).

The participants were not informed about the content or structure of the test beforehand. In order to exclude the effect of listening ability on utterances, Japanese

meanings were presented to a test taker when he or she had difficulty in comprehending the interviewer.

Table 3.4
Tasks in the Speaking Test

Task 1 Please introduce yourself for 90 seconds. Please talk about many things. When you have finished reading, please raise your head. Topic Examples: name, school year, school, favorites, family, and friends
Task 2 You are talking with your friend. Express what you're talking about as shown in the picture. Elicited sentences: *I want to go to Maruya to buy a book./Good bye./Hello./What time is it now?/I am sorry.*
Task 3 You are a reporter for your school newspaper. You are going to interview a boy who transferred from another school last week, and then write a report. Look at your notes and ask him questions about himself. The teacher in front of you will play the role of the new student. Things to ask the boy ①*Do you like this school?* ②*Where do you live now?* ③*Where did you live before?*
Task 4 Tell me about your favorite singer, TV programs, or animal for 90 seconds. You have one minute to prepare. Example: reasons, how popular they are
Task 5 Your brother is mischievous. While you were at school, he scattered your things about your room. When you scolded him about it, he said, "nothing has changed at all." Tell him how the room has changed by comparing how it was before with how it is now. You have 90 seconds to speak.

Note. All was originally written in Japanese. See Appendixes 3.4 to 3.6.

3.2.3 Procedures

3.2.3.1 Test Administration Procedures

The students took the two tests in July of 2002. The speaking test was conducted after school, whereas the written tests were administered in English classes except at the prefectural high school, where it was conducted after school.

The participants took the Productive VKT, the Receptive VKT, and the Grammar Test, which were combined and administered together. Basically, 45 minutes was given to solve questions, but the test ended early when every student said that he or she had finished.

Some of the students took the speaking test within one week of the three

paper-and-pencil tests. In administering the 15-minute Speaking Test, 11 Japanese interviewers with sufficient English speaking ability attended a practice session to learn the interview procedures and conducted a test. All students were informed that they would speak English in the test beforehand, but they were not informed about the test contents. Test takers were instructed in Japanese and told to speak as much as possible. In order to avoid pressuring the students too much, the interviewers moved to the next section when there was a silence for 15 seconds.[3.4] This 15-second period was determined as appropriate because STEP Test interviewers wait for 10 seconds at most and because this study wanted to add another five seconds to make sure that test takers had stopped talking. During the speaking test, all the utterances were tape-recorded.

3.2.3.2 Analysis Procedures

Analysis procedures were described in Chapters 4 to 6.

3.3 Summary of Chapter 3

This chapter describes the method used in Chapters 4 to 6. Four tests used were (a) the Productive Vocabulary Knowledge Test (Productive VKT), (b) the Receptive Vocabulary Knowledge Test (Receptive VKT), (c) the Grammar Test, and (d) the Speaking Test.

Chapter 4 Study 1A: Validation of the Productive Vocabulary Knowledge Test[4.1]

4.1 Purposes, Hypotheses, and a Research Question

The aims of the present chapter are twofold: (a) to examine the validity of interpretation based on the scores of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English, and (b) to investigate the qualities of three scoring methods of the Productive VKT (Methods 1, 2, and 3). The examination of the validity is needed in order to use tests with more validity for Study 1C (Chapter 6). The second aim is set in order to devise scoring methods to take the oral aspect of productive vocabulary knowledge into account for the purpose of reducing possible effects of mode differences (i.e., oral vs. written) on results. Figure 4.1 shows where the present chapter (i.e., Chapter 4 Study 1A: Validation of the Productive Vocabulary Knowledge Test) is located in the overall research.

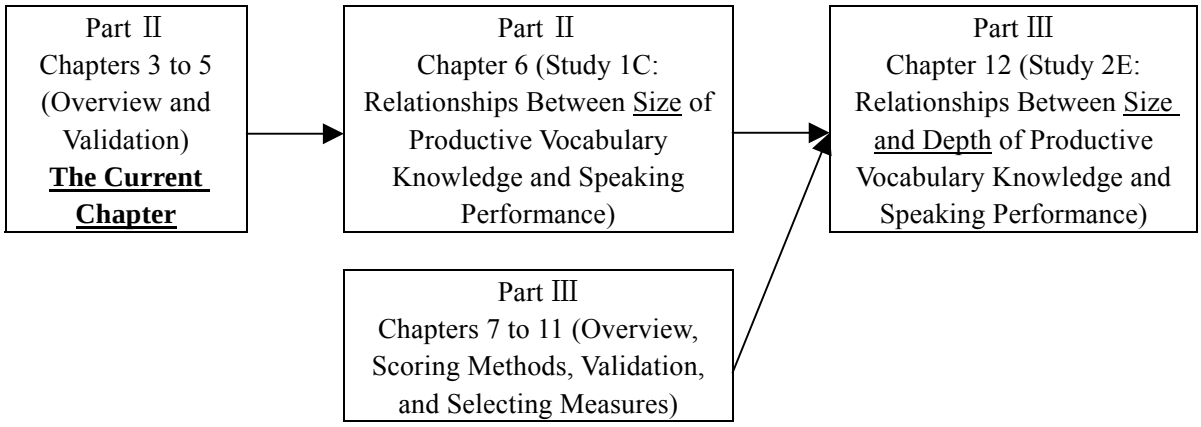


Figure 4.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

One Main Hypothesis and a Research Question were posed in relation to the validity and scoring methods. In order to examine the Main Hypothesis, five Hypotheses for Validity were made. Table 4.1 illustrates the organization of the current

study (Study 1A).

Main Hypothesis: The Productive VKT developed in this study has positive evidence for validity.

Hypothesis for Validity 1 (the structural aspect): The data on the Productive VKT fit an item response theory (IRT) model. In other words, the Productive VKT satisfies unidimensionality.

Hypothesis for Validity 2 (the structural aspect): The reliability of items and persons is high in the Productive VKT.

Hypothesis for Validity 3 (the generalizability aspect): The reliability of raters is high in the Productive VKT.

Hypothesis for Validity 4 (the external aspect): Correlations between the Productive VKT and the Receptive Vocabulary Knowledge Test (Receptive VKT) are higher than those between the Productive VKT and the Grammar Test, and those between the Receptive VKT and the Grammar Test.

Hypothesis for Validity 5 (the content aspect): The Productive VKT has positive evidence for the content aspect.

Research Question: Do Methods 2 and 3 have more positive evidence for validity than Method 1?

The Hypothesis for Validity 4 was based on the principle of internal construction validation (Henning, 1987) that correlations between similar aspects are stronger than those between different aspects.

Table 4.1
Organization of Study 1A: Validation of the Productive Vocabulary Knowledge Test

	Validity aspect	Method to be used	Section reported
Main Hypothesis			4.3.4
Hypothesis for Validity 5	Content	Expert professional judgment	4.3.4
Hypotheses for Validity 1 and 2	Structural	Classical true-score reliability methods and item response theory (IRT) methods	4.3.2
Hypothesis for Validity 3	Generalizability	Reliability in showing consistency of performance across raters	4.3.2
Hypothesis for Validity 4	External	Correlations	4.3.3
Research Question			4.3.5

4.2 Method

4.2.1 Participants

The participants in this study were 343 students studying English in Japan. This group contained 182 junior high school students and 161 senior high school students (see Table 4.2). They were aged from 14 to 16 and all had studied English as a foreign language for approximately two or three years. Their first language (L1) was Japanese. All students in School D belonged to an English course. All students in the same school year in Schools B, C, and E and those in the same class in School A participated. Approximately 60% of students in School E and in the English course in School D go to university directly after graduation. Therefore, a wide range of learners in terms of

Table 4.2
Number of Participants at Each School

Type of school (School year)	*n*
Public junior high school [School A] (2nd year)	33
Public junior high school [School B] (3rd year)	73
Public junior high school [School C] (3rd year)	76
Prefectural senior high school [School D] (1st year)	5
Senior high school affiliated with a national university [School E] (1st year)	156[a]
Total	343

Note. [a]Originally, 162 students took the three paper-and-pencil tests, but six students were not used for analysis because four of them went abroad for an extended period of time, and two were from foreign countries.

English ability participated. They were considered to be novice Japanese learners of English in this research.

4.2.2 Instruments

The three tests were used in this chapter: (a) the Productive Vocabulary Knowledge Test (Productive VKT), (b) the Receptive Vocabulary Knowledge Test (Receptive VKT), and (c) the Grammar Test. The details were described in sections 3.2.2.1 to 3.2.2.3.

4.2.3 Procedures

4.2.3.1 Test Administration Procedures

The details were described in section 3.2.3.1.

4.2.3.2 Scoring Procedures

In scoring the Productive VKT, there were three scoring methods: Methods 1, 2, and 3 (see Table 4.3). With Method 1, one point was given only when the spelling was correct (i.e., only when the meaning and the written form [spelling] matched exactly). With Methods 2 and 3, one point was provided even when the spelling was not correct but the spelling suggested that a test taker knew the correct spoken form (pronunciation). Methods 2 and 3 were different in that Method 2 had three conditions to be met to get a score, whereas Method 3 had only one. In other words, Method 3 was the most lenient, followed by Method 2, while Method 1 was the strictest. For example, those who wrote *pich* in the Productive VKT gained one point in Method 3, whereas they obtained zero points in Methods 1 and 2.

Table 4.3
Scoring Methods of the Productive Vocabulary Knowledge Test in Study 1A

Method	Criteria	Responses scored as correct
1	Correct only when the spelling is correct.	*peach; bag*
2	Correct when the spelling is correct OR when there is a misspelling but the following three conditions are met. (a) Only one letter is different from the correct spelling. (b) The pronunciation of a test taker's response read aloud by a rater is almost the same as that of the correct answer when the rater does not know the correct answer. (c) There are no other English words with similar spelling.	*peach; bag* *peech; bak*
3	Correct when the spelling is correct OR when there is a misspelling but the pronunciation of the test taker's response read aloud by a rater is almost the same as that of the correct answer when the rater compares it with the correct answer.	*peach; bag* *peech; bak* *pich, pici; bog*

Note. Both singular and plural forms were scored as correct.

Method 2 was based on Waring (1997) but is different in the following two respects. First, for the criterion of (b) in Table 4.3, Waring scored a response as correct if the response was similar overall to the target word, but the criterion of this study was refined and made more concrete. Second, Waring scored a response as correct when there was a tense mistake, while this study did not, and scored a response as correct only when a present tense form was written because all the prompts in the test used the present tense.

The three scoring methods were made and compared in this chapter because in order to examine relationships between productive vocabulary knowledge and speaking performance, knowledge of a spoken form (pronunciation) may need to be assessed in the Productive VKT rather than a written form (spelling). However, giving credit for a correct spoken and a correct written form (Methods 2 and 3) requires

judgment and takes more time to score than a correct written form (Method 1), which results in a lower degree of practicality. One question here is whether Methods 2 and 3 are more valid than Method 1, and whether Methods 2 and 3 are worth using even when their practicality is lower. This issue was raised in the Research Question. The answers in the Productive VKT were evaluated by two raters, both of whom had majored in language testing.

As for the Receptive VKT and Grammar Test, selecting one correct answer resulted in the award of a point.

4.2.3.3 Analysis Procedures

The Main Hypothesis and the Research Question were tested using results of five Hypotheses for Validity. The Hypotheses for Validity 1, 2, and 3 were analyzed using the Rasch multi-faceted measurement in item response theory (IRT). IRT was described in section 2.6.4.2. The Hypothesis for Validity 4 was examined using Pearson's product-moment correlation coefficients (r) for the purpose of investigating the strength of the relationship between two variables. The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1). When correlation coefficients were compared and one coefficient was higher than the other in a predicted direction, hypotheses were considered to be partially met. When the effect sizes (q) of differences between the correlation coefficients were medium (q = .30 to .49) or strong (q = .50 or more; Cohen, 1988; see section 2.6.4.1), the hypotheses were considered to be confirmed. The Hypothesis for Validity 5 was analyzed based on a professional judgment. The Main Hypothesis 5 was tested using results of seven Hypotheses for Validity.

4.3 Results and Discussion

4.3.1 Receptive Vocabulary Knowledge Test and Grammar Test

As in Table 4.4, the first Rasch analysis of the Receptive VKT revealed that two items (Items 6 & 23) were misfitting (5%), leading to their exclusion. In the second analysis, two items (Items 11 & 27) were misfitting, but the percentage of misfitting items (5%) was less than 10% (Stansfield & Kenyon, 1995; see section 2.6.4.2). Thus, the results of the remaining 38 items in the second analysis were used for the subsequent analysis.

After the Grammar Test was examined using the Rasch analysis, it was found that one item (Item 9) was misfitting (7%), which was then excluded in the second analysis. In the second analysis, one item (Item 12) was misfitting and the percentage of them (8%) was less than 10%. Thus, the results of the remaining 13 items in the second analysis were used for the subsequent analysis.

Table 4.4
Misfit Criteria, Percentages of Misfit, and Reliability of the Receptive Vocabulary Knowledge Test and the Grammar Test

		Infit Mean Square *M* (*SD*) [Misfit Criterion]	Percentage of misfit (The number beyond the criterion/Total number)	Reliability
Receptive first	Test takers	1.0 (0.3) [1.6]	3.79% (13/343)	.89
	Items	1.0 (0.2) [1.4]	5.00% (2/40)	.99
Receptive second	Test takers	1.0 (0.3) [1.6]	3.50% (12/343)	.89
	Items	1.0 (0.2) [1.4]	5.26% (2/38)	.99
Grammar first	Test takers	1.0 (0.3) [1.6]	1.46% (5/343)	.76
	Items	1.0 (0.2) [1.4]	7.14% (1/14)	.98
Grammar second	Test takers	1.0 (0.3) [1.6]	1.17% (4/343)	.75
	Items	1.0 (0.1) [1.2]	7.69% (1/13)	.98

Note. *n* = 343; Test takers = Test takers' ability; Item = Item difficulty. Reliability = "Rasch equivalent to the KR-20 or Cronbach Alpha statistic, i.e., the ration [*sic*] of 'True variance' to 'Observed variance'" (Linacre, 1991, pp. 85-86).

Since the criterion of less than 10% (Stansfield & Kenyon, 1995; see section 2.6.4.2) was met, the fit of the data to the model (i.e., unidimensionality) was

considered to be satisfied, which leads to one type of evidence for validity of the two tests. In addition, the reliability of items and persons was sufficiently high. Although the proportion of misfitting persons was slightly higher than 2% (McNamara, 1996; see section 2.6.4.2) in the Receptive VKT , the difference between 2% and 3.5% was marginal and all test takers were included in the analysis.

4.3.2 Investigating the Hypotheses for Validity 1 to 3

The analysis of the Productive VKT was done for each of the three scoring methods. As in Table 4.5, For Method 1, the first Rasch analysis of the Productive VKT revealed that one item (Item 5) was misfitting (3%), leading to its exclusion. In the second analysis, no items were misfitting and the percentage of misfitting items

Table 4.5
Misfit Criteria, Percentages of Misfit, and Reliability of the Productive Vocabulary Knowledge Test

			Infit Mean Square *M* (*SD*) [Misfit Criterion]	Percentage of misfit (The number beyond the criterion/Total number)	Reliability
Method first	1	Test takers	1.0 (0.3) [1.6]	2.62% (9/343)	.98
		Items	1.0 (0.2) [1.4]	2.56% (1/39)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {99.4%}
Method second	1	Test takers	1.0 (0.4) [1.8]	2.33% (8/343)	.98
		Items	1.0 (0.2) [1.4]	0.00% (0/38)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {99.4%}
Method first	2	Test takers	1.0 (0.4) [1.8]	2.04% (7/343)	.98
		Items	1.0 (0.2) [1.4]	2.56% (1/39)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {99.0%}
Method second	2	Test takers	1.0 (0.4) [1.8]	2.33% (8/343)	0.98
		Items	1.0 (0.2) [1.4]	0.00% (0/38)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {99.0%}
Method first	3	Test takers	1.0 (0.4) [1.8]	2.33% (8/343)	.97
		Items	1.0 (0.2) [1.4]	2.56% (1/39)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {98.5%}
Method second	3	Test takers	1.0 (0.4) [1.8]	2.92% (10/343)	0.97
		Items	1.0 (0.2) [1.4]	0.00% (0/38)	1.00
		Raters	1.0 (0.0) [NVC]	NVC	.00 {98.5%}

Note. *n* = 343; Test takers = Test takers' ability; Item = Item difficulty; { } = The percentage of the two raters' agreement. .00 in the rater reliability means 1.00 (Linacre, 1991, p. 86). NVC = No values computed. See Appendixes 4.1 to 4.5.

(0%) was less than 10%. Thus, the results of the remaining 38 items in the second analysis were used for the subsequent analysis.

After the initial calibration of Method 2, it was found that one item (Item 5) was misfitting (3%), which was then excluded in the second analysis. In the second analysis, no items were misfitting and the percentage of them (0%) was less than 10%. Thus, the results of the remaining 38 items in the second analysis were used for the subsequent analysis.

As for Method 3, the first Rasch analysis of the Productive VKT indicate that one item (Item 33) was misfitting (3%), leading to its exclusion. The second analysis revealed that no items were misfitting and the percentage of them was less than 10% (0%). Thus, the results of the remaining 38 items in the second analysis were used for the subsequent analysis.

It should be noted that Methods 1 and 2 produced the same misfitting item judgment (Item 5 as misfitting), whereas Method 3 produced a different misfitting item judgment (Item 33 as misfitting), which suggests the possibility that Method 3 assesses a rather different dimension.

Since the criterion of less than 10% was met, the fit of the data to the model (i.e., unidimensionality) was considered to be satisfied. The Hypothesis for Validity 1 was supported, which leads to the structural aspect of validity evidence of the three scoring methods on the Productive VKT. Although the proportion of misfitting persons was slightly higher than 2% (McNamara, 1996; see section 2.6.4.2), the differences between 2% and 2.33% in Methods 1 and 2 and the one between 2% and 2.92% in Method 3 were marginal, and all test takers were included in the analysis.

Next, three types of reliability in the second analysis were examined here: (a)

Rasch item reliability (equivalent to the KR-20 or Cronbach's alpha; Linacre, 1991, p. 86), (b) Rasch person reliability, and (c) interrater reliability. First, Rasch reliability of items was found to be very high for Method 1 (1.00), Method 2 (1.00), and Method 3 (1.00). Second, Rasch reliability of persons was found to be very high for Method 1 (.98), Method 2 (.98), and Method 3 (.97). Third, Rasch reliability of raters was found to be very high for Methods 1 to 3 (.00, which means 1.00; Linacre, 1991, p. 86). The percentage of the two raters' agreement was very high for Method 1 (99.4%), Method 2 (99.0%), and Method 3 (98.5%). These results show that the three methods provide stable and consistent scores. Therefore, the Hypotheses for Validity 2 and 3 were supported.

4.3.3 Investigating the Hypothesis for Validity 4

In this section, associations are examined between logit scores in the Productive VKT (Productive VK scores, hereafter) obtained from the three scoring methods, the Receptive VKT (Receptive VK scores), and the Grammar Test scores (Grammar scores) in the second analysis. Table 4.6 shows means and standard deviations for the test scores. It should be noted that the logit scores cannot be compared, for example, between Method 1 of the Productive VKT and the Receptive VKT because the logit scores of item difficulty for each test were different. When raw scores were examined, the proportions correct showed that the Receptive VKT was easier than the Productive VKT (67% vs. 43 to 57%; see Table 4.7).

Table 4.6
Descriptive Statistics of the Logit Scores of Test Takers' Ability Estimates on the Productive and Receptive Vocabulary Knowledge Test and the Grammar Test

	M	*SD*	Skewness	Kurtosis	Minimum	Maximum
Productive Method 1 (38 [= 40-1-1] items)	-0.77	2.54	0.07	-0.52	-5.83	6.93
Productive Method 2 (38 [= 40-1-1] items)	-0.79	2.51	0.20	-0.23	-6.07	6.96
Productive Method 3 (38 [= 40-1-1] items)	-0.73	2.36	0.29	-0.26	-5.75	6.57
Receptive (39 [= 40-2] items)	1.48	1.49	0.35	-0.46	-2.32	4.94
Grammar (13 [= 15-1-1] items)	0.42	1.49	0.28	-0.87	-2.87	2.94

Note. n = 272. Productive = Productive Vocabulary Knowledge Test (the results in the second analysis). The number of participants (n) decreased (from n = 343 to n = 272) because test takers with zero or full scores for at least one test were excluded. See Appendix ? *M* and *SD* for item difficulty estimates.

Table 4.7
Descriptive Statistics of the Raw Scores on the Productive and Receptive Vocabulary Knowledge Test and the Grammar Test

	M	%	*SD*	Skewness	Kurtosis	Minimum	Maximum
Productive Method 1 (38 [= 40-1-1] items)	16.19[a]	.43	20.27	0.21	-1.24	2.00	74.00
Productive Method 2 (38 [= 40-1-1] items)	18.03[a]	.47	20.32	0.06	-1.25	2.00	74.00
Productive Method 3 (38 [= 40-1-1] items)	21.51[a]	.57	17.83	-0.33	-0.87	4.00	74.00
Receptive (39 [= 40-2] items)	26.16	.67	6.75	-0.17	-0.78	7.00	38.00
Grammar (13 [= 15-1-1] items)	7.35	.57	3.17	0.03	-1.24	1.00	12.00

Note. n = 272. % = M/Full score = Item facility; [a]Average mean of the two raters' rating; Productive = Productive VKT (the results in the second analysis). The number of participants (n) decreased (from n = 343 to n = 272) because test takers with zero or full scores for at least one test were excluded. See Table 4.6 for test takers' ability estimates. % shows that the Receptive VKT is the easiest, followed by the Productive VKT and Grammar Test.

In the Hypothesis for Validity 4, it was hypothesized that the correlation between the Productive VK scores and the Receptive VK scores would be higher than that between the Productive VK scores and the Grammar scores, and also that between the Receptive VK scores and the Grammar scores, because the Productive and Receptive VK scores should reflect more similar aspects of vocabulary knowledge. Table 4.8 shows that when correlation coefficients were compared, all the hypotheses

were met except in cases of first-year senior high school students. The effect sizes of differences between the correlation coefficients were medium (q = .35 to .47; Cohen, 1988; see section 2.6.4.1) in cases of third-year junior high school students. In the other cases, the effect sizes of differences between the correlation coefficients were small or almost zero (q = -.07 to .27; Cohen, 1988; see section 2.6.4.1). Therefore, the Hypothesis for Validity 4 was partially supported. When correlation coefficients were compared among first-year senior high school students, two of the hypotheses were not satisfied in Method 3 (see two opposite signs of inequality [<] in Table 4.8), while one of the hypotheses was not satisfied in Method 1 (see one opposite sign of inequality [<] in Table 4.8) although the effect sizes of differences between the correlation coefficients in Methods 1 to 3 were small or almost zero (q = -.07 to .00).

Table 4.8
Correlations Between the Scores on the Productive and Receptive Vocabulary Knowledge Test and the Grammar Test

	Productive & Receptive	>[a] (q)	Productive & Grammar	>[ab] (q)	Receptive & Grammar
Total (n =272)					.80**
Method 1	.85**	> (.13)	.81**	> (.16)	
Method 2	.86**	> (.17)	.81**	> (.19)	
Method 3	.85**	> (.13)	.81**	> (.16)	
1st-year senior (n = 98)					.46**
Method 1	.45**	= (.00)	.45**	< (-.01)	
Method 2	.46**	= (.00)	.46**	= (.00)	
Method 3	.41**	< (-.07)	.47**	< (-.06)	
3rd-year junior (n = 143)					.43**
Method 1	.71**	> (.35)	.49**	> (.43)	
Method 2	.73**	> (.39)	.49**	> (.47)	
Method 3	.73**	> (.39)	.49**	> (.47)	
2nd-year junior (n =31)					.44*
Method 1	.60**	> (.20)	.46**	> (.22)	
Method 2	.62**	> (.16)	.51**	> (.25)	
Method 3	.61**	> (.27)	.41*	> (.24)	

Note. [a]Hypothesized direction; [b]Comparison between the second column and the sixth column. Senior = Senior high school students; junior = junior high school students.
*p < .05. **p < .01.

As for the relationship between Methods 1, 2, and 3, very high correlations were found (r = .87-.99; see Table 4.9). Overall, the three scoring methods tend to assess very similar aspects of productive vocabulary knowledge. In particular, the adjacent scoring methods produced higher correlations than other methods. In other words, correlations between Methods 1 and 2 were always higher than the ones between Methods 1 and 3, whereas correlations between Methods 2 and 3 were always higher than the ones between Methods 1 and 3. This suggests that similar scoring criteria produce more similar results.

Table 4.9
Correlations Between the Scores Derived From the Three Scoring Methods

	Methods 1 & 2	Methods 1 & 3	Methods 2 & 3
Total (n =272)	.99**	.96**	.98**
Senior 1st year (n = 98)	.95**	.91**	.95**
Junior 3rd year (n = 143)	.98**	.90**	.94**
Junior 2nd year (n = 31)	.94**	.87**	.95**

Note. **p < .01.

Another point is that, as can been seen in Table 4.8, overall, the Productive VK scores, the Receptive VK scores, and the Grammar scores were strongly associated with each other when all the data was included (n = 272). These results indicate that a few types of knowledge are closely related and predictable from one another when a rather wide range of novice learners is participating. When an ability range is limited (e.g., only first-year senior high school students), the correlations are lower, which results in the more necessity of assessing each component separately.

4.3.4 Investigating the Hypothesis for Validity 5 and the Main Hypothesis: Validity Argument of the Productive Vocabulary Knowledge Test

As seen in sections 4.3.2 to 4.3.3, it was found that the Hypotheses for Validity 1, 2, and 3 were supported, while the Hypothesis for Validity 4 was partially supported. As for the Hypotheses for Validity 5 regarding the content aspects, since the original Mochizuki Test was made following a clear procedure of selecting words from a frequency list (Mochizuki, 1998), and the modification was done with explicit reasons in mind, content relevance was considered to be satisfied to some degree, which supports the Hypotheses for Validity 5. Although the results were not perfect, almost all the hypotheses were confirmed, which produced enough positive evidence for validity. Therefore, the Main Hypothesis was supported and it is concluded that there is an acceptable level of validity for the Productive VKT for novice Japanese learners of English developed in the current research, although this conclusion does not end the process of validation because "validation is a continuing process" (Messick, 1989, p. 13).

4.3.5 Investigating the Research Question: Scoring Methods of the Productive Vocabulary Knowledge Test

Generally speaking, Methods 1, 2, and 3 tended to measure similar knowledge and produced similar results (see Table 4.9). This conclusion is similar to results of previous studies concerning cloze tests showing that an exact word scoring method (similar to Method 1) is correlated strongly to acceptable word scoring methods (similar to Methods 2 and 3; e.g., Kobayashi, 2002, p. 575).

However, strictly speaking, two differences were seen through the analyses: (a) regarding misfitting items and (b) regarding relationships in relation to the Receptive VKT and Grammar Test. First, the misfitting item detected in Methods 1 and 2 was different from the one in Method 3, which indicates that Methods 1 and 2 assess more

similar knowledge. Second, it was revealed that when the selection of target test takers was narrower (i.e., among first-year senior high school students), Method 3 had two diversions of the hypotheses, whereas Method 1 had one. In this sense, in relation to receptive vocabulary knowledge and grammatical knowledge, Method 2 seems to be better than Methods 1 and 3 at least in one school year. Therefore, as an answer to the Research Question asking if Methods 2 and 3 have more positive evidence for validity than Method 1, generally all the methods have similar positive evidence but in certain cases, Method 2 has more positive evidence for validity. In addition, Method 3, which was the most lenient criterion, has the least evidence for validity, which suggests that too much leniency can distort results. Furthermore, Method 1 is better than the other two methods in terms of practicality because Methods 2 and 3 require more time to score, and finer judgment in scoring. Therefore, test users may need to select the three scoring methods according to (a) test purposes, that is, whether or not the target of assessment includes knowledge of spoken forms, (b) the range of target test takers, and (c) resources that test users have (e.g., time and raters).

4.4 Summary of Chapter 4

The present chapter (Study 1A) examines the validity of interpretation based on the scores of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English, and investigated the qualities of three scoring methods of the Productive VKT (Methods 1, 2, and 3). It is concluded that there is an acceptable level of validity for the Productive VKT. It is also shown that generally speaking, Methods 1, 2, and 3 tend to measure similar knowledge and produce similar results. This study decided to use Method 2 in Chapter 6 (Study 1C).

Chapter 5 Study 1B: Validation of the Speaking Performance Measures[5.1]

5.1 Purpose and Hypotheses

The purpose of this chapter is to examine the qualities of the various Speaking Performance Measures, with a specific focus on Japanese beginning learners of English. Specifically this study investigates (a) whether the AS-unit length (the number of tokens per AS-unit) indicates fluency or syntactic complexity and (b) whether the Speaking Performance Measures of fluency, accuracy, syntactic complexity, and lexical complexity have positive evidence for validity. The examination of the meaning of the measure and the validity is necessary in order to use measures with more validity for Study 1C (Chapter 6). Figure 5.1 shows where the present chapter (i.e., Chapter 5 Study 1B: Validation of the Speaking Performance Measures) is located in the overall research.

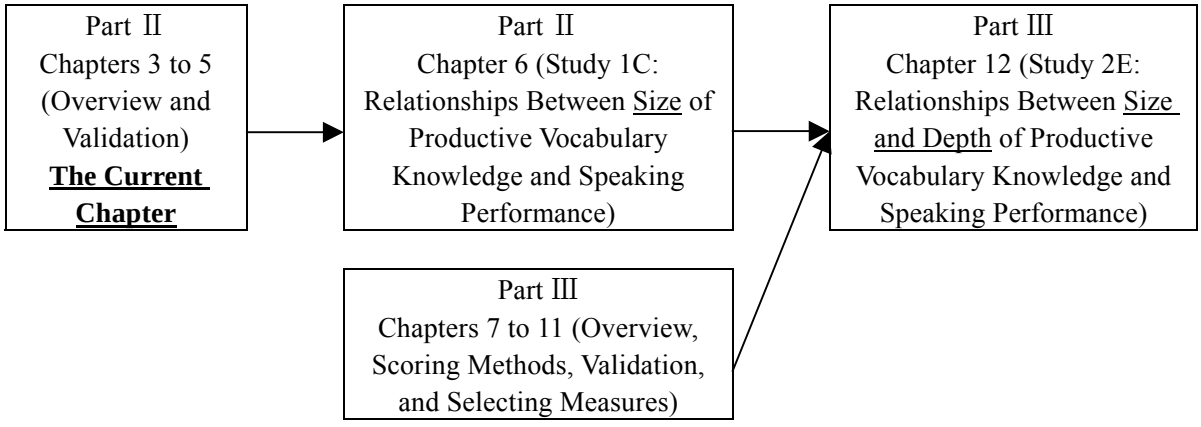


Figure 5.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

Two Main Hypotheses were posed in relation to the interpretation and validity of the Speaking Performance Measures. In order to examine the Main Hypothesis 2, three Hypotheses for Validity were made. Table 5.1 illustrates the organization of the current study (Study 1B).

Main Hypothesis 1: The AS-unit length (the number of tokens per AS-unit) is a syntactic complexity measure rather than a fluency measure. In other words, correlations between the AS-unit length and syntactic complexity measures are stronger than those between the AS-unit length and fluency measures.

Main Hypothesis 2: The Speaking Performance Measures have positive evidence for validity.

Hypothesis for Validity 1 (the generalizability aspect): The Speaking Performance Measures have strong correlations with the same measures of different tasks.

Hypothesis for Validity 2 (the generalizability aspect): The reliability of raters is high on the Speaking Performance Measures.

Hypothesis for Validity 3 (the content aspect): The Speaking Performance Measures have positive evidence for the content aspect.

The Main Hypothesis 1 was based on the previous studies (see section 2.3.2) and the principle of internal construction validation (Henning, 1987) that correlations between similar aspects are stronger than those between different aspects. The

Table 5.1
Organization of Study 1B: Validation of the Speaking Performance Measures

	Validity aspect	Method to be used	Section reported
Main Hypothesis 1			5.3.2
Main Hypothesis 2			5.3.4
Hypothesis for Validity 3	Content	Expert professional judgment	5.3.4
Hypotheses for Validity 1 and 2	Generalizability	Reliability and correlations in showing consistency of performance across tasks and raters	5.3.3 and 5.3.4

Hypothesis for Validity 3 investigated the relationships between “one measure of one task” and “the same measure of one of the other tasks” (e.g., a correlation between “F1 [the number of tokens per minute] of Task 1” and “F1 of Task 3”).

5.2 Method

5.2.1 Participants

The participants analyzed in this study were 74 Japanese lower and upper secondary school students (see Table 5.2). They included third-year students at two public junior high schools (n = 57) and first-year students at a prefectural senior high school (n = 4) and at a senior high school affiliated with a national university (n = 13). All of them were part of those who were analyzed in Chapter 4 (n = 343) except one test taker who did not take the Productive Vocabulary Knowledge Test (Productive VKT) but took the Speaking Test. There were originally 171 students who took the Speaking Test, but the following people were excluded: (a) students whose utterances were not usable due to mechanical problems (e.g., poor quality of recording) or problems concerning interviewing procedures (n = 24), and (b) those who did not generate a single clause in one or more of the three monologic tasks (Tasks 1, 4, or 5; see section 3.2.2.4; n = 73). The students in the second category (b) were excluded because it was not possible to calculate measures that used clauses as denominators (e.g., the number of error-free clauses per clause [A1]) when there were no clauses in the utterances. This leads to differences in the number of participants for each measure, which makes it difficult to compare the results between the measures. This seemed especially important in examining and comparing measures on the same basis. Although there is an alternative way to impute zero when measures were impossible to

calculate, this was not done because the appropriateness of this imputation method was unknown. For example, whether an utterance with no error-free clauses and no clauses (0/0 = an indeterminate value) can be treated in the same way as an utterance with no error-free clauses and three clauses (0/3 = 0) was unknown. The participants' speaking ability ranged approximately from the Smattering (or below) level to the Waystage Plus level[5.2] based on North's (2000) proficiency scale results (Koizumi, 2005a).[5.3]

Table 5.2
Number of Participants at Each School

Type of school (School year)	*n*
Public junior high school [School B] (3rd year)	29
Public junior high school [School C] (3rd year)	28
Prefectural senior high school [School D] (1st year)	4
Senior high school affiliated with a national university [School E] (1st year)	13
Total	74

5.2.2 Instruments

The Speaking Test was used. In order to compute the Speaking Performance Measures for analysis, three monologic tasks (Tasks 1, 4, and 5) designed to elicit long utterances were utilized (see section 3.2.2.4). The details were described in section 3.2.2.4.

5.2.3 Procedures

5.2.3.1 Test Administration Procedures

The details were described in section 3.2.3.1.

5.2.3.2 Coding Procedures

This section consisted of two phases: transcription of utterances and coding to compute the Speaking Performance Measures.

Utterances from the Speaking Test were transcribed. When there was a 15-second silence, the utterances produced after the silence, if any, were not transcribed. When there was a conversation between a student and an interviewer in the monologue tasks, the utterances were not used for analysis. The transcription was done with a policy of listening to the talk sympathetically by taking the context into consideration. In other words, pronunciation errors were not strictly reflected in transcripts. After the utterances were transcribed by the author, two graduate students who were majoring in applied linguistics checked them, taking half each, and then the author checked them a second time. All words were used for analysis rather than selecting a certain amount of text (e.g., the shortest text) because the volume of the data was rather small. This study analyzed the data for each task separately because the information from each task seemed more beneficial for future selection of measures.

In order to compute the Speaking Performance Measures for analysis, three monologic tasks (Tasks 1, 4, and 5) were utilized. As in Table 5.3, 17 measures (six for fluency, four for accuracy, three for syntactic complexity, and four for lexical complexity) were selected from 31 measures (see Table 2.20). The criteria for selecting measures were as follows. The fluency measures using pauses were not used because a specialized tool was necessary for strict measurement (Griffiths, 1991). Three types of dysfluency markers (i.e., functionless repetitions, self-corrections, and false starts) for F4, F5, and F6 were combined and used as in Takiguchi (2003) because they were considered to assess similar aspects of fluency since there were relatively strong associations (r = .60 to .87; Skehan & Foster, 1999) between different types of dyfluency markers (i.e., repetitions, false-starts, reformulations, and replacement). Specific measures of accuracy were not examined because the three tasks did not have target language structures as in Takiguchi (2003) and that Fujimori (2004) showed that

there were high correlations between specific measures and general measures of accuracy (r = .73 to .74). LC1 was selected rather than type token ratio (TTR) because LC1 was devised to reduce the problem of type token ratio being influenced by text length (e.g., Daller et al., 2003; Wolfe-Quintero et al., 1998; Read, 2000) by "rewarding the presence of many types in a longer composition, despite the increased repetition of those types" (Wolfe-Quintero et al., p. 107). Mean segmental type-token ratio and the index of lexical diversity (D) were not used because the spoken texts in this study were very short. The software requires at least 50 words to compute D (Malvern & Richards, 2002; Meara & Miralpeix, 2004).

Table 5.3
Summary of 17 Speaking Performance Measures Used in Study 1B

Construct	Code	Measure
Fluency	F1	No. of tokens per minute
	F2	No. of tokens per AS-unit
	F3	No. of clauses per minute
	F4+	No. of dysfluency markers per minute
	F5+	No. of dysfluency markers per token
	F6+	No. of dysfluency markers per AS-unit
Accuracy	A1	No. of error-free clauses per clause
	A2	No. of error-free AS-units per AS-unit
	A3+	No. of errors per token
	A4+	No. of errors per AS-unit
Syntactic complexity	SC1	No. of clauses per AS-unit
	SC2	No. of subordinate clauses per AS-unit
	SC3	No. of tokens per AS-unit
Lexical complexity	LC1	Guiraud index: No. of types divided by the square root of No. of tokens
	LC2	Lexical density: No. of lexical tokens per token
	LC3	Weighted lexical density: No. of sophisticated lexical tokens and No. of basic lexical tokens given half the weight divided by No. of tokens
	LC4	No. of sophisticated word types per token

Note. + = Measure that has a negative meaning when it has a higher value. F2 and SC3 are the same measures so there are actually 16 measures. Dysfluency markers consist of functionless repetitions, self-corrections, and false starts. See Appendix 5.1.

In order to compute 17 performance measures used, coding was carried out for the following aspects: AS-units, clauses, dysfluency markers, speaking time, tokens

and types, lexical tokens, sophisticated words, and errors. After the definition of each aspect was presented, a procedure to code it was described in this section. Before raters began independent coding, they practiced coding based on the definitions and examples. Examples of coding and computing measures can be seen in Appendix 5.1.

First, the definition of AS-units (Analysis of Speech Unit) was the same as Foster, Tonkyn, and Wigglesworth (2000), who defined it as "a single speaker's utterance consisting of *an independent clause, or sub-clausal unit,* together with any *subordinate clause*(s) associated with either" (p. 365; see also section 2.3.2). Fragmentary utterances were considered to be AS-units if utterances could be understood when a coder took information on task instruction (including pictures), content, and contexts into consideration. Japanese (L1) words were taken into account only when test takers used Japanese to substitute their unknown English words.

Second, clauses had two types: an independent clause and a subordinate clause. An independent clause was "minimally a clause including a finite verb" (Foster et al., 2000, p. 365). A subordinate clause included an adverbial clause, an adjective clause, and a nominal clause, and "consisted minimally of a finite or non-finite Verb element plus at least one other clause element (Subject, Object, Complement or Adverbial)" (Foster et al., p. 366). The following shows some examples.

|Hello.| |Open the door.| (no independent clauses)

|I like English.| |I like tennis and play it every day.| (1 independent clause)

|I study English :: because I like it.| |I like :: studying English.|
(1 independent clause, 1 dependent clause)

When there were no verbs, subordinate clauses were not counted (e.g., |{I} I like … Ayumi Hamasaki {because} … because {she} … she … pretty.| [1 clause]).

Third, dysfluency markers were marked as in Foster et al. (2000).

As for AS-units, clauses, and dysfluency markers, most (77%, n = 57) of the utterances were first coded by the author and another graduate student majoring in language testing. The interrater reliability of AS-units was very high (r = .99, p < .01). When there were disagreements, they discussed the differences until they reached an agreement, and the result was used as the final coding. There were no inconsistent codings regarding clauses and dysfluency markers between the author and the other rater. Since the definitions and coding systems were considered to be established, the rest of the utterances were judged by the author.

Fourth, speaking time was calculated twice by the author with a stopwatch. Time was calculated from when students started speaking to when they stopped speaking (i.e., to the last word before they produced a silence of 15 seconds) or until 90 seconds had passed. The intrarater reliability was very high for all tasks (r = .99, p < .01).

Fifth, tokens were pruned tokens after dysfluency markers were excluded. They did not include not only dysfluency markers but also filled and unfilled pauses. Types were counted based on lemma as in Daller et al. (2003). That is, a base form and inflected forms were considered to be the same type. For example, the following were considered to one type: *play, plays, playing*; *be, is, am, are, was, were, been*; *Taro, Taro's*.

Sixth, the definitions of lexical tokens and grammatical tokens were based on O'Loughlin (2001). Lexical words were content words and grammatical words were function words (e.g., all forms of *be, do, have*[5.4] and auxiliaries). However, some modifications were made from O'Loughlin as follows. Firstly, O'Loughlin used "an item" as a unit whereas this study used "one word" as a basic unit in order to be consistent with counting methods of types and tokens. Therefore, lexicalized clauses

(e.g., *I mean*), meta-talk (e.g., *the point is*) were counted individually. Secondly, while contractions were counted as two items in O'Loughlin (e.g., *they're* was counted as two items), this study counted the contractions as one word and judged whether they were lexical or grammatical based on the words before the apostrophe (e.g., *they're* as grammatical and *let's* as lexical). Third, in judging sophisticated words, O'Loughlin counted "repetition of low frequency lexical items … including alternate word forms of the same items (e.g., *study, student*)" as high frequency words a second time (p. 107). In contrast, this study took lemma count and examined the words independently and individually without considering repetitions.

The judgment of lexical and grammatical words from all the three tasks was conducted by the author and another rater who were majoring in applied linguistics. The interrater reliability was very high (r = .99, p < .01). When there were disagreements, they discussed and came to an agreement, which was utilized as the final coding.

Seventh, sophisticated words in this study were words not in the list of 1,000 words in the JACET8000 (JACET Basic Words Revision Committee, 2003), which uses lemma count. Numerals are included in the JACET8000. Proper nouns and Japanese words were excluded from the list of sophisticated words.[5.5] The criterion of beyond 1,000 words was used because lower secondary school students were supposed to learn 1,000 words (Ministry of Education, Science & Culture, 1989). The JACET 8000 Level Marker (Shimizu, 2005) was used for coding.

Eighth, errors included both major and minor errors of grammar and vocabulary and they were judged by a target-like criterion. Discourse errors in terms of coherence were not considered because the tasks were not intended to elicit such aspects. Errors of pronunciation were not assessed because of the difficulty in making an explicit

criterion (Ano, 2002, p. 41) and because error detection was done using transcripts, not recorded tapes. Major (i.e., global) and minor (i.e., local) errors were not separated for two reasons. First, there were no speaking studies that distinguished the two types of error. Second, according to Rifkin and Roberts (1995), previous studies regarding error gravity have not had consistent research designs and constructs. Errors from all the three tasks were coded by the author, and two graduate students who were majoring in applied linguistics examined the transcription, each checking half the total. The interrater reliability was very high (r = .94, p < .01). When there were disagreements, they discussed the discrepancies and reached an agreement after discussion, using the result as the final coding. After the coding, the values of the Speaking Performance Measures were computed.

5.2.3.3 Analysis Procedures

In order to investigate the Main Hypothesis 1 and the Hypotheses for Validity 1 and 2, simple correlations (Pearson's product-moment correlation coefficients) were used. The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1). The reason why simple correlations were utilized rather than not factor analysis or structural equation modeling (SEM) was that the number of participants was not large enough. When the effect sizes (q) of differences between the correlation coefficients were medium (q = .30 to .49) or strong (q = .50 or more; Cohen, 1988; see section 2.6.4.1), the hypotheses were considered to be confirmed. The Hypothesis for Validity 3 was analyzed based on a professional judgment. The Main Hypothesis 2 was tested using results of three Hypotheses for Validity.

5.3 Results and Discussion

5.3.1 Descriptive Statistics of the Speaking Performance Measures

Table 5.4 shows means and standard deviations for the basic data and the Speaking Performance Measures. The results demonstrate that the number of subordinate clauses and sophisticated word types was small as the participants were

Table 5.4
Descriptive Statistics of the Basic Data and the Speaking Performance Measures

	Task 1		Task 4		Task 5	
	M	*SD*	*M*	*SD*	*M*	*SD*
Speaking time (second)	41.92	25.15	34.97	23.96	58.24	30.26
No. of tokens (pruned tokens)	26.68	18.48	19.47	15.52	30.28	21.47
No. of AS-unit	4.89	2.86	3.38	2.09	5.20	2.91
No. of clauses	5.22	3.11	3.84	2.60	3.74	2.94
No. of dysfluency markers	3.47	4.81	3.68	6.17	6.46	8.61
No. of error-free clauses	3.32	2.07	1.96	1.76	0.73	1.56
No. of error-free AS-units	2.76	1.94	1.50	1.44	0.78	1.53
No. of errors	3.18	2.71	2.58	2.29	9.50	6.69
No. of subordinate clauses	0.11	0.39	0.58	0.78	0.11	0.51
No. of types	19.27	11.74	13.89	8.90	17.32	9.73
No. of lexical tokens	15.00	9.93	10.58	7.61	12.05	7.20
No. of lexical tokens within basic tokens	12.68	8.41	9.51	6.80	8.88	5.76
No. of sophisticated word types	1.96	1.78	0.95	0.99	2.46	1.58
F1 (No. of tokens per minute)	46.75	30.55	40.03	26.73	34.42	20.00
F3 (No. of clauses per minute)	9.83	6.72	8.70	7.48	5.05	4.95
F4 (No. of dysfluency markers per minute)	4.15	4.47	4.93	5.70	5.71	6.86
F5 (No. of dysfluency markers per token)	0.12	0.13	0.17	0.22	0.29	0.73
F6 (No. of dysfluency markers per AS-unit)	0.64	0.71	0.89	1.04	1.30	1.93
A1 (No. of error-free clauses per clause)	0.68	0.22	0.47	0.30	0.11	0.21
A2 (No. of error-free AS-units per AS-unit)	0.61	0.25	0.41	0.32	0.11	0.21
A3 (No. of errors per token)	0.11	0.09	0.15	0.11	0.41	0.27
A4 (No. of errors per AS-unit)	0.59	0.45	0.82	0.55	0.81	0.70
SC1 (No. of clauses per AS-unit)	1.08	0.23	1.17	0.35	0.74	0.31
SC2 (No. of subordinate clauses per AS-unit)	0.02	0.06	0.21	0.30	0.04	0.24
SC3 or F2 (No. of tokens per AS-unit)	5.22	1.16	5.67	1.99	5.48	1.67
LC1 (Guiraud index)	3.65	1.05	3.09	0.84	3.14	0.91
LC2 (Lexical density)	0.58	0.08	0.58	0.11	0.44	0.16
LC3 (Weighted lexical density)	0.33	0.06	0.32	0.07	0.28	0.11
LC4 (No. of sophisticated word types per token)	0.08	0.07	0.05	0.07	0.11	0.09

Note. $n = 74$.

beginning level learners. It seems that tasks in the current study may not have been suitable to elicit such features and that the criterion of considering any words beyond the 1,000 high-frequency words sophisticated words was too challenging for the target participants (see Appendixes 5.2 and 5.3 for all the correlational results).

5.3.2 Investigating the Main Hypothesis 1: Meaning of the AS-Unit Length

In the Main Hypothesis 1, it was hypothesized that if the AS-unit length (i.e., the number of tokens per AS-unit) were a syntactic complexity measure rather than a fluency measure, the correlations between the AS-unit length and syntactic complexity measures would be stronger than those between the AS-unit length and fluency measures.

Table 5.5 illustrates correlations between the AS-unit length and syntactic complexity measures and those between the AS-unit length and fluency measures. The former correlations were compared with the latter ones to examine the relative strength. For example, the correlation between the AS-unit length in Task 1 (Task 1 AS-unit length) and SC1 (the number of clauses per AS-unit; see Table 5.5) in Task 1 (Task 1 SC1; r = .34) was stronger than that between Task 1 AS-unit length and Task 1 F1 (the number of tokens per minute; r = .14) but weaker than that between Task 1 AS-unit length and Task 1 F4 (the number of dysfluency markers per minute; r = .42). As this example illustrates, the results were not clear-cut. Some correlations between the AS-unit length and syntactic complexity were stronger than correlations between the AS-unit length and fluency, and vice versa.

Table 5.5
Correlations Between the AS-Unit Length and Fluency Measures and Between the AS-Unit Length and Syntactic Complexity Measures

	T1F1	T1F3	T1F4	T1F5	T1F6	(F Averaged)	T1SC1	T1SC2	(SC Averaged)
T1 ASUL	.14	-.07	.42**	.17	.41**	(.22)	.34**	.44**	(.39**)
T4 ASUL	.14	.11	.32**	.16	.25*	(.20)	.34**	.33**	(.34**)
T5 ASUL	.18	.12	.38**	.20	.30**	(.24*)	.23*	.18	(.21)
	T4F1	T4F3	T4F4	T4F5	T4F6		T4SC1	T4SC2	
T1 ASUL	.15	.00	.34**	.10	.40**	(.20)	.24*	.16	(.20)
T4 ASUL	.18	-.06	.05	-.16	.19	(.04)	.77**	.74**	(.76**)
T5 ASUL	.25*	.09	.14	-.05	.19	(.13)	.42**	.32**	(.37)
	T5F1	T5F3	T5F4	T5F5	T5F6		T5SC1	T5SC2	
T1 ASUL	.25*	.04	.27*	-.07	.08	(.12)	.12	.00	(.06)
T4 ASUL	.34**	.07	.14	-.12	.01	(.09)	.27*	-.12	(.08)
T5 ASUL	.32**	-.10	.22	-.25*	-.04	(.03)	.28*	.12	(.20)
Among the same tasks						(.10)			(.47**)
Among all the three tasks						(.14)			(.31**)

Note. n = 74. T = Task. F = Fluency; SC = Syntactic complexity. ASUL = AS-unit length. Correlations were averaged after using Fisher z transformation (Henning, 1987). Underlined = correlations among the same tasks.
*p < .05. **p < .01.

However, when the correlations were compared among the same tasks (see underlined values in Table 5.5), most combinations were consistent with the hypothesis that correlations between the AS-unit length and syntactic complexity will be stronger than correlations between the AS-unit length and fluency. In Task 1, 80% (8 cases out of 10)[5.6] of the combinations satisfied the conditions, whereas in Task 4, 100% (10/10) and in Task 5, 70% (7/10) did. When all three tasks were combined, 83% (25/30)[5.7] satisfied the conditions. Also, an alternative way to examine is to use the average of correlations. As can be seen in Table 5.5, the averaged correlations between the AS-unit length and syntactic complexity measures were stronger than those between the AS-unit length and fluency measures in each task (i.e., Task 1: r = .39 > .22; Task 4: r = .76 > .04; Task 5: r = .20 > .03; 3 tasks combined: r = .47 > .10). Even when the correlations were compared among all the three tasks, in which task differences can induce some variations unrelated to the interpretation of the

AS-unit length, a majority of combinations (i.e., 70%, 63 cases out of 90)[5.8] satisfied the hypothesis, and the averaged correlation between the AS-unit length and syntactic complexity measures (r = .31) was stronger than that between the AS-unit length and fluency measures (r = .14). The effect size of differences between the correlation coefficients was medium (q = .41) when the correlations were compared among the same tasks (i.e., r = .10 vs. .47). In contrast, the effect size of differences between the correlation coefficients was small (q = .18) when the correlations were compared among all the three tasks (i.e., r = .14 vs. .31). Since the results of examining correlations among the same tasks were considered more crucial in this analysis (because the results of examining correlations among all the three tasks were affected also by task effects), it is concluded that the Main Hypothesis 1 was confirmed and that the AS-unit length tends to have stronger relationships with syntactic complexity than fluency, and that the AS-unit length can be considered a syntactic complexity measure rather than a fluency measure. This interpretation of the AS-unit length as syntactic complexity (SC3) was used and F2 was excluded for the subsequent analysis (see section 11.3.3 for further discussion).

5.3.3 Investigating the Hypothesis for Validity 1

In relation to Hypothesis for Validity 1, Table 5.6 shows correlations of the same measure between tasks. For instance, there was a moderate correlation of F1 between Task 1 and Task 4 (r = .49), but there were little correlations of F1 between Task 1 and Task 5 (r = .07) and of F1 between Task 4 and Task 5 (r = .18). It was shown that there were 17 combinations of moderate or strong (r = | .30| or above) correlations: F1 between Tasks 1 and 4; F3 between Tasks 1 and 4; F4 between Tasks 1 and 4, between Tasks 1 and 5, and between Tasks 4 and 5; F5 between Tasks 1 and 4; F6 between

Tasks 1 and 4; A3 between Tasks 4 and 5; SC3 between Tasks 1 and 4, between Tasks 1 and 5, and between Tasks 4 and 5; LC1 between Tasks 1 and 4, between Tasks 1 and 5, and between Tasks 4 and 5; L2 between Tasks 1 and 4 and between Tasks 4 and 5; LC3 between Tasks 4 and 5. However, the moderate or strong correlations of F1, F3, F5, F6, A3, LC2, and LC3 did not appear between all tasks but only between certain tasks, and only the results of F4 (the number of dysfluency markers per minute), SC3 (the AS-unit length; the number of tokens per unit), and LC1 (the Guiraud index: the number of types divided by the square root of the number of tokens) were consistent across tasks, which suggests that only the three measures have moderate generalizability.

Table 5.6
Correlations Between the Values of the Same Measure Derived From the Three Tasks for 16 Speaking Performance Measures

	F1	F3	F4	F5	F6	A1	A2	A3	A4	SC1	SC2	SC3
Task 1 & Task 4	.49**	.36**	.45**	.34**	.58**	-.09	-.01	.11	.08	.29*	.15	.47*
Task 1 & Task 5	.07	-.04	.37**	-.04	.16	.02	.06	.08	.01	.21	-.05	.48**
Task 4 & Task 5	.18	-.10	.53**	.05	.27*	.19	.01	.38**	.05	.09	-.08	.60**

Note. n = 74. Underlined = .30 or more.
*p < .05. **p < .01.

	LC1	LC2	LC3	LC4
Task 1 & Task 4	.59**	.35**	.16	.13
Task 1 & Task 5	.38**	.18	.07	.05
Task 4 & Task 5	.35**	.44**	.32*	-.13

Moreover, there were a limited number of measures that correlated strongly (r = | .50| or more) with the same measure (i.e., F4 between Tasks 4 and 5; F6 between Tasks 1 and 4; SC3 between Tasks 4 and 5; LC1 between Tasks 1 and 4), and the strong correlations did not appear between all tasks but only between certain tasks. This indicates that tasks affect the results of the measures to a large degree and that the measures are unlikely to be strongly generalizable across tasks. It should be noted that

this lack of strong generalizability of measures may have been because the three tasks had very different levels of difficulty (Tasks 5 [the most difficult] > 4 > 1; Koizumi, 2005a). Furthermore, a consistent lack of generalizability of measures within the same construct (e.g., accuracy) may suggest that the construct itself is variable or easily changeable across tasks, as indicated in SLA research (e.g., Ellis, 2001; Fulcher, 1997). If so, this characteristic of lack of strong generalizability can be beneficial because these measures can reflect the constructs well. This is because they may be appropriate for investigating small differences of task characteristics, and because they can be used as "context-sensitive" measures (Selinker & Han, 2001). Whether this characteristic is useful or problematic seems to depend on the purposes of measure use (i.e., whether results are to be generalized across tasks or whether the interpretation is to be limited only to a certain task is acceptable).

5.3.4 Investigating the Hypotheses for Validity 2 and 3 and the Main Hypothesis 2: Validity Argument of the Speaking Performance Measures

In order to examine the Main Hypothesis 2, three Hypotheses for Validity were set. Hypothesis for Validity 3 dealt with the content aspect. The Speaking Performance Measures used in this study were derived from the measures that met all the following criteria: ratio measures used in two or more empirical L2 speaking studies published since 1990, and measures that used the category of "fluency," "accuracy," or "complexity" except for "lexical complexity" (see section 2.3.2). Except for the AS-unit length (the number of tokens per AS-unit), of which the meaning was explored in section 5.3.2, all the measures had one single meaning and there was no disagreement among experts in the previous literature. Therefore, it can be concluded that this hypothesis was satisfied.

As for the Hypothesis for Validity 2, as seen in section 5.2.3.2.2, all the reliability of raters in calculating the measures was high on the Speaking Performance Measures, which leads to support of this hypothesis.

The Hypothesis for Validity 1 was examined in section 5.3.2. There were no measures that correlated strongly (r = | .50| or more) with the same measure across tasks, so all the measures were found to lack strong generalizability across tasks and therefore, this hypothesis was not supported. Thus, interpretation of the Speaking Performance Measures needs to be limited to each task and interpreted as such. Although this study originally intended to generalize results derived from the Speaking Performance Measures, the initial intention was changed to the position that the results of the measures were interpreted in combination with tasks and that only when results were consistent across tasks, they were discussed generally, without confining the interpretation to the case of tasks used.

Therefore, among the three Hypotheses for Validity, two were supported, and there were two types of positive evidence for validity. As an answer to the Main Hypothesis 2, the results demonstrate that the Speaking Performance Measures have positive evidence for validity and since there is an acceptable level of validity, they can be used for the subsequent research.

5.4 Summary of Chapter 5

This chapter (Study 1B) investigates (a) whether the AS-unit length (the number of tokens per AS-unit) indicates fluency or syntactic complexity and (b) whether the Speaking Performance Measures of fluency, accuracy, syntactic complexity, and lexical complexity have positive evidence for validity. The results suggest that the AS-unit length tends to mean syntactic complexity and that the Speaking Performance

Measures have positive evidence for validity. Since there is an acceptable level of validity, they can be used for Study 1C.

Chapter 6 Study 1C: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance[6.1]

6.1 Purpose and a Research Question

The purpose of this chapter is to examine relationships between size of productive vocabulary knowledge and speaking performance, which is the main focus in Part Ⅱ. In order to ensure this research is of a high standard, two studies were conducted in the previous chapters (Chapters 4 to 5). Figure 6.1 shows where the present chapter (i.e., Chapter 6 Study 1C: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance) is located in the overall research. One Research Question was addressed for this purpose.

Research Question: What aspects of speaking performance are related to size of productive vocabulary knowledge to a moderate or strong degree?

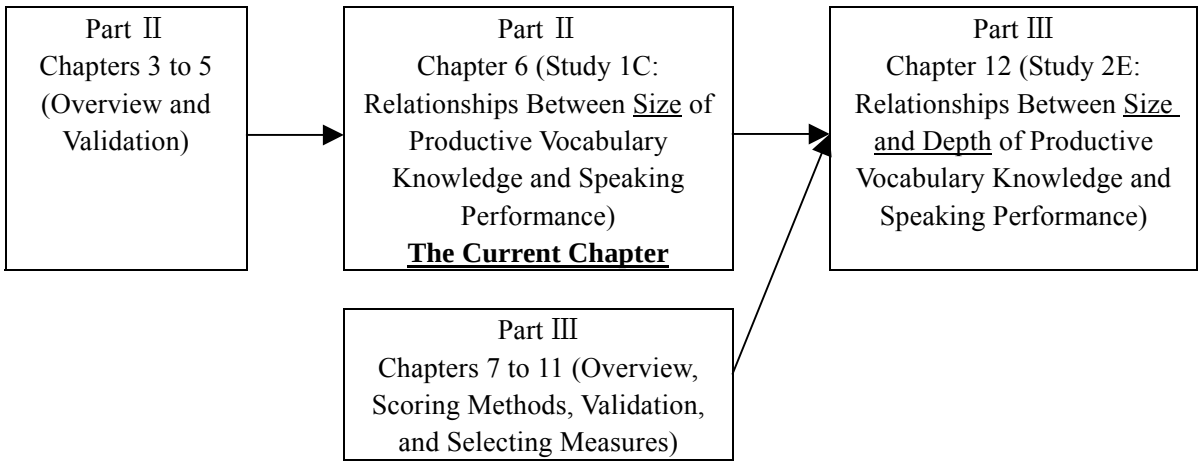


Figure 6.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

6.2 Method

6.2.1 Participants

The participants analyzed in this study were 73 Japanese lower and upper secondary school students (see Table 6.1). They were almost the same as those who were analyzed in Chapter 5 (n = 74). The only difference was that one test taker had to be excluded who did not take the Productive Vocabulary Knowledge Test (Productive VKT).

Table 6.1
Number of Participants at Each School

Type of school (School year)	*n*
Public junior high school [School B] (3rd year)	28
Public junior high school [School C] (3rd year)	28
Prefectural senior high school [School D] (1st year)	4
Senior high school affiliated with a national university [School E] (1st year)	13
Total	73

6.2.2 Instruments

The Productive VKT and the Speaking Test were used. The details were described in sections 3.2.2.1 and 3.2.2.4.

6.2.3 Procedures

6.2.3.1 Test Administration Procedures

The details were described in section 3.2.3.1.

6.2.3.2 Scoring and Coding Procedures

The Productive VKT and the Speaking Performance Measures were scored and computed as described in sections 4.2.3.2 and 5.2.3.2. In this study, the logit scores derived from Method 2 of the Productive VKT was used. Sixteen Speaking Performance Measures were utilized, as seen in Table 6.2.

Table 6.2
Summary of 16 Speaking Performance Measures Used in Study 1C

Construct	Code	Measure
Fluency	F1	No. of tokens per minute
	F3	No. of clauses per minute
	F4	No. of dysfluency markers per minute
	F5	No. of dysfluency markers per token
	F6	No. of dysfluency markers per AS-unit
Accuracy	A1	No. of error-free clauses per clause
	A2	No. of error-free AS-units per AS-unit
	A3	No. of errors per token
	A4	No. of errors per AS-unit
Syntactic complexity	SC1	No. of clauses per AS-unit
	SC2	No. of subordinate clauses per AS-unit
	SC3	No. of tokens per AS-unit
Lexical complexity	LC1	Guiraud index: No. of types divided by the square root of No. of tokens
	LC2	Lexical density: No. of lexical tokens per token
	LC3	Weighted lexical density: No. of sophisticated lexical tokens and No. of basic lexical tokens given half the weight divided by No. of tokens
	LC4	No. of sophisticated word types per word

Note. F2 were excluded based on the results in section 5.3.2.

6.2.3.3 Analysis Procedures

Simple regression analyses were utilized with the Speaking Performance Measures as dependent variables and with productive vocabulary knowledge as an independent variable. Forty-eight (i.e., 16*3) analyses were performed repeatedly. Simple regression analyses, not simple correlations, were used in order to utilize a similar method to that used in Study 2 (2E). The interpretation of adjusted R^2 (the proportion explained) was based on Cohen (1988, pp. 413-414): Lower than small < .0196; .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large (see section 2.6.4.1).

6.3 Results and Discussion

Table 6.3 shows means and standard deviations for basic statistics, the Speaking Performance Measures, and the Productive VKT. The standard deviations of speaking time were large across the three tasks.

Table 6.3
Descriptive Statistics of the Basic Data, the Speaking Performance Measures, and the Productive Vocabulary Knowledge Test

	Task 1		Task 4		Task 5	
	M	*SD*	*M*	*SD*	*M*	*SD*
Speaking time (second)	42.33	25.07	35.21	24.04	58.89	29.95
No. of tokens (pruned tokens)	26.95	18.47	19.68	15.52	30.64	21.39
No. of AS-unit	4.93	2.86	3.40	2.09	5.26	2.89
No. of clauses	5.26	3.11	3.88	2.59	3.78	2.95
No. of dysfluency markers	3.52	4.83	3.70	6.21	6.51	8.66
No. of error-free clauses	3.36	2.07	1.99	1.76	0.74	1.56
No. of error-free AS-units	2.78	1.94	1.51	1.44	0.79	1.54
No. of errors	3.19	2.73	2.59	2.30	9.59	6.69
No. of subordinate clauses	0.11	0.39	0.59	0.78	0.11	0.52
No. of types	19.44	11.73	14.03	8.88	17.51	9.67
No. of lexical tokens	15.14	9.93	10.70	7.60	12.19	7.15
No. of lexical tokens within basic tokens	12.79	8.41	9.62	6.79	8.99	5.72
No. of sophisticated word types	1.97	1.79	0.96	0.99	2.48	1.58
F1 (No. of tokens per minute)	46.91	30.73	40.39	26.72	34.60	20.08
F3 (No. of clauses per minute)	9.82	6.77	8.77	7.51	5.05	4.98
F4 (No. of dysfluency markers per minute)	4.21	4.47	4.91	5.74	5.56	6.79
F5 (No. of dysfluency markers per token)	0.12	0.13	0.16	0.22	0.29	0.74
F6 (No. of dysfluency markers per AS-unit)	0.65	0.71	0.89	1.04	1.28	1.93
A1 (No. of error-free clauses per clause)	0.68	0.22	0.48	0.30	0.11	0.21
A2 (No. of error-free AS-units per AS-unit)	0.61	0.25	0.41	0.32	0.11	0.21
A3 (No. of errors per token)	0.11	0.09	0.15	0.10	0.40	0.27
A4 (No. of errors per AS-unit)	0.58	0.45	0.81	0.56	0.78	0.66
SC1 (No. of clauses per AS-unit)	1.08	0.23	1.18	0.35	0.74	0.32
SC2 (No. of subordinate clauses per AS-unit)	0.02	0.06	0.21	0.30	0.04	0.24
SC3 or F2 (No. of tokens per AS-unit)	5.24	1.15	5.72	1.96	5.50	1.67
LC1 (Guiraud index)	3.66	1.05	3.10	0.84	3.15	0.91
LC2 (Lexical density)	0.58	0.08	0.58	0.11	0.44	0.16
LC3 (Weighted lexical density)	0.33	0.06	0.32	0.07	0.28	0.11
LC4 (No. of sophisticated word types per token)	0.08	0.07	0.05	0.07	0.11	0.09
Productive Vocabulary Knowledge Test (logit)	-0.62	2.11				

Note. n = 73. KR21 (Kuder-Richardson 21 formula) of the Productive VKT = 1.00 (.999) using CALS (n.d.).

As seen in Table 6.4, some values of the Speaking Performance Measures were predicted by the Productive VKT scores to some degree, whereas the other values were not. For example, 17% of Task 1 F4 (the number of dysfluency markers per minute)

was predicted by productive vocabulary knowledge. On the other hand, only 3% of Task 1 F5 (the number of dysfluency markers per token) was predicted by productive vocabulary knowledge.

Table 6.4
Simple Regression Analysis Relating Productive Vocabulary Knowledge to Each Speaking Performance Measure

	Task 1		Task 4		Task 5	
	Adjusted R^2	β	Adjusted R^2	β	Adjusted R^2	β
F1	-.01	-.01	.00	.11	.05	.25
F3	-.01	-.09	-.01	-.02	-.01	-.01
F4	.17	.42	.06	.27	.11	.35
F5	.03	.20	-.01	.05	-.01	-.09
F6	.10	.33	.09	.32	.00	.10
A1	.01	-.17	.11	.35	.20	.46
A2	.00	-.11	.01	.14	.23	.49
A3	.00	-.13	.10	-.33	.29	-.55
A4	-.01	.08	-.01	-.07	.15	-.41
SC1	.04	.24	.09	.32	.06	.28
SC2	.01	.16	.10	.33	-.01	.08
SC3	.33	.58	.28	.54	.41	.65
LC1	.26	.52	.41	.65	.21	.47
LC2	.14	-.38	.08	-.31	.12	-.37
LC3	.03	-.20	.04	-.24	.10	-.34
LC4	-.01	.05	-.01	-.02	.02	-.17

Note. n = 73. Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large. See Appendixes 6.1 and 6.2.

Medium or strong effects (adjusted R^2 = .13 or more) of productive vocabulary knowledge were observed in Task 1 F4, Task 1 SC3, Task 1 LC1, Task 1 LC2, Task 4 SC3, Task 4 LC1, Task 5 A1, Task 5 A2, Task 5 A3, Task 5 A4, Task 5 SC3, and Task 5 LC1. In particular, strong effects (adjusted R^2 = .26 or more) were found in Task 5 A3, Task 1 SC3, Task 4 SC3, Task 5 SC3, Task 1 LC1, and Task 4 LC1.

Consistent moderate or strong effects of productive vocabulary knowledge across tasks were found with SC3 (the number of tokens per AS-unit) and LC1 (the

Guiraud index: the number of types divided by the square root of the number of tokens). This suggests that those who have more productive vocabulary knowledge tend to produce a larger number of words within one AS-unit and more lexical variation (i.e., better syntactic complexity and lexical complexity in speaking performance). Additionally, when the formula of computing the two measures was examined, both SC3 ([the number of tokens]/[the number of AS-units]) and LC1 ([the number of types]/√[the number of tokens]) had vocabulary volume (i.e., the number of tokens and types) on the numerators. Therefore, it can be interpreted that those who have more size of productive vocabulary knowledge may produce more tokens and types and may excel in vocabulary aspects of speaking performance. This was used for setting the hypotheses in Study 2 (2E).

One problem with the hypothesis above is that results of F1 (the number of tokens per minute) were not in line with this interpretation. However, this may be related to how speaking time was measured in this study (Study 1C). F1 was calculated using the following formula: [the number of tokens]/[speaking time, as measured by seconds]*60. The large standard deviations of speaking time (see Table 6.3) indicate that speaking time varied substantially according to test takers. That is, some of them began speaking fast but soon stopped speaking, resulting in shorter speaking time with fast speech (high values of F1). In contrast, others tried to speak as much as possible by exploring words and structures in order to express what they had in mind, which led to longer speaking time with slower speech (low values of F1). The same logic can be applied to F3 (the number of clauses per minute) and F4 (the number of dysfluency markers per minute), which have speaking time in the denominator. The method of measuring speaking time may also have affected not only F1, F3, and F4 but also the other results because the speaking time may affect the number of tokens, AS-units, and

clauses uttered. Although the speaking time was measured in this way in order to avoid putting excessive pressure to speak on test takers for a pedagogical reason, for research purposes, it may be necessary to measure speaking time from the beginning of the speech until the specified time in order to avoid very different values in the denominator according to test takers. In Study 2 (2D and 2E), this problem was overcome by modifying the method of measuring time.

Moderate or strong effects of productive vocabulary knowledge revealed only with a certain task were found with Task 1 F4, Task 5 A1, Task 5 A2, Task 5 A3, Task 5 A4, and Task 1 LC2. As for Task 1 F4 (the number of dysfluency markers per minute), because of the problem discussed above, the result may be inconclusive. Regarding Task 1, LC2 (Lexical density: the number of lexical tokens per token) was predicted by productive vocabulary knowledge to a moderate degree. Concerning Task 5, all the accuracy measures (A1, A2, A3, and A4) were predicted by productive vocabulary knowledge to a moderate degree. Why only these measures were predicted more than others is difficult to interpret because the three tasks varied not only in difficulty (Tasks 5 [the most difficult] > 4 > 1; see Koizumi, 2005a) but also in other task characteristics (e.g., Tasks 1 and 5 with no planning time and Task 4 with planning time of 60 seconds). Study 2 (2E) was conducted to enhance interpretability by including tasks with similar difficulty and characteristics.

6.4 Limitations of Study 1C and Modifications for Study 2 (2E)

This study (Study 1C) was conducted as a pilot study of Study 2 (2E). Since there were the following six limitations in Study 1C, Study 2E needed to be conducted by improving the methodology used in Study 1C (see Table 6.5 for a summary).

Table 6.5
Comparisons Between Studies 1C and 2E

Study 1C	Study 2E
Only vocabulary size was dealt with.	Both vocabulary size and depth were dealt with.
The number of the measures of lexical complexity was limited.	The number of the measures of lexical complexity was increased.
Only three speaking tasks of varying difficulty and characteristics were used.	Tasks of similar difficulty and characteristics were used. The same two tasks in Study 1C were also used to enhance interpretation.
Speaking time varied substantially according to test takers. The face-to-face interview was used to elicit speech.	A specified time (60 seconds) was given to all test takers. The tape-mediated format was used to elicit speech.
Aspects examined in validation were rather limited.	Aspects examined in validation were expanded.
The Productive VKT suffered a lack of content representativeness.	The item selection procedure of selecting words randomly from a vocabulary list was taken to satisfy content representativeness.

First, Study 1C included size of productive vocabulary knowledge in examining relationships with speaking performance, while Study 2E added depth aspects of productive vocabulary knowledge to the investigation. This expansion of research targets (from only size to both size and depth) is essential in Study 2E because vocabulary depth as well as size can be considered integral to speaking performance based on Levelt (1993) and because, to my knowledge, there were no empirical examinations dealing with this topic directly.

Second, Study 1C used the measures used in the previous speaking studies (see section 5.2.3.2.2). The main results of Study 1C were that there were moderate or strong relationships across tasks between size of productive vocabulary knowledge and the vocabulary-related measures (i.e., SC3 [the AS-unit length: the number of tokens per AS-unit] and L1 [the Guiraud index: the number of types divided by the square root of the number of tokens]). However, the number of measures of lexical complexity was limited and only four measures of lexical complexity were used. Based on the results of Study 1C, which shows the potential importance of examining

relationships between productive vocabulary knowledge and vocabulary in speaking performance, it was decided that Study 2E used more measures of lexical complexity in order to focus more on vocabulary performance. In addition, some of the Speaking Performance Measures of accuracy and syntactic complexity and the criteria of sophisticated words were also added or modified (see section 10.2.3.2 for specific reasons).

Third, Study 1C used only three tasks to elicit speaking performance, which had its own difficulty and characteristics, as mentioned in section 6.3. Therefore, it was rather difficult to interpret the different results of the tasks. Study 2E included tasks of similar difficulty and characteristics by using only tasks without pre-task planning time to enhance interpretability. Moreover, the same two tasks in Study 1C (Tasks 1 and 5) were also used in Study 2E so that the interpretation may deepen by comparing results of Study 1 and Study 2E.

Fourth, the potential problematic nature of the measuring method of speaking time in Study 1 was discussed in section 6.3. Study 2E gave a specified time (60 seconds) to all test takers. One problem with this method is the possibility of pressuring test takers who cannot keep speaking until the specified time finishes. In order to overcome this problem, the elicitation of speaking performance was changed from the face-to-face interview (Study 1C) to a tape-mediated format (Study 2E). The latter format is one of the semi-direct formats and has another advantage in the following two aspects: (a) in reducing the number of factors affecting speaking performance (e.g., effects of interlocutors; e.g., Brown, A., 2003, 2004; Lazaraton, 1992, 1996; O'Sullivan, 2002, 2004), and (b) in practicality, because collecting and training interviewers and finding rooms for each interviewer to conduct interviews were not needed. Since both Studies 1C and 2E focused on monologues of speaking

performance, the lack of interviewers and inability to elicit interactive speaking performance were not considered to be serious problems (see section 2.3.2). Furthermore, practicality is important when a large number of participants need to be gathered in order to obtain more representative samples of the target population (Japanese learners of English at the novice level), and to get stable results in using multivariate analyses in Study 2E. There is one disadvantage in changing the formats, however. Differences in formats can affect speaking performance, resulting in different results in Studies 1C and 2E. Despite this weakness, it was decided to change the formats because the advantages outweighed the disadvantages.

Fifth, although Study 1 (1A and 1B) was conducted after examining the validity of inferences based on the scores of Productive VKT and the values of the Speaking Performance Measures, aspects examined in validation were rather limited. Study 2 (2A and 2B) expanded the aspects to be covered in validation so that the test scores and values of the measures can be used with more appropriacy and meaningfulness. Additionally, as for scoring methods in the Productive VKT, three scoring methods were compared in Study 1 (1A), but the criteria of scoring methods were improved and compared again in Study 2 (2A) in order to increase the validity, specifically the generalizability aspect of the validity.

Sixth, one limitation of the Productive VKT in Study 1C was a lack of content representativeness of vocabulary size, which leads to the difficulty of generalizing results and obtaining vocabulary size estimates, because words tested were not selected randomly. In contrast, Study 2E changed the item selection procedure and developed a new version of the Productive VKT that can estimate vocabulary size of test takers by selecting words randomly from a vocabulary list. Additionally, the vocabulary list used in Study 1C was based on the Hokkaido University English

Vocabulary List and the *Longman Dictionary of Contemporary English* (3rd ed.; see section 3.2.2.1). On the other hand, the Productive VKT used in Study 2E was developed based on the JACET8000, which was specifically made for Japanese learners of English (JACET Basic Word Revision Committee, 2003; see section 7.2.2.1). Since the target learners in this study were Japanese learners of English at the novice level, using the JACET8000 can be considered more appropriate, which can help enhance the content aspect of validity.

6.5 Summary of Chapter 6

This chapter examines relationships between size of productive vocabulary knowledge and speaking performance. As a result, consistent moderate or strong effects of productive vocabulary knowledge across tasks were found with SC3 (the number of tokens per AS-unit) and LC1 (the Guiraud index: the number of types divided by the square root of the number of tokens). This suggests that those who have more productive vocabulary knowledge tend to produce a larger number of words within one AS-unit and more lexical variation (i.e., better syntactic complexity and lexical complexity in speaking performance).

Part Ⅲ Study 2: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance

Chapter 7 Overview of Study 2 [7.1]

7.1 Purpose

The main purpose of Part Ⅲ (from Chapters 7 to 12) is to examine relationships between size and depth of productive vocabulary knowledge and speaking performance, with a specific focus on Japanese learners of English at the novice level. In order to achieve this purpose, Chapter 7 (this chapter) describes the method used in the subsequent chapters. Chapter 8 examines which scoring method of the written version of the Productive Vocabulary Knowledge Test (Productive VKT) is better in terms of relationships with the oral version of the Productive VKT. Chapter 9 examines the validity of inferences based on the scores of the Productive VKT including the Size and Depth Sections. Chapter 10 examines which Speaking Performance Measures need to be included in the research. Chapter 11 investigates the validity of inferences based on values of the Speaking Performance Measures. Then relationships between size and depth of productive vocabulary knowledge and speaking performance are investigated in Chapter 12. Figure 7.1 shows where the

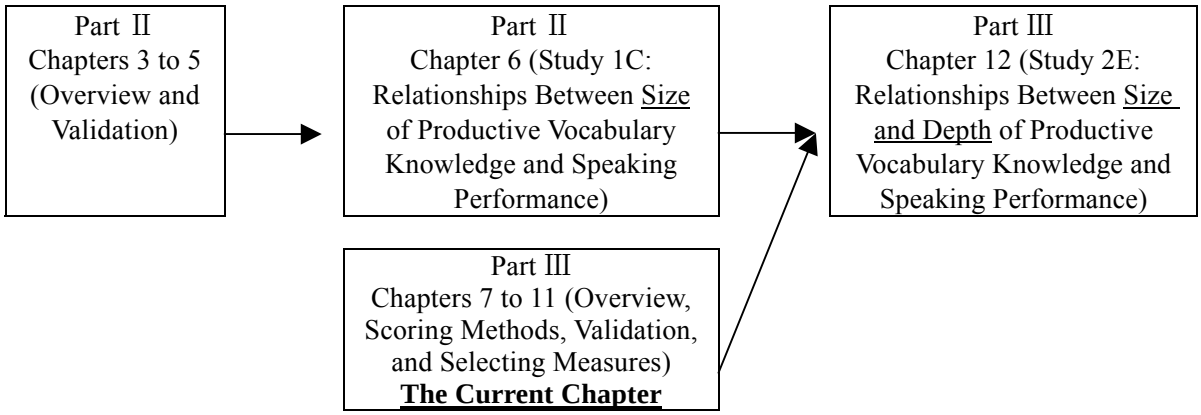


Figure 7.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

present chapter (i.e., Chapter 7 Overview of Study 2) is located in the overall research.

7.2 Method

7.2.1 Participants

Participants were described in detail in Chapters 8 to 12 (see Appendixes 7.1 to 7.4).

7.2.2 Instruments

This section described the two main instruments. Instruments used only in certain chapters were described in each chapter.

Two tests were used: (a) the Productive VKT and (b) the Speaking Test. For each test, a questionnaire (the Productive VKT Questionnaire and the Speaking Test Questionnaire) was conducted.

7.2.2.1 Productive Vocabulary Knowledge Test

In order to assess productive vocabulary knowledge, there were two sections in the Productive VKT: the Size and Depth Sections (see Appendix 7.6).

7.2.2.1.1 Size Section

The Size Section of the Productive VKT was designed to assess size of productive vocabulary knowledge for beginning learners of English. Size of productive vocabulary knowledge consists of (1) vocabulary size in Chapelle's (1994) second dimension. To be more specific, size of productive vocabulary knowledge is defined as having knowledge of (d) a written form and (h) form and meaning in Nation's (2001)

vocabulary knowledge framework (see section 2.2.1 Table 2.7), and as having knowledge of producing (recalling) a form related to a primary meaning in Laufer et al.'s (2004) classification. The detailed description of test purpose and design was based on Read and Chapelle's (2001) vocabulary assessment framework (see Appendix 7.5).

As the method to assess productive vocabulary knowledge, a translation format was selected, in which a Japanese (L1) meaning was presented to elicit an equivalent English (L2) word (see Table 7.1). The first letter of the word was provided as a clue to avoid a word that had the same meaning but was different from the answer. This format was the same as that used in Study 1 (see section 3.2.2.1 for details).

Table 7.1
Examples in the Size Section of the Productive Vocabulary Knowledge Test

Write the English word that best corresponds to the Japanese meaning on your answer sheet. The first letter of the English word is already given. Example: 町 (t_____) The answer is *town* (*towns*). ふれる、さわる (t) The answer is (touch).

Note. All was originally written in Japanese. Both singular and plural forms are correct answers.

In developing the Size Section of the Productive VKT, four steps were taken. First, words tested were selected from Levels 1 to 3 (up to 3,000 the most frequent lemmas; see section 2.2.1 for the methods used to count words) in the JACET8000, which was specifically tailored for Japanese learners of English (JACET Basic Word Revision Committee, 2003). The JACET8000 was selected as a basis of making the Productive VKT because it was made for Japanese learners of English based on both the British National Corpus and a list of words ranked according to the frequency in which Japanese learners encounter English words. The JACET8000 is based on the corpus of spoken and written language. Levels 1 to 3 were selected because they were

considered to cover all necessary words in speaking for novice learners. As seen in section 2.4.1, Adolphs and Schmitt (2003) found that approximately 95% of words in everyday (interactive) speaking performance belonged to 2,000 word families or 3,000 lemmas. Word selection of 32 words was done for each level using the RAND function in an Excel 2000 sheet (see Mochizuki et al., 2003, pp. 203-206, for the random word selection procedure). In total, 96 words were selected (i.e., 32 words * 3 levels). In addition, in order to check the validity of the Productive VKT, 14 keywords were included that appear commonly in the Size Section of the Productive VKT and the Speaking Test (see also section 7.2.2.2). Therefore, 108 words were tested in total (i.e., 96 + 12). The following two types of randomly selected words were not used: (a) words of which English sounds and Japanese translated meanings were similar (e.g., *cup* and *ah*) and (b) words that started with the same first letter and had similar meanings (e.g., *tale/talk*, *ally/alliance*, and *inclusive/included*). Then, the next candidate words were chosen after being selected from the random selection.

Second, the primary meaning (i.e., the meaning that appeared first in dictionaries) was searched for using the three dictionaries: (a) the *Longman Dictionary of Contemporary English* (4th ed.; Summers et al., 2003), (b) the *Oxford Advanced Learner's Dictionary* (Oxford University Press, 2004), and (c) the *Cambridge Dictionary of American English* (Cambridge University Press, 2004). They were selected for three reasons: First, their meanings were ordered according to the frequency of words, not according to historical order. Second, the three dictionaries covered both British and American English. Third, (b) and (c) could be used on the internet. Words were not used for the Size Section when one or more dictionaries out of the three did not have the target word in their entry or when all the dictionaries did not share the same meaning in the first entry (e.g., fix, fortune, and farming).

The third step in making the Size Section was to select appropriate translations of the primary meaning by searching the following three English-Japanese dictionaries: (a) the *Taishukan's Unabridged Genius English-Japanese Dictionary* (Konishi & Minamide, 2002), (b) the *New English-Japanese Dictionary* (Kenkyusha, 2005), and (c) the *Progressive English-Japanese Dictionary* (Yahoo Japan, 2005). The three were selected because they were dictionaries of middle size or more and because (b) and (c) could be used on the internet. Translations that were easy to understand were selected for the Size Section. Words were not used for the Size Section when the primary meaning had several translations in Japanese (e.g., flight) or when the primary meaning seemed different from the meaning that Japanese learners remember first (e.g., recipe, engineering, and plain), which may be confusing to learners.

Fourth, the words were ordered in the Size Section as follows: Level 1, Level 2, Level 3, Level 1, Level 2, and so on. Then, the position of 14 keywords was selected randomly and the words already in the position were placed at the end of the Size Section.

7.2.2.1.2 Depth Section

The Depth Section of the Productive VKT was designed to assess depth of productive vocabulary knowledge for beginning learners of English. Depth of productive vocabulary knowledge consists of (2) the knowledge of word characteristics in Chapelle's (1994) second dimension. To be more specific, depth of productive vocabulary knowledge is defined as having knowledge of (d) a written form, (f) word parts, (l) associations, and (p) collocations in Nation's (2001) vocabulary knowledge framework (see section 2.2.1 Table 2.7), and as having knowledge of producing (recalling) a form related to a stimulus word based on Laufer et al.'s (2004)

classification. The detailed description of test purpose and design was based on Read and Chapelle's (2001) vocabulary assessment framework (see Appendix 7.5).

Three aspects of depth (i.e., derivations, antonyms, and collocations) were selected for the following three reasons. First, among depth aspects, associations between words were selected because speaking involves forming an utterance by connecting words. According to Aitchison (2003), the first and second most common responses in word association tests (i.e., the first and second most common and important word associations) are antonyms ("coordination" is Aitchison's term) and collocations (p. 86). Thus, this study decided to use the two aspects.

Second, synonyms and superordinates were not used because knowledge of synonyms and superordinates may overlap knowledge of size to a large degree. For example, the word *beautiful* has a primary meaning of "extremely attractive to look at" (Summers et al., 2003, p. 137) and knowing a written form in connection with a primary meaning is knowledge of vocabulary size. Additionally, one of the synonyms of the word *beautiful* is *attractive* and this is considered knowledge of vocabulary depth.[7.2] In contrast, antonyms may be much more independent of the primary meaning of the word than synonyms. For instance, one of the antonyms of the word *beautiful* is *ugly*, but *not ugly* does not mean the same as *beautiful*, so *ugly* does not belong to the primary meaning of *beautiful*. From another perspective, some may say that in writing antonyms (and also collocates), the (primary) meanings may be involved in the process, so size aspects are also assessed. For example, when test takers write the word *ugly* as an antonym of *beautiful*, they search for the meaning of *beautiful* first, and think of the opposite of the meaning of *beautiful* next, and then the word *ugly* among many possible opposite words is elicited (*beautiful* → attractive → not attractive → *ugly*). Since the knowledge of form and meaning is the most basic aspect of vocabulary

knowledge (Laufer et al., 2004; Laufer & Goldstein, 2004; Webb, 2005), it is difficult not to assess any of the size aspects at all. However, this study attempted to avoid shared areas of size and depth as much as possible and to select depth aspects that certainly have independent elements from size aspects.

Third, attempts were made to cover the three aspects of Nation's (2001) vocabulary knowledge framework (i.e., form, meaning, and use) in order to increase content coverage of productive vocabulary knowledge. If antonyms are considered to belong to "meaning" and synonyms are regarded as belonging to "use," some aspects included in "form" were sought. The only area that can be assessed as depth in the category of "form" was "word parts." There were two candidates: derivations and inflections. Inflections were not tested because the range of possible answers was limited (e.g., adding *s* or *es* to make plural forms of nouns). Therefore, derivations were chosen.

This study did not assume there is hierarchy between the three aspects (i.e., derivatives, antonyms, collocations). Each subsection is analyzed separately, but it was interpreted that the more points test takers can obtain, the deeper productive vocabulary knowledge they have. Each subsection had both the same words and different words in the Size Section in order to enhance comparability and control the difficulty (see Appendix 7.7).

7.2.2.1.2.1 Derivation Subsection

Derivations are often defined as "the formation of new words by adding affixes to other words or morphemes" (Richards & Schmidt, 2002, p. 151). This study included both derivations and back formations under the category of derivations and defined derivations as "the formation of new words by adding affixes to other words or

morphemes" (Richards & Schmidt, p. 151) and by removing "an affix from an existing word" (p. 45).

There were 20 items that can be categorized into two types in the Derivation Subsection: (a) items to add a derivational suffix to a stimulus word and make a derivative (e.g., from *work*, write *worker*; 12 items) and (b) items to take a derivational suffix from a stimulus word and make a back formation word (e.g., from *comfortable*, write *comfort*; 8 items). Derivational suffixes were selected from the Levels 2 to 4 in Bauer and Nation (1993; see Appendix 7.8). All selected words used as stimuli and their derivational and back formation words were both within Levels 1 to 3 in the JACET8000 (JACET Basic Word Revision Committee, 2003). Only derivational suffixes, not derivational prefixes, were included in the Derivation Subsection, although some derivational prefixes were tested in the Antonym Subsection. Words that have more than one answer were not included in the Derivation Subsection (e.g., *inform* [Noun: a person who does …], Answer: *informant*, *informer*)

As the method to assess productive vocabulary knowledge, the format was selected in which a stimulus word was presented to elicit a derivative or a back formation word (see Table 7.2 for examples). Due to lack of testing time, this study specified a part of speech of the answer (e.g., a noun) and presented a hint of a concrete meaning (e.g., a person who does …) in order to elicit one single derivational or back formation word for a stimulus word, instead of having test takers write three possible derivatives (Schmitt & Meara, 1997). One possibility was that test takers did not know parts of speech, so the explanation of parts of speech was put in the instruction of the Derivation Subsection. Explanation of parts of speech was based on Swan (1995) and Mochizuki (1996, pp. 28-29). Among parts of speech, nouns, verbs, and adjectives were tested because they are "the three major word classes in English"

(Aitchison, 2003, p. 112). Additionally, adverbs were not tested because they seemed

Table 7.2
Examples in the Derivation Subsection of the Productive Vocabulary Knowledge Test

Make the English word the word that has the part of speech written in [ ]. Write only one word. Do not write words with *ing* and *ed*.
Example: kind [Noun: the nature of …] The answer is (kindness).
＊Nouns are words that mainly express names of people or things (e.g., *cat*).
Example: This is my cat. → *cat* is a noun.
Example: introduction [Verb: do the action of …] The answer is (introduce).
＊Verbs are words that mainly express actions (e.g., *run*) and states (e.g., *live*)
Example: I live in Japan. → *live* is a verb.
Example: wonder [Adjective: the state of …] The answer is (wonderful).
＊Adjectives are words that describe people, things, and events (e.g., *cute*)
Example: I have a cute cat. → *cute* is an adjective.

Note. All was originally written in Japanese.

difficult, especially for junior high school students. Parts of speech of the answers were presented but not those of the stimulus words because parts of speech can be difficult to understand for junior high school students and because attempts were made not to use elements of parts of speech as far as possible.

7.2.2.1.2.2 Antonym Subsection

Antonyms is defined as "a word which is opposite in meaning to another word" (Richards & Schmidt, 2002, p. 27) and it can be classified into four types (Mochizuki et al., 2003, pp. 54-56): (a) complementary (e.g, *alive* and *dead*), (b) converseness (e.g., *buy* and *sell*), (c) antonymy (e.g., *hungry* and *full*), and (d) incompatibility (e.g., *spring, summer, fall*, and *winter*). The Antonym Subsection had three types ((a), (b), and (c)) out of the four. Items that tested knowledge of (c) antonymy contained the most items because it is the most common antonym type (Mochizuki et al., p. 55). The reason (d) incompatibility was not included was that its concept seemed to be more difficult for junior high school students. Words with at least one possible answer belonging to

Levels 1 to 3 in the JACET8000 (JACET Basic Words Revision Committee, 2003) were selected. There were two types of answer included in the subsection: (x) antonyms whose forms were totally different from stimulus words (e.g., *high* and *low*), and (y) ones that could be derived taking derivational prefixes from the stimulus words or adding them to the stimulus words (e.g., *independence* and *dependence*; see Appendix 7.9).

The Antonym Subsection and the Collocation Subsection were both based on Shimamoto (2005). One difference from her format was that she asked test takers to write as many answers as possible whereas this study required them to write only one answer (see Table 7.3 for examples). The reason was that, first, the latter format may not be affected by test takers' individual characteristics (see section 2.2.2) and can elicit the most appropriate associations that test takers think they are, and second, less time was needed to administer the test (see Appendix 7.2).

Preliminary content analysis of the Antonym Subsection showed that one item (Item 9, *like*) may be problematic because it is a homograph (i.e., a word with the same spelling but with a different meaning from another). Therefore, it was excluded from further analysis. Therefore, 17 items were analyzed.

Table 7.3
Examples in the Antonym Subsection of the Productive Vocabulary Knowledge Test

Write one word that has the opposite meaning to the word presented. Example: good ( ) An example answer is (bad).
1. open ( ) (An example answer: close, shut)

Note. All was originally written in Japanese.

7.2.2.1.2.3 Collocation Subsection

Collocations refer to "the tendency of two or more words to co-occur in

discourse" (Schmitt, 2000, p. 76). According to Schmitt (2000), collocations can be divided into three types: (a) grammatical/syntactic collocations (e.g., *access to*), (b) semantic/lexical collocations (e.g., *spend money*), and (c) collocations not "based on grammatical or semantic patterning" (e.g., [not *at* but] *on Monday*; p. 77). The current study only used the second type. Schmitt (2000) defined collocations as words (x) occurring together, and (y) with "varying degrees of exclusivity" (p. 77),[7.3] but this study considered two words next to each other as collocations if the first requirement was satisfied and accepted more types of word combinations as collocations. In other words, free combinations (e.g., *strong dog*, *play soccer*) as well as two words next to each other with "varying degrees of exclusivity" (e.g., *hot dog*, *pay attention*) were included as collocations in this study. Furthermore, as for transparency of meaning, both transparent and opaque words were considered collocations.[7.4]

The test format based on Shimamoto (2005) elicited a word that comes with a stimulus word by specifying a part of speech (see Tables 7.4 to 7.6). There were three types. First, an adjective or a noun was elicited in response to a noun (Format A; see Table 7.4). Second, a noun was elicited in response to an adjective (Format B; see Table 7.5). Third, a noun was elicited in response to a verb (Format C; see Table 7.6). Formats A and B dealt with noun phrases (e.g., *clean air*), whereas Format C dealt with verb phrases (e.g., *play baseball*). All three types were combined and analyzed together and test scores were interpreted as knowledge of collocations. As for Format A, some adjectives (e.g., *good, better*) that can occur with many nouns were not allowed as answers in Format A. These adjectives were derived from a pilot study.

As for stimulus words, the following two types were selected: (a) words that can occur with a variety of words (e.g., *dog* and *buy*) and (b) ones that can only occur with a limited number of words (e.g., *tragedy* and *religion*).[7.5]

Table 7.4
Examples in the Collocation Subsection of the Productive Vocabulary Knowledge Test (Format A: (Adjective or Noun) + Noun)

Write one English word that fits (　) (an adjective or noun). Do not write *good, better, best, bad, nice, wonderful, great, big, small,* or *old*. The same words can be used as answers. Example of an adjective: (　　) air　An example answer: (clean) air Example of a noun:　(　　) cream　An example answer: (ice) cream

Note. All was originally written in Japanese.

Table 7.5
Examples in the Collocation Subsection of the Productive Vocabulary Knowledge Test (Format B: Adjective + (Noun))

Write one English word that fits (　) (a noun). The same words can be used as answers. Example: clean (　　)　An example answer: clean (air)

Note. All was originally written in Japanese.

Table 7.6
Examples in the Collocation Subsection of the Productive Vocabulary Knowledge Test (Format C: Verb + (Noun))

Write one English word that fits (　) (a noun). The same words can be used as answers. Example: play (　　)　An example answer: play (baseball) Note: (a/an/the) means that nouns with *a*, *an*, and *the* or without them can be answers.

Note. All was originally written in Japanese.

Contexts were not provided in items in the Collocation Subsection (e.g., Fill in the blank with an appropriate word. *It was a* (　) *birthday.*; see section 2.2.2) because giving contexts was not in line with the definition of productive vocabulary knowledge (i.e., knowledge to produce vocabulary forms without the context of language use; see section 2.6.3; see also Appendix 7.2).

7.2.2.2 Speaking Test

The Speaking Test was designed to elicit speaking performance, specifically, fluency, accuracy, syntactic complexity, and lexical complexity, from Japanese beginning learners of English. The target speaking performance was monologues without pre-task planning time. The task types investigated were description and comparison (see section 2.3.1 for reasons; see Appendix 7.10).

Study 2 narrowed aspects and types of speaking performance in investigation (from "description, comparison, and explanation with and without planning time" in Study 1 to "description and comparison without planning time" in Study 2) but used multiple (two or three) tasks (e.g., a self-introduction task and two description tasks using two pictures) to assess one aspect (e.g., monologue without planning time) and one type (e.g., description). Since Study 1 revealed that there are some task differences in relationships between productive vocabulary knowledge and speaking performance (see section 6.3), using multiple tasks was considered more appropriate than widening aspects and types of speaking performance in investigation and using a single task to assess one aspect and one type.

This study used a tape-mediated format in eliciting speaking performance. The reason for this choice was explained in section 6.4. Since test takers did not have a person to talk to in front of them in this format, a more detailed and possibly more authentic speaking situation than Study 1 was set as follows: "An assistant language teacher (Ms. Smith, whose picture is on the test booklet) is a new teacher. She wants to know your English ability, so you will have an English speaking test. Speak as if you were speaking to her." In other words, the listener and the purpose of speaking and of conducting the speaking test were presented.

There were six tasks used in Study 2. Two were from Study 1 (Tasks 1 and 5) to compare results across Studies 1 and 2. Task 4 in Study 1 was not utilized in Study 2 because it involved planning time. The other four were from test tasks of University of Cambridge Local Examinations Syndicate (UCLES) for the following two reasons. First, the difficulty of tasks can be predicted to some degree without pretests if tasks are derived from tests at the same level although the nature of the tasks may be different from the original test since these tasks were originally used in speaking tests

in which a pair of test takers spoke. Second, similar pictures of similar difficulty and similar task characteristics can be selected in order to examine generalizability (Tasks 3 and 4). Pictures that everyone can understand with general knowledge were selected. Since some of these pictures were on the web, pictures for Tasks 2 and 6 were drawn by hand while all the elements of the pictures were maintained. Pictures for Tasks 3 and 4 did not go through this process because describing the details in the picture seemed to be a fundamental element of the tasks.

As seen in Table 7.7 (see Table 7.8 for the actual tasks), Task 1 was a self-introduction task, which had test takers speak about their school life, family, friends or other relevant topics. Tasks 2 and 5 involved explaining differences between two pictures. Tasks 3 and 4 involved describing a picture. Task 6 asked test takers to

Table 7.7
Tasks in the Speaking Test

Task	Speaking type	Content	Keywords	Source (North's Level)
1	Description	Self-introduction	family [1], friend [1]	Study 1 (Tourist)
2	Picture comparison	Comparing pictures on the left and the right	another [1], open [1]	Movers (Breakthrough)
3	Picture description	Washing dishes	wash [1]	PET (Threshold)
4	Picture description	Riding bicycles	ride [1]	PET (Threshold)
5	Picture comparison	Comparing Taro's rooms before and after	name [1], half [1]	Study 1 (Waystage)
6	Description of differences	Identifying the different picture and giving a reason	fruit [2], person [1], cook [2], dance [1], sing [1], buy [1]	Movers (Breakthrough)

Note. [ ] = Level in the JACET8000; PET = Preliminary English Test. Exact source: Task 2 from UCLES (1999, p. 29); Tasks 3 and 4 from UCLES (2001, p. Ⅲ, Ⅵ); Task 6 from University of Cambridge ESOL Examinations (n.d.-c, p. 30). North's levels: (from easier to more difficult) Tourist, Breakthrough (A1), Waystage (A2), and Threshold (B1). See examples of North's (2000) proficiency level descriptors for Note 5.2. See Koizumi (2005a) for relationships between Tasks 1 and 5 and North's (2000) levels including CEF (Common European Framework) levels. Fourteen keywords included seven nouns (*fruit, person, family, friend, bike,* and *name*), four verbs (*cook, buy, write, ride,* and *wash*), two adjectives (*open* and *half*), and one determiner [quantifier] (*another*).

Table 7.8
Instructions of Tasks in the Speaking Test

Task 1 Please introduce yourself to Ms. Smith. Please talk about your name, family, and friends first. If you do not know what to say, please talk about anything (e.g., your school and favorite things). You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.
Task 2 There are differences between the two pictures. Please explain the differences. Please talk about the marked objects first. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.
Tasks 3 & 4 Describe this picture in as much detail as possible so that Ms. Smith, who is not looking at the picture, can understand what is in it. Please talk about the marked behaviors first. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.
Task 5 Please unfold the folded picture. There are pictures above and below. Your brother (Jiro) is mischievous. While you were away at school, he scattered your belongings in your room. Say how and what in the room has changed by saying "something was something before, but now something is something." If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.
Task 6 One picture out of the four is different. Please tell me what is different and why. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Note. All was originally written in Japanese. See Appendix 7.10.

identify differences between four pictures. Before the test began, test takers answered some easy questions as a warm-up. The difficulty of tasks was predicted as follows, based on UCLES (n.d.-a) and Koizumi (2005a): (from easier to more difficult) (1) Task 1 (Tourist in North's [2000] levels), (2) Tasks 2 and 6 (Breakthrough), (3) Task 5 (Waystage), and (4) Tasks 3 and 4 (Threshold).

There were 14 keywords common to the Productive VKT and the Speaking Test. These words were selected from the picture used in the Speaking Test and from words that belonged to Levels 1 to 3 in the JACET8000 (JACET Basic Word Revision Committee, 2003). Once keywords were decided, they were put in the Productive VKT (see section 7.2.2.1). The same parts of speech were chosen within the same task to avoid confusion among test takers. In Tasks 2, 3, 4, and 5 of the Speaking Test, the pictures for these words were marked (by a circle or a rectangle) and the instruction

said that test takers should describe the marked thing or behavior first (see Appendix 7.10).

The participants were not informed about the content or structure of the test beforehand. They were assured that the results of the Speaking Test were not included in the class evaluation. During the test, instructions were given in Japanese. Although some may say that this procedure is unnatural, it was done for the following two reasons. First, having test takers understand the instructions is essential in conducting tasks and using L1 is one of the most efficient ways of ensuring this. Second, even if the instructions are given in Japanese, the speaking process, starting from concepts and expressing them in English, does not seem very different.

At the end of each task, the following instruction was given in order to elicit unplanned speech: When Ms. Smith says “Please start” please say “Yes” and start speaking right away. Examples of responses to the tasks were not provided due to lack of time and because, if there were examples, test takers were likely to mimic the structures in the examples if possible. In Study 1 and the pilot study of Study 2, examples were not given but utterances could be elicited without difficulty. Test takers were instructed not to pass on the content of the test to anyone.

The time allowed to complete each task was all set as 60 seconds in order to make instructions consistent and thus easier for test takers. Since target learners were beginning learners, more words could be elicited by using more tasks with shorter speaking time for each task. As a result, the two tasks from Study 1 (Tasks 1 and 5) went through minor modifications in terms of time (from speaking time of 90 seconds to 60 seconds) and instructions in order to be consistent with the other tasks and the testing situations.

Of the six tasks, one task (Task 6) was not analyzed in Study 2 because many

test takers misunderstood the instructions.

7.2.2.3 Questionnaires

Two questionnaires were used: the Productive VKT Questionnaire and the Speaking Test Questionnaire (see Appendixes 7.11 and 7.12). They were conducted in order to obtain background information (e.g., experiences of going abroad) and information concerning tests (e.g., the substantive and consequential aspect of validity [i.e., face validity and washback effects]). An option "I don't know" was put among other options based on Cumming, Grant, Mulcahy-Ernt, and Powers (2004) because responses with this option may be more meaningful since some test takers may not be conscious of the test quality and without the "I don't know" option, they may be forced to select other options.

As for the method of determining face validity in the questionnaires, the following phrase was used: Do you think this test measures English vocabulary (or speaking) ability? This phrase was used to ensure understanding by junior high school students and because if the phrase "Do you think this test measures 'your' English vocabulary (or speaking) ability?" were used, some test takers would think that they do not have enough vocabulary (or speaking) ability, so no tests can measure "their" vocabulary (or speaking) ability.

7.2.3 Procedures

7.2.3.1 Test Administration Procedures

Before administering the Productive VKT and Speaking Test, a pilot study was conducted. As for the Productive VKT, 84 students took the pilot test (38 public junior high school students, 42 public senior high school students, and 4 undergraduates at a

national university). As for the Speaking Test, 11 out of the 84 students took the pilot test (8 public junior high school students and 3 undergraduates at a national university; see Appendix 7.1). The instructions and task difficulty were examined and the tests were modified.

The Productive VKT was conducted during English classes or after school (see Appendix 7.1). Forty-five minutes were needed to administer the tests (25 minutes for the Size Section and 20 minutes for the Depth Section). After solving questions in the Size Section, test takers waited until the test administrator asked them to go to the next section. In answering the Depth Section, test takers were not allowed to go back to the previous Size Section. These procedures were taken because some items in one section could be answered using information in the other section. When the time was over, test takers were asked to mark where they had finished in order to obtain information on who did not finish solving questions. After the Productive VKT, the questionnaire was answered when there was time.

The Speaking Test was administered within 15 minutes. As seen in Table 7.9, there were two versions of different orders of tasks. For both versions, Task 1 came first. Then in Version A, Tasks 2 and 3 came next, followed by Tasks 6, 4, and 5. In Version B Tasks 6, 5, and 4 came next, followed by Tasks 2 and 3. This was done for two reasons: (a) to counterbalance order effects and (b) to decrease the chances of cheating by copying other test takers' utterances. Although there were six tasks, counterbalancing all the six tasks was not done because it was too complex. Therefore, two task groups ("Tasks 2 and 3" and "Tasks 4, 5, and 6") were counterbalanced and included tasks of various predicted difficulties (see section 7.2.2.2). In the beginning of the Speaking Test, Task 1 (a self-introduction task) was put first because it is normal to conduct a self-introduction first in the specified testing situations that the Speaking

Test had (e.g., talking to Ms. Smith; see Appendix 7.13).

Table 7.9
Order of Sets in Each Version in the Speaking Test

	First set	Second set	Third set
Version A	Task 1	Tasks 2 and 3	Tasks 6, 4, and 5
Version B	Task 1	Tasks 6, 5, and 4	Tasks 2 and 3

The Speaking Test was administered using either of the following two types of facility: (a) language laboratory facilities (LL Kit) when the schools had them, and (b) the Speaking Test Kit that the researcher brought to school. The Speaking Test using either of the two types was almost the same in terms of the instructions and what test takers did. The first tape (instruction tape) was run including instructions and pauses for test takers to speak. They listened to the tape with headphones and spoke into a microphone. Their utterances were recorded onto the second tape (recording tape). There were two minor differences in testing situations between the LL Kit and the Speaking Test Kit: (a) the time of starting the test and the test version, and (b) distances from other test takers. First, in using the LL Kit, a teacher ran the instruction tape of Version A or Version B, and all the test takers in the room listened to the tape of the same version at the same time and they spoke at the same time. In contrast, in using the Speaking Test Kit, the start button of the instruction tape was pushed by the test administrator for each student after the student put on the headphones and was ready. Therefore, the starting time of the tape and the version of the tape were different for each test taker. Second, test takers using the LL Kit sat at their desks, so the distances from other test takers varied. On the other hand, the test administrator put the Speaking Test Kit as far away as possible so that cheating could not be done easily. There were eight students at most in a room. The distances depended on the size of the

room used but there was no case in which test takers sat right next to each other. In addition, different versions of the Speaking Test were set at adjacent places. Moreover, when there was evidence of a test taker repeating the utterances of others, his or her data was not used for analysis.

7.2.3.2 Analysis Procedures

Analysis procedures were described in Chapters 8 to 12.

7.3 Summary of Chapter 7

This chapter describes the method used in Chapters 7 to 12. Two tests used were (a) the Productive Vocabulary Knowledge Test (Productive VKT) and (b) the Speaking Test. For each test, a questionnaire (the Productive VKT Questionnaire and the Speaking Test Questionnaire) was conducted.

Chapter 8 Study 2A: Scoring Methods of the Productive Vocabulary Knowledge Test

8.1 Purpose and Research Questions

The present chapter (Chapter 8) compares the written version with the oral version of the Productive VKT and also examines the relative qualities between several scoring methods of the written version. This is investigated because Study 2 (2E) examines relationships between size and depth of productive vocabulary knowledge and speaking performance and may need to assess the oral aspect of productive vocabulary knowledge, but a written (paper-and-pencil) format is used since the oral version lacks practicality in administration. Thus, the examination of possible effects of mode differences (i.e., oral vs. written) is necessary.

Therefore, the aims of the present chapter are twofold: (a) to examine relationships between the written version and the oral version of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English, and (b) to investigate which scoring method of the written version of the Productive VKT is better in terms of relationships with the oral version of the Productive VKT. These examinations were done in order to explore a more appropriate scoring method with more validity for the Productive VKT, which is used in the subsequent chapters. Figure 8.1 shows where the present chapter (i.e., Chapter 8 Study 2A: Scoring Methods of the Productive Vocabulary Knowledge Test) is located in the overall research. Two Research Questions were posed in relation to the two aims.

Research Question 1: How large a gap is there between the written version and the oral version of the Productive VKT?

Research Question 2: Which scoring method of the written version of the Productive

VKT has a stronger relationship with the oral version of the Productive VKT?

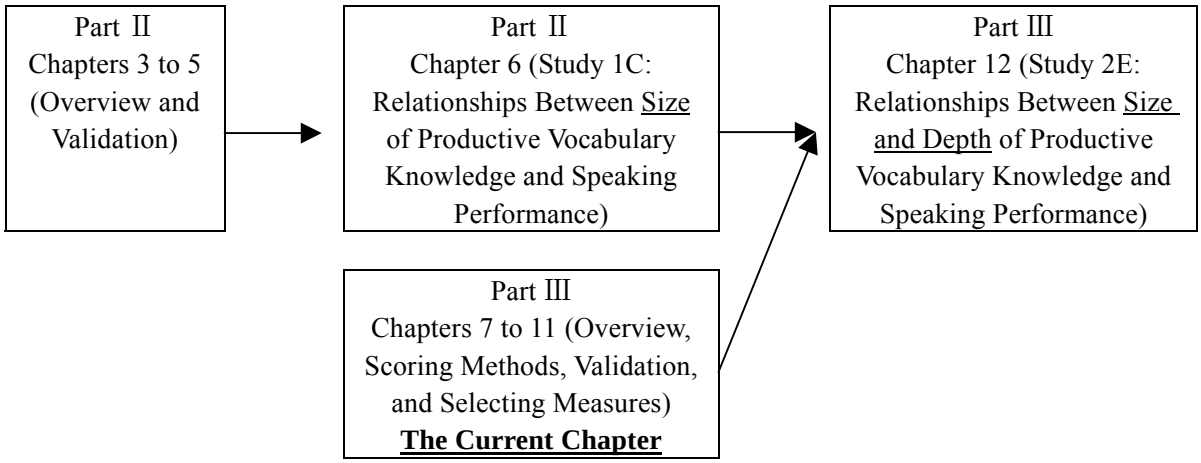


Figure 8.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

The two Research Questions were posed in order to devise a scoring method to take the oral aspect of productive vocabulary knowledge into account for the purpose of reducing possible effects of mode differences (i.e., oral vs. written) on results. Chapter 4 examined the three scoring methods in order to investigate the same point, and it was found that Method 2 provided the most evidence for validity of the interpretation based on test scores. However, the three conditions (see Table 8.1) in the

Table 8.1
Scoring Criteria of the Productive Vocabulary Knowledge Test in Study 1A

Method	Criterion	Responses scored as correct
2	Correct when the spelling is correct OR when there is a misspelling but the following three conditions are met. (a) Only one letter is different from the correct spelling. (b) The pronunciation of a test taker's response read aloud by a rater is almost the same as that of the correct answer when the rater does not know the correct answer. (c) There are no other English words with similar spelling.	*peech; bak*

Note. Both singular and plural forms were scored as correct.

criteria of Method 2 in Study 1 (1A) based on Waring (1997) have the following three problems.

First, one condition is met when (a) only one letter is different from the correct spelling, but Waring (1997) and other research have not examined empirically why a one letter difference is acceptable but a two letter difference is not. Second, how to use the criterion (b) can be variable according to raters, so other rigid criteria may be needed. Third, as for the criterion (c), the degree of strictness varies depending on whether other words with similar spelling exist or not, which might be problematic in interpreting results on the whole consistently. As for the third problem, it is possible to exclude (c) in the criteria but this seems much more problematic because the criteria become too lenient and it is possible for test takers to get a point by writing a response satisfying (a) and (b) when they actually intended to write a different word of similar spelling. This study compared six scoring methods in order to address the first and the second problems.

8.2 Method

8.2.1 Participants

There were 71 participants in Study 2A. They took both the written version and the oral version of the Productive VKT. Originally, 79 students participated but eight were excluded (see Table 8.2 for the reasons). The participants in Study 2 (2B, 2D, and 2E) were basically junior and senior high school students, but Study 2A involved university undergraduates and graduate school students because it was difficult to get participants from senior high school students. It was assumed that their tendency of misspelling was similar to that of senior high school students because they had learned

English in the same educational system and had the same mother tongue, although their proficiency was probably beyond the novice level. All junior high school students were part of those who were analyzed in Chapter 9.

Table 8.2
Participants at Each School

School	School type	School year	*n* who took the oral version	*n* for Study 2A	*n* excluded (Reasons for the exclusion)
JHS D	Public	3rd	39	37	2 (Abroad 1, Name 1)
JHS F	Public	3rd	22	16	6 (Abroad 3, Only oral 1, Name 2)
JHS Total			61	53	
University	National	1st to 4th	15	15	
Graduate School	National	1st	3	3	
Total			79	71	8 (Abroad 4, Only oral 1, Name 3)

Note. JHS = Junior high school; Abroad = A student who had experience of living abroad for one year or more; Name = A student whose name was not detected on the oral version because s/he forgot to say his/her name or his/her voice was too weak; Only oral = A student who took the oral version but not the written version.

8.2.2 Instruments

Two versions of the Productive VKT were used: the written version and the oral version.

8.2.2.1 Written Version of the Productive Vocabulary Knowledge Test

The details were described in section 7.2.2.1. While the whole Productive VKT consisted of the Size and Depth Sections, Study 2A used only the Size Section because there was insufficient time to administer the oral version of the Depth Section and also because assessing depth using the oral version was rather difficult.

Preliminary content analysis of the Size Section showed that there were five words that started with the same letter and had similar meaning, and these were excluded in the subsequent analyses. Thus, in Study 2A, there were 103 items (i.e., 108

– 5) in total.

8.2.2.2 Oral Version of the Productive Vocabulary Knowledge Test

The items in the oral version of the Productive VKT were exactly the same as in the written version. However, there were two minor differences between the written and oral versions. First, in the oral version, when test takers did not know the answer, they said, “Pass” and went on to the next item. They were not allowed to go back to the previous items because they might pick up others’ responses and then go back and repeat the responses that they had just heard. Second, there were four types of the oral version with the same items but with different orders of presenting the items. The first type started with Item 1 (to Item 108); the second version started with Item 25 (through Item 108 to Item 24); the third version started with Item 50 (through Item 108 to Item 49), and the fourth version started with Item 75 (through Item 108 to Item 74). This was done in order to avoid cheating from other students nearby (see Appendixes 8.1 and 8.2). As in the written version, 103 items were used for the analysis.

8.2.3 Procedures

8.2.3.1 Test Administration Procedures

The written version was conducted, as explained in section 7.2.3.1.

The oral version was conducted within 10 minutes. Although this is shorter than the written version of 25 minutes, 10 minutes was the maximum time possible because of time constraints. Test takers took the oral version as follows. First, they listened to the instruction tape from the tape recorder speaker located in the center of the room. Second, they said their names into the microphone located on each desk. Third, they said the item numbers and their responses for each item orally into the microphone. All

the names, item numbers, and responses were tape-recorded. The instruction said test takers did not need to hurry and that it was okay not to finish all the items within 10 minutes. However, most junior high school students were influenced by the speed of the others and maintained a fast speed throughout the test, finishing in under 10 minutes. In contrast, there were three undergraduates who did not finish the test. Because most students maintained about the same level of speed as others and also because they could not go back to the previous items (see section 8.2.2.2), some unintended aspects of speed of lexical retrieval were assessed in the oral version in addition to the intended construct (productive vocabulary knowledge; see section 7.2.2.1.1). Two graduate students who majored in language testing administered the oral version after the training.

Before the test administration, the desks to be used were positioned as far apart as possible to minimize opportunities for cheating, and then the test materials (e.g., the test booklet, tape recorder, and microphone) were placed on the desks. There were seven students at most in one room. The distances from other test takers depended on the size of the room used, but there was no case in which test takers sat right next to each other. In addition, different types of the oral version (see section 8.2.2.2) were given at adjacent places so that neighboring students did not have the same type. Moreover, when there was evidence of a test taker repeating others' utterances, his or her data was not used for analysis, although this situation did not arise in Study 2A. Undergraduates and graduate school students took the test alone in a room with only the test administrator.

At two junior high schools, the written version was first conducted during English lessons. A few weeks later, the oral version was conducted during English lessons or after school. It was considered that there were few order effects because of

the time distance. On the other hand, undergraduates and graduate school students took the two versions within the same day, so the order of the two versions was counterbalanced. Half took the written version first and the rest took the oral version first.

8.2.3.2 Scoring Procedures

The responses in the written version were categorized into three: (a) words with correct spelling, (b) words with misspelling, and (c) no responses. When some judgments were necessary (e.g., a letter could be read as both *a* and *u*), responses were interpreted as the ones closer to the answer. When deciding correct or incorrect spelling, spelling was judged as correct if it occurred in the *Taishukan's Unabridged Genius English-Japanese Dictionary* (Konishi & Minamide, 2002). This dictionary was used because it included many varieties of English spellings (e.g., British, American, Australian, and Canadian English, and eye dialects [i.e., words spelled based on pronunciations]; Konishi & Minamide, 2002, preface). Mistakes of making plurals (e.g., *sockes*) were judged as correct if spellings other than plural forms were correct. Study 2A scored a response as correct only when a present tense form was written because all the prompts in the test used the present tense, as in Study 1 (1A and 1C).

The responses in the oral version were categorized into three: (a) words with correct pronunciation in terms of vowels and consonants, (b) words with mistakes of pronunciation regarding vowels and consonants, and (c) no responses. The criterion of judgment of a mistake in pronunciation was whether native speakers and proficient learners of English could recognize the word when hearing the pronunciation. Since the recording conditions were not very good because of noises outside and the quiet

voices of test takers, the raters scored as correct when they were not very sure. When test takers said more than one word, the last word was used for the judgment.

Whether there was a mistake in pronunciation was judged independently by two raters, who practiced rating, and there was 99.08% agreement of the judgment. When there was disagreement, rationales of judgment were discussed and the agreed judgment was used for the analyses.

After the categorization of responses in the oral and written versions, the scoring of the written version was done using six scoring methods (see Table 8.3 and Appendix 8.3). Only one rater was used because no subjectivity was involved. The six methods can broadly be divided into two types: the criteria of giving a point (a) only when the spelling is correct (Method 1), and (b) when a few conditions are met (Methods 2 and 3). Method 2 was divided into four: Methods 2a, 2b, 2c, and 2d. Methods 2 and 3 were different in two points. First, in Method 2, the number of misspelled letters was a concern, but in Method 3, there were no criteria related to the number of misspelled letters. Second, in Method 2, (b) the correctness of pronunciation knowledge was judged by raters, but in Method 3, the criterion was based on spelling rules in English.

To be strict, a correspondence between the pronunciation and the spelling depends on the context (e.g., a certain spelling occurs only in a certain part of a word). However, the rule used in this study was summarized without thinking about the context because when the context was taken into consideration, correspondence rules between the pronunciation and the spelling were rather limited, and these rules seemed too strict to make scoring criteria for the purpose of the current study.

Table 8.3
Scoring Criteria of the Written Version of the Productive Vocabulary Knowledge Test in Study 2A

Method	Criteria
1	(same as Method 1 in Study 1) Correct only when the spelling is correct.
2a	(same as Method 2 in Study 1): Correct when the spelling is correct OR when there is a misspelling but the following three conditions are met. (a) Only one letter is different from the correct spelling. (b) The pronunciation of a test taker's response read aloud by a rater is almost the same as that of the correct answer when the rater does not know the correct answer. (c) There are no other English words with similar spelling. (e.g., *mouce* for *mouse*)
2b	Same as Method 2a except (a) Only two letters are different from the correct spelling. (e.g., *brrid e* for *bridge*)
2c	Same as Method 2a except (a) Only three letters are different from the correct spelling. (e.g., *o kestra* for *orchestra*; *ple ser* for *pleasure*)
2d	Same as Method 2a except (a) Only four letters are different from the correct spelling. (No responses belonged to this category.)
3	Correct when the spelling is correct OR when there is a misspelling but the following two conditions are met. (b) The misspelling is in line with the English spelling rules. In other words, based on the spelling rules, the pronunciation of a misspelled word can be considered the same as that of the correct spelling (e.g., *mouce, brige*) and/or in misspelled words, silent letters are deleted or added (e.g., *mous, washe*). (c) There are no other English words with similar spelling.

Note. Both singular and plural forms were scored as correct. The pronunciation of the correct spelling was based on Konishi and Minamide (2002). The English spelling rules were summarized based on six books: two books regarding phonics and four books regarding phonology (Gotoh, 1991, pp. 17-36; Ikeura, Izumi, & Itakura, 1990, pp. 20-22; Pennington, 1996, pp. 192-199; Spalding, 1986, pp. iv-vii; Takebayashi, 1988, pp. 200-205; Yasui, 1992, pp. 10-13, 216). The rules of silent letters were based on Pennington (1996, p. 197).

8.2.3.3 Analysis Procedures

The two Research Questions were answered using the results of the categorization of responses in the oral and written versions of the Productive VKT and Pearson's product-moment correlation coefficients (r) between the oral version and the written version scored by six scoring methods. The interpretation of $|r|$ was as follows:

Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1).

8.3 Results and Discussion

The oral version and the written version were both categorized into three types. Table 8.4 shows nine categories combined. For each item, all the test takers were classified into nine categories. For example, as seen in Table 8.5, in Item 2, 33 students wrote a word with correct spelling in the written version and said a word with correct pronunciation in the oral version (as seen in (a)), whereas 31 students wrote a word with misspellings in the written version and said a word with correct pronunciation in the oral version (as seen in (d)).

Table 8.4
Nine Categories of Responses From the Written and Oral Versions

Written \ Oral	Correct pronunciation (oral○)	Pronunciation mistake included (oral△)	No response (oral×)
Correct spelling (written○)	(a)	(b)	(c)
Misspelling included (written△)	(d)	(e)	(f)
No response (written×)	(g)	(h)	(i)

Table 8.5
Number of Test Takers who Belonged to Each Category

No.	Answer	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	Total
1	dog	71	0	0	0	0	0	0	0	0	71
2	mouse (mice)	33	0	0	31	2	0	2	0	2	70
3	sock(s) (sox)	46	0	6	6	2	1	3	0	6	70
	(Others omitted)										
	Total	2735	87	84	548	421	332	154	96	2155	6612
	Proportion	.41	.01	.01	.08	.06	.05	.02	.01	.33	1.00

Note. See Appendixes 8.4 and 8.5 for the whole table.

Among the nine categories, the cases where written and oral responses

converged were a), e), and i). The frequency and proportion in the other six cells indicates diversions between the oral version and the written version. Although it is interesting to examine what items diverged and why, this study focused on the overall tendency by looking at the proportion in total at the bottom of Table 8.5. First, the total proportion of a), e), and i), which means the proportion of the written and oral responses that converged, was .80 (i.e., .41 + .06 + .33; to be strict, .8032), whereas the total proportion of the other six (i.e., the proportion of the written and oral responses that diverged) was .1968. In other words, there was a gap of about 20% (19.68%) between the oral version and the written version.

Next, scoring was done based on the six scoring methods. Then, correlations between total scores of the oral version and the six types of total scores of the written version were examined.

As seen in Table 8.6, there were very high correlations between the total scores of the oral version and the six types of the total scores of the written version (r = .89 to .90). It should be noted that although very high correlations existed, they were not perfect ones. The shared variances between the oral version and the written version were approximately 80% (i.e., 79.21% to 81.00%; calculated from $(.89)^2$ to $(.90)^2$). This is consistent with the result above. Therefore, as an answer to the Research Question 1 (How large a gap is there between the written version and the oral version of the Productive VKT?), there is a gap of about 20% between the two versions. The results suggest that although the 20% gap is not very small, the substantial part (80%) is shared by the two versions. If the criterion of strong generalizability across the two versions is set at r = .50 or above, as was done in Study 1B, the two versions of the Productive VKT have very strong generalizability, which is one piece of evidence of the generalizability aspect of validity of the Productive VKT (see section 2.5). Reasons

for the 20% gap in the constructs that the two versions measure may be that knowledge of written forms and oral forms is inherently slightly different, and that the oral version also assesses access speed of productive vocabulary knowledge (see section 8.2.3.1), which needs to be examined further.

Table 8.6
Correlations Between the Scores on the Oral Version and the Six Types of Scores of the Written Version

	M1	M2a	M2b	M2c	M2d	M3	Oral
Method 1	--	.99**	.99**	.99**	.99**	1.00**	.89**
Method 2a (one letter)		--	1.00**	1.00**	1.00**	1.00**	.90**
Method 2b (two letters)			--	1.00**	1.00**	1.00**	.90**
Method 2c (three letters)				--	1.00**	1.00**	.90**
Method 2d (four letters)					--	1.00**	.90**
Method 3 (spelling rule)						--	.90**
Oral version							--

Note. n = 71. M = Method; Oral = Oral version.
**$p < .01$.

As for the comparisons between the six scoring methods, there were very high correlations between the six types of the total scores of the written version (r = .99 to 1.00). When differences of the relationships between the oral and written versions were examined, the difference was only .01 between Method 1 and the other methods, whereas there were no differences between Methods 2 (2a to 2d) and 3. Although the effect sizes of differences between the correlation coefficients of .99 and 1.00 were very strong (q = 1.15), the difference of .01 were considered to be very small. The similar results of Methods 2a to 2d and of Methods 2 (2a to 2d) and 3 indicate that the scoring criteria of one letter difference or more letter difference, and the scoring based on the English spelling rules or that based on human judgments do not matter very much. As an answer to the Research Question 2 (Which scoring method of the written version of the Productive VKT has a stronger relationship with the oral version of the Productive VKT?), Methods 2 (2a to 2d) and 3 had slightly higher correlations than

Method 1. However, it seems that the similarities are greater than the differences, and all the six methods tend to measure very similar productive vocabulary knowledge and produce similar results. Although there was a small difference between Method 1 and the other methods, this difference can be considered very marginal. Therefore, it is concluded that there is little difference between Method 1 and the other methods, at least in the case of the novice learners targeted in this study, which led to the decision to use Method 1 in the subsequent study and to score words with correct spelling as correct. The advantage of using Method 1 is that the third problem posed in section 8.1 concerning Waring's (1997) method is not related and that scoring can be easier.

8.4 Summary of Chapter 8

The present chapter examines relationships between the written version and the oral version of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English, and (b) investigated which scoring method of the written version of the Productive VKT was better in terms of relationships with the oral version of the Productive VKT. It is shown that there is a gap of about 20% between the two versions. As for the comparisons between the six scoring methods, it is demontrated that there is little difference between Method 1 and the other methods, at least in the case of the novice learners targeted in this study, which led to the decision to use Method 1 in Chapters 9 and 12.

Chapter 9 Study 2B: Validation of the Productive Vocabulary Knowledge Test

9.1 Purpose and Hypotheses[9.1]

The purpose of the present chapter is to examine the validity of interpretation and uses based on the scores of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English, which is used in the main examination of relationships between productive vocabulary knowledge and speaking performance in Chapter 12 (Study 2E). Figure 9.1 shows where the present chapter (i.e., Chapter 9 Study 2B: Validation of the Productive Vocabulary Knowledge Test) is located in the overall research.

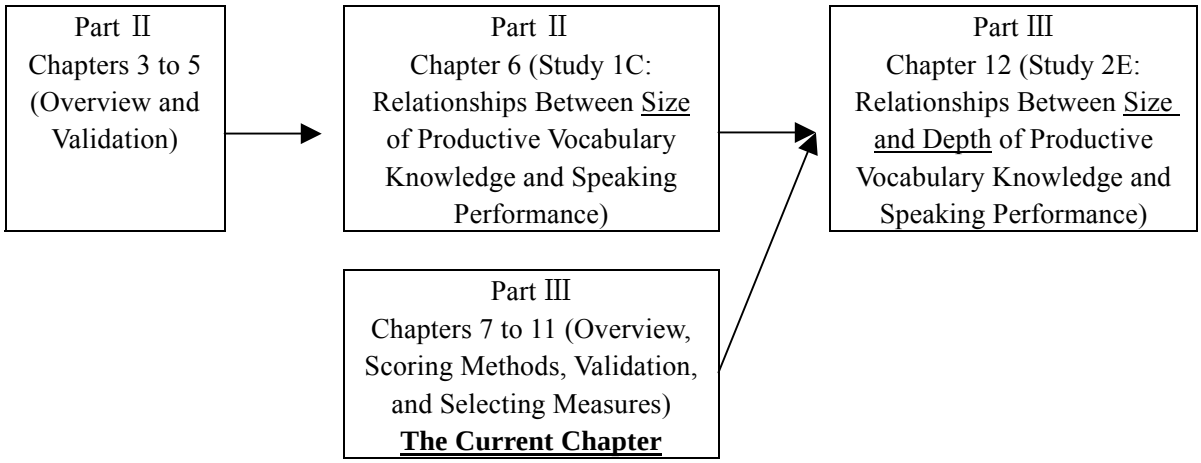


Figure 9.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

One Main Hypothesis was posed in relation to the validity. In order to examine the Main Hypothesis, 14 Hypotheses for Validity were made. Table 9.1 illustrates the organization of the current study (Study 2B).

Main Hypothesis: The Productive VKT developed in this study has positive evidence for validity.

Hypothesis for Validity 1 (the substantive aspect): There is a positive response from test takers concerning the quality of the Productive VKT.

Hypothesis for Validity 2 (the consequential aspect): There is a positive response from test takers concerning the effect of the Productive VKT on learning English.

Hypothesis for Validity 3 (the substantive aspect): Those who do not know the keywords (as measured by the Size Section of the Productive VKT) do not use them in the Speaking Test.

Hypothesis for Validity 4 (the structural aspect): The data in each section/subsection of the Productive VKT fit an item response theory (IRT) model. In other words, the Productive VKT satisfies unidimensionality).

Hypothesis for Validity 5 (the structural aspect): Each section/subsection of the Productive VKT has one factor structure. In other words, the Productive VKT satisfies unidimensionality).

Hypothesis for Validity 6 (the structural aspect): The reliability of items and persons is high in each section/subsection of the Productive VKT.

Hypothesis for Validity 7 (the structural aspect): When response patterns are examined individually in the Size Section of the Productive VKT, the number of correct items is the largest in Level 1, followed by Level 2. The number of correct items is the smallest in Level 3 (e.g., the following example is the case in which a hypothesis is supported: 15 items correct in Level 1 > 8 items correct in Level 2 > 3 items correct in Level 3).

Hypothesis for Validity 8 (the structural aspect): The three depth subsections can be combined and expressed under the construct “depth” (because the three

depth subsections assess the same construct of "depth" although each also assesses unique aspects of depth).

Hypothesis for Validity 9 (the external aspect): Those who have the higher STEP (Society for Testing English Proficiency) Test grade obtain higher scores in each section/subsection of the Productive VKT (because vocabulary knowledge is considered to be part of communicative language ability [e.g., Bachman & Palmer, 1996], which is assessed by the STEP Test [STEP, 2004], and the two are expected to be related to some degree).

Hypothesis for Validity 10 (the external aspect): There are moderate or strong correlations between the Productive VKT section/subsection scores and the CASEC (Computerized Assessment System for English Communication) total scores (because vocabulary knowledge is considered to be part of communicative language ability [e.g., Bachman & Palmer, 1996], which is assessed by the CASEC and the two are expected to be correlated to some degree).

Hypothesis for Validity 11 (the external aspect): Correlations between the Productive VKT section/subsections and the CASEC Section 1, which intends to assess vocabulary knowledge, are higher than those between the Productive VKT section/subsections and the CASEC Sections 2 and 4.

Hypothesis for Validity 12 (the external aspect): Correlations between the Productive VKT section/subsections and the CASEC Section 4, which has a similar test format to the Productive VKT section/subsections, are higher than those between the Productive VKT section/subsections and the CASEC Sections 2 and 4.

Hypothesis for Validity 13 (the content aspect): Each section/subsection of the

Productive VKT has positive evidence for the content aspect.

Hypothesis for Validity 14 (the generalizability aspect): There is a strong correlation between the oral version and the written version of the Size Section of the Productive VKT.

Table 9.1
Organization of Study 2B: Validation of the Productive Vocabulary Knowledge Test

	Validity aspect	Method to be used	Section reported
Main Hypothesis			9.3.4
Hypothesis for Validity 13	Content	Expert professional judgment	9.3.4
Hypotheses for Validity 1 and 3	Substantive	・Questionnaire ・Discourse analysis by focusing on the language produced by a test taker	9.3.1
Hypotheses for Validity 4 to 8	Structural	・Classical true-score reliability methods and item response theory (IRT) methods ・Factor analysis, & structural equation modeling (SEM)	9.3.2
Hypothesis for Validity 14	Generalizability	Correlations in showing consistency of performance across modes	9.3.4
Hypotheses for Validity 9 to 12	External	Correlations ANOVA	9.3.3
Hypothesis for Validity 2	Consequential	Questionnaire	9.3.1

The Hypothesis for Validity 7 was set because Level 1 items were the highest frequency words out of the three levels and thus Level 1 may be the easiest overall. The Hypotheses for Validity 10 to 12 were posed based on the principle of internal construction validation (Henning, 1987) that correlations between similar aspects are stronger than those between different aspects.

Although examination of the same Main Hypothesis was conducted in Study 1A, this study (Study 2B) investigated them again for the following two reasons. First, in

addition to the Size Section, the Depth Section was made. Second, the validation of interpretation and uses based on scores and values is always needed, especially when the range of target test takers varies, which is the case in Study 2B.

9.2 Method

9.2.1 Participants

The participants in this study were 913 students studying English in Japan. This group contained 330 (both public and national) junior high school students and 583 (both public and national) senior high school students, including 481 males (52.68%), 428 females (46.88%), and 4 students who did not write their sex on the questionnaires (0.44%; see Table 9.2). Their first language (L1) was Japanese. They were aged from 14 to 18 and all had studied English as a foreign language for approximately two to five years. Therefore, learners of a wide range of English ability participated. In fact, the range in Study 2 (2B) was much wider than that in Study 1 (1A). They were considered to be novice Japanese learners of English in this research. Originally, 1066 students took the Productive VKT, but 153 students were not used for analysis (see Table 9.2 *Note*. for the reasons).

Table 9.2
Participants at Each School

School	School type	School year	*n* who took the Productive VKT	*n* for Study 2B	*n* excluded (Reasons for the exclusion)
JHS A	Public	3rd	20	14	6 (Abroad 2, Blank 2, Misunderstanding 1, Foreign 1)
JHS B	Public	3rd	73	58	15 (Abroad 7, Blank 4, Misunderstanding 3, Foreign 1)
JHS C	Public	3rd	34	26	8 (Abroad 3, Blank 1, Misunderstanding 4)
JHS D	Public	3rd	35	26	9 (Blank 1, Misunderstanding 8)
JHS E	Public	3rd	39	33	6 (Abroad 1, Blank 4, Misunderstanding 1)
JHS F	Public	3rd	39	36	3 (Abroad 3)
JHS G	National	3rd	160	137	23 (Abroad 16, Blank 2, Misunderstanding 4, Native 1)
JHS Total			400	330	70 (Abroad 32, Blank 14, Misunderstanding 21, Others 3)
SHS A	Public	3rd	36	32	4 (Blank 2, Misunderstanding 1, Not finished 1)
SHS B	Public	1st to 3rd	155	136	19 (Blank 7, Misunderstanding 11, Not finished 1)
SHS C	Public	1st	38	36	2 (Misunderstanding 2)
SHS D	Public	1st	57	57	0
SHS E	Public	2nd to 3rd	92	84	8 (Abroad 1, Blank 3, Misunderstanding 4)
SHS F	Public	2nd to 3rd	56	46	10 (Abroad 2, Blank 2, Misunderstanding 3, Twice 3)
SHS G	National	1st	79	61	18 (Abroad 15, Misunderstanding 3)
SHS H	National	2nd	153	131	22 (Abroad 21, Foreign 1)
SHS Total			667	583	83 (Abroad 39, Blank 14, Misunderstanding 24, Others 6)
Total			1066	913	153 (Abroad 71, Blank 28, Misunderstanding 45, Others 9)

Note. JHS = Junior high school; SHS = Senior high school. Abroad = A student who had the experience of being abroad for one year or more; Blank = There was a complete blank in one section or subsection; Misunderstanding = A student who misunderstood the instruction. Foreign = A student from a foreign country. Twice = A student who took the test twice (When a student took the test twice, the responses in the first test that they took were used.). Not finished = A student who did not finish the test within the time. Native = A student whose mother was a native speaker of English. The third-year junior high school students were in the ninth-grade, whereas the first-year senior high school students were in the tenth-grade.

9.2.2 Instruments

The two tests and one questionnaire were used in this chapter: (a) the Productive Vocabulary Knowledge Test (Productive VKT), (b) the Speaking Test, and (c) the Productive VKT Questionnaire. The Productive VKT was the main focus in this study

(Study 2B), whereas the Speaking Test and the Productive VKT Questionnaire were used for validation of the Productive VKT. The Productive VKT consisted of two sections: the Size Section and the Depth Section. The Depth Section had three subsections: the Derivation Subsection, Antonym Subsection, and Collocation Subsection. The details were described in sections 7.2.2.1 to 7.2.2.3.

Preliminary content analysis of the Size Section showed that there were 12 words that started with the same letter and had similar meaning, and these were excluded in the subsequent analyses. In order to make the number of words in one level the same, additional six items were excluded (see Appendix 7.6 for excluded items). In the end, there were 26 items for one level and 78 items (i.e., 26 items * 3 levels) in total.

9.2.3 Procedures

9.2.3.1 Test Administration Procedures

The details were described in section 7.2.3.1.

9.2.3.2 Scoring and Coding Procedures

9.2.3.2.1 Scoring Procedures of the Productive Vocabulary Knowledge Test

In scoring the Productive VKT, All items in the Size Section were scored as either correct or incorrect. Only words with correct spelling received a point, based on the results of Study 2A (see section 8.3). Spelling was judged as correct if it was listed in the *Taishukan's Unabridged Genius English-Japanese Dictionary* (Konishi & Minamide, 2002). For example, *frend/friend*, *socks/sox*, and *work/wark* were scored as correct (see Appendix 7.2 for all the correct answers). Mistakes of making plurals (e.g., *sockes*) were judged as correct if spellings other than plural forms were correct. Study

2 (2B) scored a response (from test takers) as correct only when a present tense form was written. Vocabulary size was estimated using the following formula: [the number of correct words]/78*3,000, based on Laufer (1992).

With regards to the Depth Section, the same principle as that used in the Size Section of giving a point only to correct spelled words was maintained. The following six steps were taken based on Shimamoto (2005). First, seven dictionaries were consulted in order to develop scoring criteria for the three subsections (i.e., Derivation, Antonym, and Collocation Subsections): (a) the *Taishukan's Unabridged Genius English-Japanese Dictionary* (Konishi & Minamide, 2002), (b) the *Random House Roget's College Thesaurus* (Pearsons, Braham, Stein, & Flexner, 2000), (c) the *Webster's New World Dictionary & Thesaurus* (1998), (d) the *Longman Dictionary of Contemporary English* (4th ed.; Summers et al., 2003), (e) the *Oxford Advanced Learner's Dictionary* (Oxford University Press, 2004), (f) the *Cambridge Dictionary of American English* (Cambridge University Press, 2004), and (g) *The Kenkyusha Dictionary of English Collocations* (Ichikawa, Dutcher, Boyd, & Sawamura, 1995). Specifically, (b) and (c) were used for the criteria of antonyms, and (d) to (g) were utilized for the criteria of collocations. The criteria included answers for all possible parts of speech and meanings.

Second, all the responses were classified into two types based on the criteria: words on the criteria (scored as correct) and words not on the criteria.

Third, the words not covered by the criteria were examined if they existed in an entry of Konishi and Minamide (2002). When there was no entry, the responses were considered incorrect.

Fourth, when there was an entry, in scoring the Derivation Subsection, only words that matched the part of speech and meaning specified in the items were

considered correct. Mistakes of making plurals (e.g., *sockes*) were judged as correct if spellings other than plural forms were correct. These rules were also applied to the Antonym and Collocation Subsections. In scoring the Antonym and Collocation Subsections, although one word was required to be written in the test, some test takers wrote two words as a response. In that case, when at least one word was appropriate to the instruction (e.g., nouns) and could be scored as correct, the responses were judged as correct. For instance, *shut down* as the response of the antonym of *open* was scored as correct, whereas *flied egg* as the response of the collocation of *cook* was scored as correct. When written responses were not grammatically connected, they were judged strictly (e.g., *my father birthday* was considered incorrect).

Fifth, the responses were excluded that were completely wrong. Then, the responses were put into a list. In summary, in the list, there were responses that met the following conditions: (a) responses that appeared in Konishi and Minamide (2002) and (b) that were not completely wrong. The words on the list were judged by three raters, who were two native speakers of English and one Japanese advanced learner of English. They rated the appropriateness of the words without considering the context in which antonyms and collocations were used, and they scored the responses as incorrect when the responses were wrong, unacceptable, redundant, unusual, unlikely, or weird.[9.2] Responses acceptable only in a limited context (e.g., acceptable only in a fairy tale or in a science-fiction story) were scored as incorrect. The raters were allowed to consult their dictionaries if needed. The agreement and internal consistency (α) of the three raters were not very high but moderate (In antonyms, the proportion that the three raters agreed on was 76.16%, α = .65. In collocations, the proportion was 52.38%, α = .63.). One reason for the moderate agreement may be that the three raters judged only words that were retained after excluding (a) all the words on the scoring

criteria made in the first step, and (b) the completely mistaken words. Another reason may be that the scoring of collocations, which had a lower agreement ratio (52.38%), could be affected by raters' experiences and cultural background.

Sixth, the items about which there were disagreements among the three raters were evaluated by two additional raters: one native speaker of English and one Japanese advanced learner of English. When three out of the five raters agreed that the students' responses were either appropriate or inappropriate as antonyms or collocations, they were scored as correct.

The reasons for using five raters instead of having the three raters discuss until their opinions converged were that judgment of depth may be affected by linguistic background history so it was judged that a more diverse range of views was needed. Furthermore, both native speakers of English and Japanese advanced learners of English were used as raters because the perspectives of world Englishes (e.g., Bolton, 2004) and expert users (McCarthy & Carter, 2001, p. 71) were considered important. The expert users included native speakers and proficient learners in the current study, and the expert users of English were used as raters (see Appendixes 9.1 to 9.4).

9.2.3.2.2 Coding Procedures of the Speaking Test

Utterances from the Speaking Test were transcribed. Although each task had 60 seconds of speaking time, only 45 seconds of speech were transcribed. The speaking time was measured from the time when the speaker started speaking (excluding time to produce dysfluency markers at the beginning of the speech) until 45 seconds had passed. The time the instruction finished was not counted as the beginning of the measuring period because the instruction did not appear in many recording tapes. The transcription was done with a policy of listening to the talk sympathetically by taking

the context into consideration. In other words, pronunciation errors were not strictly reflected in transcripts. After the utterances were transcribed and 45 seconds of time was measured by the author and five undergraduates, the author checked the transcripts and the speaking time a second time.

The transcripts were then searched for the keywords that appeared both in the Productive VKT and Speaking Test. The test takers were categorized into three types: (a) those who uttered the keyword in a certain task, (b) those who did not use the keyword but did say something related, and (c) those who did not use the keyword or did not say anything related. Although there were originally 14 keywords, only eight keywords were analyzed because Task 6, which included six keywords, was not used in this study.

9.2.3.3 Analysis Procedures

The Main Hypothesis was tested using the results of 14 Hypotheses for Validity. The Hypotheses for Validity 1 to 3 were examined through questionnaire items and the language that test takers produced. The Hypotheses for Validity 4 to 8 were analyzed using the Rasch multi-faceted measurement in item response theory (IRT), factor analysis (FA), structural equation modeling (SEM; see sections 2.6.4.1 and 2.6.4.2), and analyses between the vocabulary levels. The Hypotheses for Validity 9 to 12 and 14 were examined using Pearson's product-moment correlation coefficients (r) and analysis of variance (ANOVA).[9.3] The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1). When correlation coefficients were compared and one coefficient was higher than the other in a predicted direction, hypotheses were considered to be partially met. When the effect sizes (q) of differences between the correlation

coefficients were medium (q = .30 to .49) or strong (q = .50 or more; Cohen, 1988; see section 2.6.4.1), the hypotheses were considered to be confirmed. When the effect sizes (η^2) were medium or strong, hypotheses was considered to be confirmed (the criteria of η^2: Lower than small < .0099; .0099 ≤ small < .0588; .0588 ≤ medium < .1379; .1379 ≤ large; Cohen, 1988; see section 2.6.4.1). The Hypothesis for Validity 13 was analyzed based on a professional judgment. The Main Hypothesis was tested using results of 14 Hypotheses for Validity.

For the analyses, ability estimates derived from the IRT analysis were used only in section 9.3.2 where IRT was used. In other relevant sections, vocabulary size estimates and raw scores were used because there were some students with zero or full scores and because standard errors of measurement of their ability estimates were larger than the other test takers in IRT.

9.3 Results and Discussion

9.3.1 Investigating the Hypotheses for Validity 1 to 3

Two types of analysis were conducted in this section using the Productive VKT Questionnaire and the language analysis. First, Table 9.3 shows test takers' reactions toward the Productive VKT after removal of the results of test takers who chose the option "I don't know" in the questionnaire (see Table 9.4 for the three items).

Table 9.3
Test Takers' Responses Concerning the Productive Vocabulary Knowledge Test

	Q2 (n = 775)	Q3 (n = 824)	Q4 (n = 833)
M (*SD*)	3.85 (1.07)	3.64 (1.17)	3.55 (1.19)

Table 9.4
Items on the Productive Vocabulary Knowledge Test Questionnaire

Part 2. Please choose the number that describes your feelings the most with regard to the test you have just taken.
---
0. I don't know. 1. No, I don't think so at all.
2. No, I don't think so very much. 3. Neither yes nor no.
4. Yes, I think so a little. 5. Yes, I think so very much.

Q2 (2-2) Do you think that this test measures English vocabulary ability?
Q3 (2-3) Was the test instruction easy to understand?
Q4 (2-4) Do you want to study English more from now on?

First, regarding questions about face validity (Q2), clarity of instructions (Q3), and effects of the Productive VKT on future English learning (Q4), means were above 3.5 on the 1 to 5 Likert scale, with 5 indicating a positive reaction. This suggests that test takers perceived the test as good to some degree in terms of face validity and potential washback effects. Therefore, the Hypotheses for Validity 1 and 2 (There is a positive response from test takers concerning the quality of the Productive VKT and the effect of the Productive VKT on learning.) were confirmed.

Second, comparisons between the responses in the Size Section of the Productive VKT and in the Speaking Test were made. Students were classified into six types, as can be seen in Table 9.5. For instance of *friend(s)*, 168 test takers wrote the correct answer *friend(s)* in the Size Section and used the word in the Speaking Test, whereas 3 test takers did not write the correct answer in the Size Section but used the word in the Speaking Test.

The analysis in this section focused on the underlined cells where test takers did not write the keyword in the Size Section but did use the word in the Speaking Test, which is contrary to the hypothesis. Some diversions were seen for each keyword, but the proportion was not large (i.e., 7% or less). In other words, most test takers satisfied the hypothesis. Since this result may have been caused by assessing

Table 9.5
Comparisons Between Responses in the Size Section of the Productive Vocabulary Knowledge Test and in the Speaking Test

Keyword	Speaking Test / Productive VKT	Test takers who used the keyword	Test takers who did not use the keyword but did say something related	Test takers who did not use keyword or did not say anything related
friend(s)	○	168 (.69)	2 (.01)	67 (.27)
	×	3 (.01)	0 (.00)	4 (.02)
family	○	151 (.62)	63 (.26)	14 (.06)
	×	11 (.05)	4 (.02)	1 (.00)
open	○	215 (.88)	4 (.02)	17 (.07)
	×	7 (.03)	0 (.00)	1 (.00)
another	○	9 (.04)	75 (.31)	106 (.43)
	×	1 (.00)	19 (.08)	34 (.14)
wash	○	226 (.93)	2 (.01)	0 (.00)
	×	16 (.07)	0 (.00)	0 (.00)
ride	○	121 (.50)	85 (.35)	1 (.00)
	×	8 (.03)	27 (.11)	2 (.01)
half	○	86 (.35)	75 (.31)	30 (.12)
	×	7 (.03)	38 (.16)	8 (.03)
write	○	67 (.27)	107 (.44)	56 (.23)
	×	4 (.02)	4 (.02)	6 (.02)

Note. n = 244. ( ) = Proportion of test takers who belonged to this category. Underlined = The focus of the analysis.

productive vocabulary knowledge using the written version, not the oral version, mistakes in the Size Section, if any, were further classified, as shown in Table 9.6. Although the underlined proportion of *open* decreased slightly (from 3% to 2%), the trend was the same. Therefore, to be strict, the Hypothesis for Validity 3, which predicts that those who do not know the keywords (as measured by the Size Section of the Productive VKT) do not use them in the Speaking Test, was only partially supported. However, as Table 9.7 shows, the following possible reasons for diversions that are not very much related to the validity of the Size Section can be considered, and they may explain some diversions that occurred. Further research is needed to detect reasons for the diversions, probably by interviewing test takers. It may be necessary to explore the areas for modification of the Size Section if there are problems.

Table 9.6
Further Analysis of Comparisons Between Responses in the Size Section of the Productive Vocabulary Knowledge Test and in the Speaking Test

Keyword	Speaking Test / Productive VKT	Test takers who used the keyword	Test takers who did not use the keyword but did say something related	Test takers who did not use keyword or did not say anything related
open	○	215 (.88)	4 (.02)	17 (.07)
	opening	3 (.01)	0 (.00)	0 (.00)
	×	4 (.02)	0 (.00)	1 (.00)
half	○	86 (.35)	75 (.31)	30 (.12)
	hurf, herf	0 (.00)	2 (.01)	0 (.00)
	×	7 (.03)	36 (.15)	8 (.03)

Table 9.7
Possible Reasons for Diversions of the Productive Vocabulary Knowledge Test and the Speaking Tests

(1) Problems related to the Size Section of the Productive VKT

・Time constraint: Although test takers knew the word, there was not enough time to answer the question in the Size Section. (This is unlikely because the participants who did not finish the test [i.e., who had reached only the middle of the section and put a mark indicating where they had finished] were excluded in the analysis.)

・Difference in meanings presented: Since test takers did not know the meaning in the Size Section, they were unable to write the word. However, they knew the meaning elicited in the Speaking Test and used the word successfully. (This is unlikely because the meaning in the two tests was the same.)

・Effect of the first letter: The first letter was presented in the Size Section and the information on the first letter confused some test takers when they attempted to retrieve the word (see section 3.2.2.1). In contrast, the way the word was elicited in the Speaking Test was not confusing, so test takers could use the word in speaking performance. (This is a possible explanation of diversions.)

(2) Problems related to test administration

・Order effect or time interval of the two tests: The students who took the Productive VKT before the Speaking Test did not remember the word in the Productive VKT, but later they learned or remembered the word and used it in the Speaking Test. (This is a possible explanation of diversions.)

・Difference in the degree of pressure: In the Productive VKT, which was a paper-and-pencil test, test takers did not write the word because they did not have enough confidence in their response or knowledge. However, in the Speaking Test, they felt the pressure to speak something and said the word even though they did not have enough confidence. (This is a possible explanation of diversions.)

(3) Problems related to test takers' knowledge structures

・Difference in pronunciation and spelling: Test takers knew the pronunciation already, which enabled them to use the word in the Speaking Test, but they had not acquired the spelling yet, which prevented them from writing the word on the Productive VKT. (This is a possible explanation of diversions.)

・Difference in confidence: Test takers could pronounce the word with confidence, but they did not have confidence in spelling. (This is a possible explanation of diversions.)

・Difference in clues to retrieve words: They did not remember words from the meaning or the letters used as a prompt in the Productive VKT, but they could recall the word through pictures or situations. (This is a possible explanation of diversions, but for the word *friend* and *family* in Task 1, letters were the prompt in both tests, so, as far as the two words go, this is unlikely.)

9.3.2 Investigating the Hypotheses for Validity 4 to 8

The Size and Depth Sections were examined from the following four perspectives: (a) test unidimensionality (using misfit analysis of IRT and factor analysis), (b) reliability, (c) relationships between the levels of the JACET8000 (JACET Basic Words Revision Committee, 2003) regarding the Size Section, and (d) SEM.

First, test unidimensionality was investigated using misfit analysis of IRT and factor analysis. After the initial calibration of the Size Section, it was found that misfitting students (n = 22, 2.41%) went beyond 2% (see Table 9.8). Out of the misfitting test takers, three had one or more blank pages in the Size Section, which suggests that they happened to skip pages. Thus, they were excluded from analysis and 910 (i.e., 913 - 3) test takers were analyzed.

Table 9.8
Misfit Criteria, Percentages of Misfit, and Reliability of the Productive Vocabulary Knowledge Test

		Infit Mean Square *M* (*SD*) [Misfit Criteria]	Percent of misfit (The number beyond the criteria/Total number)	Reliability
Size first (n = 913)	Test takers	1.0 (0.3) [1.6]	2.41% (22/913)	.96
	Items	1.0 (0.2) [1.4]	0.00% (0/84)	1.00
Size second (n = 910)	Test takers	1.0 (0.3) [1.6]	2.53% (23/910)	.95
	Items	1.0 (0.2) [1.4]	0.00% (0/78)	1.00
Derivation (n = 910)	Test takers	1.0 (0.4) [1.8]	3.63% (33/910)	.87
	Items	1.0 (0.2) [1.4]	5.00% (1/20)	1.00
Antonym (n = 910)	Test takers	1.0 (0.4) [1.8]	3.63% (33/910)	.81
	Items	1.0 (0.2) [1.4]	0.00% (0/20)	1.00
Collocation (n = 910)	Test takers	1.0 (0.5) [2.0]	3.63% (33/910)	.78
	Items	1.0 (0.1) [1.2]	0.00% (0/18)	1.00

Note. Test takers = Test takers' ability; Item = Item difficulty.

The second Rasch analysis of the Size Section revealed that no items (0%) were misfitting and that the percentage of them was less than 10% (Stansfield & Kenyon,

1995; see section 2.6.4.2).

After the initial calibration of the Derivation Subsection, it was found that one item (Item 5) was misfitting (5%), and the percentage of the misfitting item was less than 10%. The value of misfit was only just beyond the criterion (Infit Mean Square = 1.41), which led to the decision to retain the item in the analysis.

After the initial calibration of the Antonym and Collocation Subsections, it was found that no items (0%) were misfitting and that the percentage of them was less than 10%.

Since the criterion of less than 10% was met, the fit of the data to the model (i.e., unidimensionality) was considered to be satisfied. The Hypothesis for Validity 4 regarding unidimensionality was supported, which leads to the structural aspect of validity evidence of the Size Section and Depth Section of the Productive VKT. Although the proportion of misfitting persons was slightly higher than 2 % (McNamara, 1996; see section 2.6.4.2), the difference was marginal. In addition, examination of misfitting students did not reveal any particular problems with test items. Thus, they were not excluded and all test takers (n = 910) were used in the analysis. One possible reason for some students having misfit values was that they guessed the answers. For example, some might have added or deleted derivational morphemes from the stimulus words randomly in the Derivation Subsection.

Next, factor analysis was conducted. Since all the items of the Productive VKT were scored as either correct or incorrect, NOHARM87 (Fraser, 1988), which specializes in doing factor analysis of such data, was used. The characteristics of this software are that both exploratory and confirmatory factor analyses are possible, and that fit indices (i.e., root mean square of residuals [RMR]) can be computed. According to Fraser (1988), “If the root mean square residual is in the order of the

typical standard error of the residuals (4 times the reciprocal of the square root of the sample size) we have a rough indication that a refined test of significance would not reject the hypothesized model." In other words, if RMR is less than 0.1325 (i.e., 4*(1/√910)), it can be considered that the data fit the model. The current study used confirmatory factor analysis and set the number of factors as one.

Table 9.9 shows that all the values of RMR were far below 0.1325 and that the data fit the one factor model. Therefore, the Hypothesis for Validity 5 of unidimensionality (Each section/subsection of the Productive VKT has one factor structure) was supported.

Table 9.9
Root Mean Square of Residuals (RMR) Derived from Factor Analysis:

Size (78 items)	Derivation (20 items)	Antonym (17 items)	Collocation (18 items)
0.006	0.008	0.006	0.007

Note. n = 910.

Second, two types of reliability of each section/subsection were examined: (a) Rasch item reliability (equivalent to the KR-20 or Cronbach's alpha; Linacre, 1991, p. 86) and (b) Rasch person reliability. The reliability of items and persons, as shown in Table 9.8, was found to be high for the Size Section and Depth Section, which show that all the section/subsections of the Productive VKT provide stable and consistent scores. Therefore, the Hypothesis for Validity 6 (The reliability of items and persons is high in the Productive VKT.) was supported.

Third, relationships between the levels of the JACET8000 in the Size Section were examined based on the Hypothesis for Validity 7 (When response patterns are examined individually in the Size Section of the Productive VKT, the number of correct items is the largest in Level 1, followed by Level 2. The number of correct

items is the smallest in Level 3). As a result, most test takers matched this hypothesis (93.96%; 855/910), but there were 55 students (i.e., 910 – 855), ranging from junior to senior high school students, who did not fit the prediction. Thus, the Hypothesis for Validity 7 was partially supported. Out of the 55, 46 students (83.64%; 46/55) belonged to a senior high school affiliated with a national university and the teacher said that students (second-year students) had taken vocabulary tests of rather high difficulty in every class. There were 10 common words between the items in the Size Section and the vocabulary tests conducted in class and the diversion pattern from the predicted pattern of 45 students (97.83%; 45/46) was explained by these ten words.

Fourth, SEM was used to examine the internal structure of the Depth Section. The Hypothesis for Validity 8, which posits that the three depth subsections can be combined and expressed under the construct of "depth," was examined using two methods: (a) fit statistics, which show the degree to which the model fit the data, and (b) path coefficients from the construct of depth to the three depth subsections.

In the initial analysis, two assumptions about using SEM were examined: univariate normality and multivariate normality (Kunnan, 1998, p. 313). As for univariate normality, skewness and kurtosis of all variables were within the criteria (|±2|; Kunnan, p. 313; see Table 9.10).

Table 9.10
Descriptive Statistics of the Productive Vocabulary Knowledge Test

	M	*SD*	Minimum	Maximum	Skewness	Kurtosis	α
Size (78 items)	1152.83	530.21	38	2962	0.79	0.15	.95
Derivation (20 items)	8.43	4.84	0	20	0.43	-0.46	.88
Antonym (17 items)	6.73	3.55	0	17	0.45	-0.38	.82
Collocation (18 items)	10.69	3.35	0	18	-0.33	0.16	.79

Note. n = 910. α = Cronbach's alpha. Size = [the number of correct items]/78*3,000. Others are raw scores. See Appendix 9.6 and 9.7.

Next, Mahalanobis Distance was utilized in order to detect multivariate outliers,

which are related to multivariate normality. One student's data (students with χ^2 = 13.82 or more, df = 2, p < .001; Tabachnick & Fidell, 2001) was deleted. Before removal, multivariate normality values in the Amos output were 1.20 (the critical ratio of 2.61, n = 910), which was high enough to be significant at 5%, showing multivariate non-normality. However, after deletion, the value decreased to 0.68, which was not significant. This data of 909 students (i.e., 910 - 1) was used for the subsequent analysis. The excluded student gained a high score on the Size Section (the estimated vocabulary size of 2,231 words) but the scores in the three depth subsections were moderate (8, 6, 12 points for each subsection). This suggests that the results of this study cannot be generalized to such learners.

In order to make a model explaining relationships between size and depth of productive vocabulary knowledge and between the depth construct and the depth subsection scores, one path model was drawn. In the path model of this analysis (Figure 9.2), the measured (observed) variables (i.e., rectangles), which can be assessed directly, were raw scores of the three depth subsections and the size estimates derived from the formula using the raw scores (i.e., [the number of correct items]/78*3,000). The latent variable (i.e., an oval) was "depth," which cannot be directly assessed and is equal to factors in factor analysis. For model estimation, a maximum likelihood method was used. The results of fit statistics (see Table 9.11) suggest that the three fit statistics (i.e., χ^2, χ^2/df, and RMSEA) did not satisfy the criteria, but that the other two (i.e., CFI and GFI) did satisfy the criteria. It was concluded that this model fit the data because χ^2 is problematic since "with large samples, trivial differences between sample and estimated population covariance matrices are often significant," and "with small samples, the computed χ^2, may not be distributed as χ^2, leading to inaccurate probability levels" (Tabachnick & Fidell, 2001,

p. 698) and because χ^2/df is affected by the characteristics of χ^2. All the coefficients of the regression weights and correlations were significant.

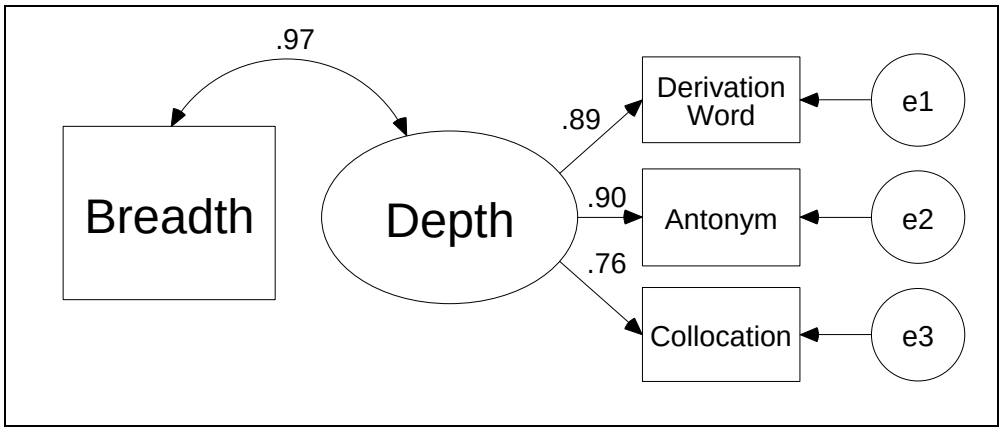


Figure 9.2. Relationship Between Size and Depth Using Structural Equation Modeling (Standardized Solution; n = 909).

Table 9.11
Fit Statistics for Model in Figure 9.2

	χ^2 (*df*), *p*	χ^2/df	CFI	GFI	RMSEA (90%CI)
Criteria	*p* > .05	< 2.0	> .90	> .90	< 0.05
Model 1	8.84 (2), .01	4.42	1.00	1.00	0.06 (0.02 to 0.11)

Note. n = 910. CFI = Comparative Fit Index; GFI = Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval. See Arbuckle & Wothke (1995) for the criteria.

Figure 9.2 shows that all the three path coefficients from the construct of depth to the three depth subsections were high, so the three depth subsections loaded highly on the depth factor. This demonstrates that the Hypothesis for Validity 8 (The three depth subsections can be combined and expressed under the construct of "depth.") was supported. One reason for having high path coefficients may be that the depth subsections assess not only vocabulary depth but also knowledge of parts of speech or cognitive ability to understand the instruction. This possibility needs to be tested in future studies.

There was a strong correlation between size and depth (.97), which indicates that,

overall, those who have wider size of productive vocabulary knowledge tend to have wider depth of productive vocabulary knowledge. This may be because when size increases, depth widens, which leads to wider vocabulary size (Mochizuki & Aizawa, 2000). It should be noted that there were simple correlations of r = .73 to .88 (see Table 9.12) between the Size Section and the Depth Section when SEM was not used, so the proportion for which the Size Section can predict the depth subsections is only 53.29 to 77.44%. The reason why simple correlation coefficients were lower than SEM results was that SEM removed errors in examination of the relationships.

Table 9.12
Correlations Between Size and Depth (Derivations, Antonyms, and Collocations)

	Derivation	Antonym	Collocation
Size	.86**	.88**	.73**
Derivation	--	.79**	.69**
Antonym		--	.70**

Note. n = 909.
**p < .01.

9.3.3 Investigating the Hypotheses for Validity 9 to 12

In this section, associations were examined between the section/subsection scores and results of other two tests: (a) the STEP Test grades (STEP, 2004) and (b) the CASEC (Computerized Assessment System for English Communication) scores. The information on test takers' STEP Test grades was obtained from the Speaking Test Questionnaire (see Table 9.13). In other words, it was a self-reported grade. In contrast, the CASEC scores were provided by a teacher at Junior High School G (see Table 9.2

Table 9.13
Item on the Speaking Test Questionnaire

Part 1. Please write your response in [ ].
(Q1-3) Do you have any English certificates? If so, please give the grade obtained below. Multiple responses are allowed.
STEP Test [ ] grade; Cambridge Test [ ]; Others [ ]

for the code).

First, in the Hypothesis for Validity 9, with regard to the STEP Test grades, it was hypothesized that those who had the higher STEP Test grade would obtain higher scores of the Productive VKT. As seen in Table 9.14, the effect sizes of grade differences were large (η_G^2 = .24 to .40) for each section/subsection, so the Hypothesis for Validity 9 was supported.

Table 9.14
Means, Standard Deviations, and One-Way Analyses of Variance (ANOVA) for Effects of the STEP Test Grade on the Size and Depth Sections

	Grade	Second	Pre-second	Third	Fourth	Fifth
	n	49	164	274	82	20
Size (78 items)	*M*	1985.10	1365.15	1002.24	864.89	898.10
	SD	427.03	418.35	325.10	376.23	519.77
	F (4, 120.41) = 75.86** [2 > pre2 > 3 = 4 = 5]; η_G^2 = .40					
Derivation (20 items)	*M*	14.63	10.29	7.27	5.70	5.70
	SD	3.73	3.96	3.63	3.63	4.52
	F (4, 584) = 63.16** [2 > pre2 > 3, 4, 5; 3 > 4; 3 = 5; 4 = 5]; η_G^2 = .30					
Antonym (17 items)	*M*	11.67	8.38	5.83	5.39	4.60
	SD	2.24	2.94	2.47	2.76	3.42
	F (4, 584) = 73.20** [2 > pre2 > 3 = 4 = 5]; η_G^2 = .33					
Collocation (18 items)	*M*	14.41	12.15	10.34	9.78	7.65
	SD	2.13	2.59	2.49	2.56	4.45
	F (4, 68.07) = 33.80** [2 > pre2 > 3 = 4 = 5]; η_G^2 = .24					

Note. n = 910. [ ] = Multiple comparisons using a Bonferroni method. When the assumption of homogeneity of variances, which is required in using ANOVA, was violated, the Brown-Forsythe modification was utilized using SPSS (2003) based on Glass & Hopkins (1996). See Appendix 9.8.
**p < .01.

Second, relationships between the CASEC scores and the test scores of the Productive VKT were examined. The CASEC intends to assess communicative ability using IRT and a computer adaptive testing system (Japan Institute for Educational Measurement, 2005). It consists of four sections that assess: (1) vocabulary knowledge, (2) knowledge and usage of expressions, (3) ability to understand main information in

listening, and (4) ability to obtain specific information in listening. The test formats of Sections 1 to 3 are multiple-choice questions, and that of Section 4 is for test takers to write down words they hear (i.e., dictation).

Table 9.15 shows means, standard deviations, and reliability (KR-21 and α) for the CASEC scores and Productive VKT. The reliability of all the CASEC section scores and the Productive VKT sections was high.

Table 9.15
Descriptive Statistics of the CASEC and the Size and Depth Sections

	CASEC 1	CASEC 2	CASEC 3	CASEC 4	CASEC Total	Size	Deri	Anto	Collo
M	89.41	99.26	104.65	104.41	397.73	1279.38	8.68	7.60	10.65
SD	23.59	22.41	23.90	30.92	84.67	412.12	3.60	2.67	3.13
KR-21 / α	.95	.96	.97	.98	1.00	.93	.79	.70	.77

Note. *n* = 133. CASEC1 = CASEC Section 1. Deri = Derivation; Anto = Antonym; Collo = Collocation; KR-21 = Kuder-Richardson Formula 21 (calculated for the CASEC scores using CALS, n.d.). α = Cronbach's alpha (for the Size and Depth Sections).

Table 9.16 shows that the Hypothesis for Validity 10 (There are moderate or strong correlations between the Productive VKT section/subsection scores and the CASEC total scores) was supported because there were high correlations between the Productive VKT section/subsections and the CASEC total scores (*r* = .54 to .77).

Table 9.16
Correlations Between the CASEC and the Size and Depth Sections

	CASEC1	CASEC2	CASEC3	CASEC4	CASEC Total
Size (78 items)	.58**	.61**	.59**	.76**	.77**
Derivation (20 items)	.52**	.54**	.50**	.67**	.68**
Antonym (17 items)	.50**	.53**	.55**	.71**	.69**
Collocation (18 items)	.47**	.38**	.46**	.50**	.54**

Note. *n* = 133. CASEC1 = CASEC Section 1.
**$p < .01$.

As for the Hypothesis for Validity 11 (Correlations between the Productive VKT section/subsections and the CASEC Section 1 are higher than those between the Productive VKT section/subsections and the CASEC Sections 2 and 4.), in the Size Section and the Antonym Subsection, there were weaker correlations with Section 1 than with Sections 2 and 3. In the Derivation Subsection, there were stronger correlations with Section 1 than with Section 3 but weaker than with Section 2. In the Collocation Subsection, there were stronger correlations with Section 1 than with Sections 2 and 3. The effect sizes of differences between the correlation coefficients were small or almost zero (q = -.05 to .09; see Table 9.17). Therefore, the Hypothesis for Validity 11 was partially supported only for the Collocation Subsection, and it was not confirmed for the other section/subsections. The reasons may be that the main difference in test items between Sections 1 and 2 was the length of options (i.e., one word or longer options), which suggests that the two sections seem to assess similar aspects of knowledge and that the vocabulary knowledge assessed by the Productive VKT was related to knowledge and ability assessed by Sections 2 and 3.

Table 9.17
Differences Between Correlation Coefficients (q) of the CASEC Sections

	$z_1 - z_2$	$z_1 - z_3$	$z_4 - z_2$	$z_4 - z_3$
Size (78 items)	-0.03	-0.01	0.15	0.17
Derivation (20 items)	-0.02	0.02	0.13	0.17
Antonym (17 items)	-0.03	-0.05	0.18	0.16
Collocation (18 items)	0.09	0.01	0.12	0.04

Note. z_1 = Transformed correlation coefficient in CASEC Section 1.

Regarding the Hypothesis for Validity 12 (Correlations between the Productive VKT section/subsections and the CASEC Section 4 are higher than those between the Productive VKT section/subsections and the CASEC Sections 2 and 4), in the Size Section and the three depth subsections, there were stronger correlations with Section

4 than with Sections 2 and 3. The effect sizes of differences between the correlation coefficients were small or almost zero (q = .04 to .18), so the Hypothesis for Validity 12 was partially supported.

Although the validity of interpretation and uses based on the grades of the STEP Test and of the CASEC was assumed in this study, it needs to be examined further in the future since the validity evidence published is limited and also validity always needs to be checked (e.g., Chapelle, 1999).

9.3.4 Investigating the Hypotheses for Validity 13 and 14 and the Main Hypothesis: Validity Argument of the Productive Vocabulary Knowledge Test

As seen in sections 9.3.1 to 9.3.3, it was found that the Hypotheses for Validity 1, 2, 4, 5, 6, 8, 9, and 10 were supported, while the Hypotheses for Validity 3, 7, 11, and 12 were partially supported.

As for the Hypotheses for Validity 13 regarding the content aspects, points to be reported for the content aspect in the checklist (see section 2.5) were satisfied to a large degree. Construct definition was made clearly and the test method was selected based on the definition. As for the Size Section, content relevance and representativeness were assured because the words were randomly selected based on Levels 1 to 3 in the JACET8000 (JACET Basic Word Revision Committee, 2003). As for the Depth Section, content relevance was assured because the words were selected based on Levels 1 to 3 in the JACET8000. Additionally the Size and Depth Sections were based on the pilot studies and the instructions and difficulty were adjusted, so they were considered to have sufficient technical quality. Therefore, there is sufficient positive evidence for the content aspect of validity, and the Hypothesis for Validity 13 was supported.

Concerning the Hypothesis for Validity 14 (There is a strong correlation between the oral version and the written version of the Size Section of the Productive VKT.), it was supported by the strong correlations between the oral version and the written version of the Productive VKT (r = .89 to .90; see section 8.3). This suggests that the results from the written version of the Productive VKT can be generalizable to results from the oral version of the Productive VKT, at least to the Size Section.

The validity was examined from 14 perspectives. Results that did not agree with the hypotheses were negative evidence for validity, and the reasons for diversions need to be examined, which may lead to further revision of the Productive VKT if necessary. However, much positive evidence for validity was observed. Therefore, the Main Hypothesis (The Productive VKT developed in this study has positive evidence for validity.) was confirmed and there is an acceptable level of validity of inferences and uses based on the test scores for novice Japanese learners of English developed in the current research. Thus, it is concluded that using the Productive VKT in this research is meaningful and appropriate.

9.4 Summary of Chapter 9

The present chapter examines the validity of interpretation and uses based on the scores of the Productive Vocabulary Knowledge Test (Productive VKT) developed for novice Japanese learners of English. It is demonstrated that there is much positive evidence for validity and that there is an acceptable level of validity of inferences and uses based on the test scores.

Chapter 10 Study 2C: Selecting the Speaking Performance Measures

10.1 Purpose, Hypotheses, and a Research Question

The purpose of this chapter is to examine the following two issues regarding lexical complexity in order to select Speaking Performance Measures with more validity. The two issues are (a) whether types can be substituted for tokens and lemmas, and (b) which criterion of sophisticated words is sufficient for analyzing speaking performance of novice learners of English. This study (Study 2C) was conducted as a pilot study of Study 2D (Chapter 11) in order to select more valid measures for the examination of the relationships between productive vocabulary knowledge and speaking performance in Study 2E (Chapter 12). Figure 10.1 shows where the present chapter (i.e., Chapter 10 Study 2C: Selecting the Speaking Performance Measures) is located in the overall research.

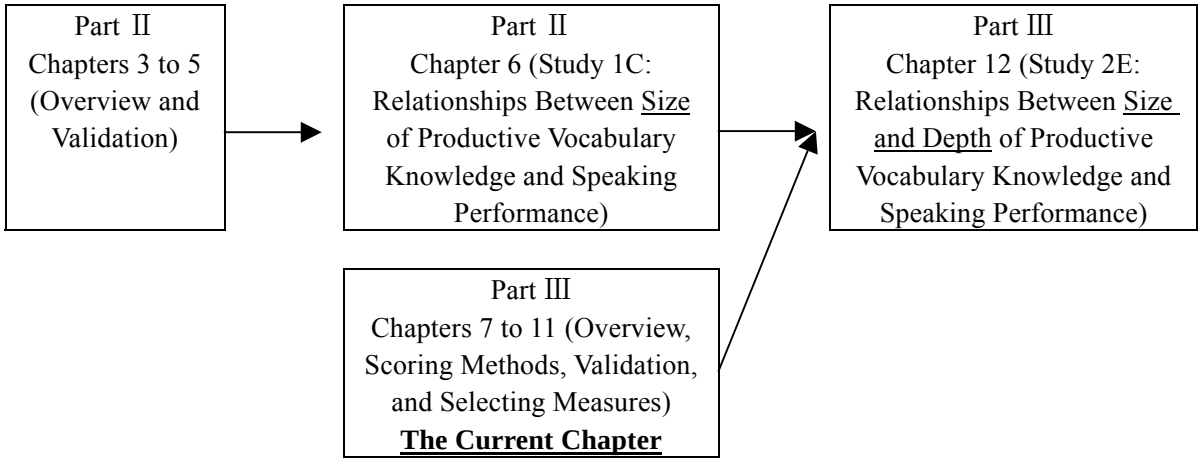


Figure 10.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

Two Hypotheses and three Research Questions were posed in relation to the two aspects examined.

Hypothesis 1: There is a strong correlation between the number of types and the number of lemmas in each task of the Speaking Test.

Hypothesis 2: There is a strong correlation between the number of types and the number of tokens in each task of the Speaking Test.

Research Question: Which criterion of sophisticated words is sufficient for analyzing speaking performance of novice learners of English?

If the two Hypotheses are supported, measures using the number of types can be considered to be substituted for measures using the number of tokens and lemmas in selecting the Speaking Performance Measures of lexical complexity.

10.2 Method

10.2.1 Participants

There were 33 participants in Study 2C. Originally, 34 students participated but one was excluded (see Table 10.1 for the reason). Participants in Study 2 (2B, 2D, and 2E) were basically junior and senior high school students, but Study 2C also involved university undergraduates and graduate school students because it was difficult to get

Table 10.1
Participants at Each School

School	School type	School year	*n* who took the Speaking Test	*n* used for Study 2C	*n* excluded (Reasons for the exclusion)
JHS E	Public	3rd	18	17	1 (The utterances from test takers were not recorded well.)
University	National	1st to 4th	12	12	
Graduate School	National	1st	4	4	
Total			34	33	1 (The utterances from test takers were not recorded well.)

Note. JHS = Junior high school.

participants from high school students. It was assumed that relationships between the Speaking Performance Measures were not very different from those of senior high school students. The participants included 5 males (15.15%) and 28 females (84.85%). Only one junior high school student was part of those who were analyzed in Study 2D.

10.2.2 Instruments

The Speaking Test was used. It used a tape-mediated format. The details were described in section 7.2.2.2.

10.2.3 Procedures

10.2.3.1 Test Administration Procedures

The Speaking Test was conducted as explained in section 7.2.3.1. One difference between Study 2C and other studies that used the Speaking Test was that in Study 2C, 18 test takers out of 33 (54.55%) took the face-to-face version of the Speaking Test first and then took the tape-mediated version of the Speaking Test, whereas in other studies, test takers only took the tape-mediated version. The two versions were conducted within the same day or two days. The face-to-face version of the Speaking Test was developed in order to examine the generalizability aspect of the Speaking Test by comparing the two versions, but the face-to-face version of the Speaking Test was not analyzed in the current study.

10.2.3.2 Coding Procedures

The procedures of transcription were mostly the same as those used in Study 1B (see section 5.2.3.2). In this section, first, all the Speaking Performance Measures used in Study 2C and rationales for including new measures were presented, and then

elements to compute values of the Speaking Performance Measures were defined and coded. Table 10.2 shows comparisons between the measures used in Studies 1 and 2 from four perspectives: fluency, accuracy, syntactic complexity, and lexical complexity.

Table 10.2
Comparisons Between the Speaking Performance Measures Used in Studies 1B and 2C

Construct	Code in Study 1	Measure of Study 1	Changes in Study 2
Fluency	F1	No. of tokens per minute	
	F2	No. of tokens per AS-unit	
	F3	No. of clauses per minute	
	F4	No. of dysfluency markers per minute	
	F5	No. of dysfluency markers per token	
	F6	No. of dysfluency markers per AS-unit	
Accuracy	A1	No. of error-free clauses per clause	
	A2	No. of error-free AS-units per AS-unit	
	A3	No. of errors per token	Excluded [1]
	A4	No. of errors per AS-unit	Excluded [1]
Syntactic complexity	SC1	No. of clauses per AS-unit	
	SC2	No. of subordinate clauses per AS-unit	
			Added: No. of subordinate clauses per clause [2]
	SC3	No. of tokens per AS-unit	
Lexical complexity	LC1	Guiraud index: No. of types divided by the square root of No. of tokens	Modified: The definition of "types" [3]
	LC2	Lexical density: No. of lexical tokens per token	
	LC3	Weighted lexical density: (No. of sophisticated lexical tokens and No. of basic lexical tokens given half the weight) divided by No. of tokens	Excluded [4]
	LC4	No. of sophisticated word types per token (The definition of "sophisticated": words not in the list of 1,250 words in the JACET8000)	Added: Other definitions of "sophisticated," which are compared in Study 2C [6] Added: 14 new measures of lexical complexity [5]

Note. [ ] = The order in which the explanations of the differences were given below. Although Study 1B found the AS-unit length (No. of tokens per AS-unit) tended to mean syntactic complexity rather than fluency, the analysis of the meaning of this measure was also conducted in order to cross validate the result. Therefore, the same measures of F2 and SC3 were included in Table 10.2.

While most measures are the same, there are six points that are different from Study 1B. First, as for accuracy measures, A3 and A4 were excluded in Study 2. The measures of the proportion of errors were not used because through Study 1B, it was found that it is difficult to make a valid method to count the number of errors. Novice speakers of English tend to produce many errors in one unit (e.g., AS-unit and clauses) and some qualitatively different errors close to each other. For example, the utterance *She like apple.* (compare with *She likes apples.*) has two types of error: an error of agreement between a subject and a verb, and an error in choosing between a singular or plural form. Should these be counted as two errors or one (because the two errors are neighbors)? Study 1B counted the adjacent errors as one. However, whether this method is valid enough seems to need further examination, although it was rather beyond the scope of Study 2C. It was decided that measures that used the number of errors would not be used and only measures with error-free clauses and AS-units were utilized.

Second, "the number of subordinate clauses per clause" was added for syntactic complexity because this measure seemed to be missing.

Third, concerning the definition of types, in Study 1B, types meant lemmas (i.e., type = lemma), based on Daller et al. (2003). In other words, a base form and inflected forms were considered to be the same type. For example, the following were considered to be one type: *play, plays, playing*; *be, is, am, are, was, were, been*; *Taro, Taro's*. In Study 2, *play, plays, playing; be, is, am, are, was, were, been*; *Taro, Taro's* were all counted as one separate type. That is, types were considered to be different if the form was different. This change was done because the latter method seemed to be more common in vocabulary research. However, if the number of lemmas and that of types are very different, this change of the definition may affect the results. The

question here is whether the number of types is closely associated with the number of lemmas. If so, the results of the measures using the number of types in Studies 1 and 2 may be comparable. This question is examined in the Hypothesis 1.

Fourth, LC3 (Weighted lexical density: [the number of sophisticated lexical tokens and the number of basic lexical tokens given half the weight] divided by the number of tokens) used in Study 1B was excluded because LC3 seemed redundant since this measure combines LC2 and LC4.

Fifth, Study 1 used the measures used in the previous speaking studies. The main results of Study 1C were that there were moderate or strong relationships across tasks between size of productive vocabulary knowledge and speaking performance as measured by SC3 (the AS-unit length: the number of tokens per AS-unit) and L1 (the Guiraud index: the number of types divided by the square root of the number of tokens). The two measures were both related to vocabulary, and based on the results of Study 1C, it was decided that Study 2 (2C) used a greater number of measures of lexical complexity to focus more on vocabulary performance. Table 10.3 shows 14 new lexical complexity measures newly added in Study 2C.

Table 10.3
Summary of 14 Lexical Complexity Measures Used in Study 2C

Code	Measure
LCN1	No. of sophisticated word types divided by No. of word types
LCN2	No. of word types per AS-unit
LCN3	No. of lexical word types per AS-unit
LCN4	No. of sophisticated word types per AS-unit
LCN5	No. of word types per clause
LCN6	No. of lexical word types per clause
LCN7	No. of sophisticated word types per clause
LCN8	No. of sophisticated tokens per token
LCN9	No. of lexical word types per token
LCN10	No. of lexical word types per word type
LCN11	No. of lexical tokens divided by the square root of No. of tokens
LCN12	No. of sophisticated word types divided by the square root of No. of tokens
LCN13	No. of sophisticated tokens divided by the square root of No. of tokens
LCN14	No. of lexical word types divided by the square root of No. of tokens

Out of the 14 measures, seven new measures (LCN1 to LCN7) were derived from Wolfe-Quintero et al. (1998), who proposed some potentially good measures to be examined in future studies. Seven additional new measures (LCN8 to LCN14) were included because they were measures that can be calculated but that were missing in the list. In particular, LCN11 to LCN14 were devised based on the background of why the Guiraud index was advocated rather than a simple type token ratio (see section 5.2.3.2). In fact, Daller et al. (2003) used LCN12 (the number of sophisticated word types divided by the square root of the number of tokens), but this measure was not included in the list because there was no other study that used it (see section 2.3.2). Now there are 17 measures of lexical complexity (i.e., 4 – 1 + 14), and some of them are very similar to each other. For example, LC5 (the number of lexical word types divided by the square root of the number of tokens) and LC10 (the number of lexical tokens divided by the square root of the number of tokens) are almost the same except that LC5 uses types and LC10 uses tokens. The question here is whether the number of types assesses similar aspects to the number of tokens and whether the former is closely associated with the latter. If so, some measures can be substituted for other measures without missing potentially important aspects of lexical complexity. This question was examined in the Hypothesis 2.

Concerning the sixth change of the measures from Study 1B to Study 2C, 15 definitions of sophisticated words were devised in Study 2C (see Table 10.4). This was done because the criterion used in Study 1B of considering 1,001 words or above as sophisticated may not have been very appropriate for novice learners. The previous study suggests that a finer criterion of using 100 words as a unit may lead to different results (Kaneko, 2004; see section 2.4.1). The finer criteria were compared and a

criterion that works well was explored in the Research Question. Furthermore, two points should be added. First, J1001 was used in Study 2 (2C to 2E) in order to maintain the comparability of Studies 1 (1B and 1C) and 2 (2C to 2E), so J1001 was not compared with other definitions, and 14 definitions were compared. Second, for the definition of sophisticated words, it was decided to use one additional vocabulary list in addition to the JACET8000 (JACET Basic Word Revision Committee, 2003): Ota, Hidai, and Kajiro (2004).

Table 10.4
Summary of 15 Definition of Sophisticated Words Examined in Study 2C

Code	Definition
J101	101 to 8,000 words in the JACET8000
J201	201 to 8,000 words in the JACET8000
J301	301 to 8,000 words in the JACET8000
J401	401 to 8,000 words in the JACET8000
J501	501 to 8,000 words in the JACET8000
J601	601 to 8,000 words in the JACET8000
J701	701 to 8,000 words in the JACET8000
J801	801 to 8,000 words in the JACET8000
J901	901 to 8,000 words in the JACET8000
J1001	1,001 to 8,000 words in the JACET8000 (Not compared in Study 2C)
O101	101 to 8,000 words based on Ota et al. and the JACET8000
O201	201 to 8,000 words based on Ota et al. and the JACET8000
O301	301 to 8,000 words based on Ota et al. and the JACET8000
O401	401 to 8,000 words based on Ota et al. and the JACET8000
O501	501 to 8,000 words based on Ota et al. and the JACET8000

Note. O101 was calculated using the following formula: [The number of words that belong to 1 to 8000 words in the JACET8000] minus [The number of words that belong to 1 to 100 words in Ota et al.'s list].

Two vocabulary lists were used because of the possibility that one could complement aspects missed by the other. According to the JACET Basic Word Revision Committee (2003), the JACET8000 was developed based on the three corpora: (a) the British National Corpus, (b) the sub-corpora derived from texts that Japanese learners of English may encounter (e.g., newspapers, magazines, authorized junior and senior high school textbooks, the STEP Test, TOEFL [Test of English as a

Foreign Language], and TOEIC [Test of English for International Communication]), and (c) the corpus of senior high school textbooks (pp. 103-105). In contrast, Ota et al.'s (2004) list was developed based on the following six corpora (p. 3): (a) authorized junior high school textbooks, (b) paper-and-pencil tests of the third, fourth, and fifth grades of the STEP Test, (c) entrance examinations to enter national, public, and private senior high schools all over Japan, (d) learner corpora, which include scripts of diaries, introductions to Japanese culture, and speeches, (e) the vocabulary list derived from the 1,000 most frequent spoken words and the 1,000 most frequent written words in the *Longman Dictionary of Contemporary English* (3rd ed.), and (f) the vocabulary list derived from about 680 most important spoken and written words in the *Collins COBUILD English Dictionary for Advanced Learners* (3rd ed.).

To summarize, while the two lists can be considered to represent a vocabulary list that took the educational environment of the Japanese into account, they seem to be different in terms of target learners. The JACET8000 is a list of words known by Japanese learners of English in general, including junior high school students and adults, whereas Ota et al. (2004) show a list of words that junior high school students may encounter and in this sense, it seems to be a list for beginners. Although the JACET8000 is appropriate enough to be used for Japanese learners of English, Ota et al. (2004) appear to have the advantage of having the data from learners' spoken words and thus it may be more appropriate for use by Japanese learners of English at the novice level, who were targeted in this study. One point to be added is that the reason the JACET8000, not Ota et al. (2004), was basically used for developing the Productive Vocabulary Knowledge Test (Productive VKT) was that Ota et al. (2004) provide a list of only 1,210 words and some of the developing process is not explicitly recorded, whereas the JACET8000 describes the developing procedures clearly, which

leads to the decision that Ota et al. (2004) should be used as a complementary source. Table 10.5 shows the proportions of words shared between Ota et al.'s (2004) vocabulary range and the JACET8000 word range. It was found that 71% of the 1 to 100 words in Ota et al. (2004) appear in the 1 to 100 word range in the JACET8000, but for the other range, only a small proportion of words are shared (see Appendix 10.1). This suggests that the two lists cover rather different areas of words, which may lead to different results of measures with sophisticated words. Although it is possible to utilize all the 14 definitions of sophisticated words, the analyses may be too complicated. It was decided to select the most appropriate definition of all the 14 definitions in Study 2C, and this issue was taken up in the Research Question.

Table 10.5
Relationships Between Ota et al. (2004) and the JACET8000

	JACET 1-100	JACET 101-200	JACET 201-300	JACET 301-400	JACET 401-500	JACET 501 or above	Plus 250	Not in the list
Ota 1-100	71 (.71)	18 (.18)	7 (.07)	2 (.02)	1 (.01)	0 (.00)	1 (.01)	0 (.00)
Ota 101-200	37 (.37)	20 (.20)	11 (.11)	4 (.04)	6 (.06)	15 (.15)	7 (.07)	0 (.00)
Ota 201-300	10 (.10)	28 (.28)	15 (.15)	13 (.13)	5 (.05)	23 (.23)	6 (.06)	0 (.00)
Ota 301-400	0 (.00)	21 (.21)	18 (.18)	11 (.11)	4 (.04)	37 (.37)	8 (.08)	1 (1.00)
Ota 401-500	2 (.02)	10 (.10)	14 (.14)	18 (.18)	8 (.08)	39 (.39)	6 (.06)	3 (.03)

Note. Ota = Ota et al. (2004); ( ) = [The number of words shared between the two lists]/100.

Table 10.6 shows the list of Speaking Performance Measures whose changes and modifications described above were taken into consideration.

Table 10.6
Summary of 29 Speaking Performance Measures

Construct	New code	Old code	Measure of Study 2C
Fluency	F1	F1	No. of tokens per minute
	F2	F2	No. of tokens per AS-unit
	F3	F3	No. of clauses per minute
	F4	F4	No. of dysfluency markers per minute
	F5	F5	No. of dysfluency markers per token
	F6	F6	No. of dysfluency markers per AS-unit
Accuracy	A1	A1	No. of error-free clauses per clause
	A2	A2	No. of error-free AS-units per AS-unit
Syntactic complexity	SC1	SC1	No. of clauses per AS-unit
	SC2	SC2	No. of subordinate clauses per AS-unit
	SC3		No. of subordinate clauses per clause
	SC4	SC3	No. of tokens per AS-unit
Lexical complexity			(Lexical diversity)
	LC1	LC1	Guiraud index: No. of types divided by the square root of No. of tokens
	LC2	LCN2	No. of word types per AS-unit
	LC3	LCN5	No. of word types per clause
			(Lexical density)
	LC4	LCN9	No. of lexical word types per token
	LC5	LCN14	No. of lexical word types divided by the square root of No. of tokens
	LC6	LCN3	No. of lexical word types per AS-unit
	LC7	LCN6	No. of lexical word types per clause
	LC8	LCN10	No. of lexical word types per word type
	LC9	LC2	Lexical density: No. of lexical tokens per token
	LC10	LCN11	No. of lexical tokens divided by the square root of No. of tokens
			(Lexical sophistication)
	LC11	LC4	No. of sophisticated word types per token
	LC12	LCN12	No. of sophisticated word types divided by the square root of No. of tokens
	LC13	LCN4	No. of sophisticated word types per AS-unit
	LC14	LCN7	No. of sophisticated word types per clause
	LC15	LCN1	No. of sophisticated word types divided by No. of word types
	LC16	LCN8	No. of sophisticated tokens per token
	LC17	LCN13	No. of sophisticated tokens per token

Note. F2 and SC3 are the same measures. Only lexical complexity measures were investigated in Study 2C.

As seen in the Research Questions, this study (Study 2C) examined four lexical aspects of speaking performance: the number of types, the number of lemmas, the number of tokens, and "the number of sophisticated word types divided by the number of types" (LC15).

Tokens were pruned tokens after dysfluency markers (i.e., functionless repetitions, self-corrections, and false starts) were excluded. Filled pauses (e.g., *mm*, *ah*) were not included in the tokens nor unfilled pauses (e.g., pauses during which speakers did not say anything). In counting types, a base form and inflected forms (e.g., the base form of *lives* is *live* and one inflected form of *live* is *lives*) were considered to be different types (e.g., *play, plays, playing*; *be, is, am, are, was, were, been*; *Taro, Taro's*). In contrast, in counting lemmas, a base form and inflected forms were considered to be the same lemma. For example, the following were considered to be one lemma: *play, plays, playing* (one lemma); *be, is, am, are, was, were, been* (one lemma); *Taro, Taro's* (one lemma). Therefore, the number was largest for tokens, followed by types, and lemmas had the smallest number. In counting the number of types and tokens, the wordlist function of Tsukamoto KWIC (Tsukamoto, 2004) was used initially and then the number of words with apostrophes (’), which Tsukamoto KWIC cannot deal with, were manually modified by the author.

Fourteen definitions of sophisticated words were compared in Study 2C (see Table 10.4). The JACET 8000 Level Marker (Shimizu, 2005) was used for checking whether words exist in the 1,001 to 8,000 word range or in the 1,000 word range of the JACET8000 list. Numerals are included in the JACET8000. Proper nouns and Japanese words were excluded from the list of sophisticated words, and the criteria were the same as those used in Study 1B. Other judgments were done twice manually using the vocabulary lists by the author. Inflected words were judged by their base forms as to whether they were sophisticated words. After counting the number of sophisticated types, values were computed using the formula: [The number of sophisticated types for each definition]/[The number of types].

10.2.3.3 Analysis Procedures

The two Hypotheses and Research Questions were answered using Pearson's product-moment correlation coefficients (r). The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1).

10.3 Results and Discussion

Table 10.7 demonstrates that the relationships between the number of types and lemmas were also very strong (r = .99 or more) in all five tasks and the relationships between the number of types and tokens were very strong (r = .89 or more) in all five tasks (supporting the Hypotheses 1 and 2). Therefore, the results of Studies 1B and 2C using different definitions of types (see section 10.2.3.2) can be considered to be comparable and the measures using tokens can be substituted by the measures using types. Table 10.8 shows the measures that were excluded in the subsequent study based on these results.

Table 10.7
Correlations Between the Number of Types, Lemmas, and Tokens

	Task 1 (n = 32)	Task 2 (n = 22)	Task 3 (n = 28)	Task 4 (n = 30)	Task 5 (n = 24)
r between No. of types and lemmas	1.00**	.99**	.99**	1.00**	.99**
r between No. of types and tokens	.93**	.89**	.96**	.97**	.91**
r between No. of lemmas and tokens	.94**	.88**	.95**	.97**	.88**

Note. **p < .01.

Table 10.8
Four Speaking Performance Measures Excluded Based on the Results of Study 2C

Construct	Code	Old code	Measure of Study 2
Lexical complexity	LC9	LC2	Lexical density: No. of lexical words per word Substituted by LC4 and LC8
	LC10	LCN11	No. of lexical tokens divided by the square root of No. of tokens Substituted by LC5
	LC16	LCN8	No. of sophisticated word tokens per token Substituted by LC11 and LC15
	LC17	LCN13	No. of sophisticated word tokens per token Substituted by LC12

Next, in order to discover a criterion of sophisticated words that works well for novice learners of English (Research Question), 14 definitions were compared using two analyses: examination of standard deviations and of correlation coefficients between the measures using different definitions of sophisticated words.

First, the criterion to be used should have more variations because without variations, relationships with other variables are unlikely to appear. This seems important when the purpose of the study is to examine relationships between the measures and other variables. For each task, the criterion with the largest standard deviations was sought. Table 10.9 shows means and standard deviations for each definition and each task.

As in Table 10.9, the criteria with the largest standard deviations regarding LC15 were different for each task. For example, in Task 1, the criterion of J201 and O101 had the largest *SD*s, whereas in Task 2, the criterion of J101, J201, and O501 had the largest *SD*s. The trends of different results for each task may be because of effects of the contents or elements in the five tasks.

Table 10.9
Descriptive Statistics of the Measures of Lexical Sophistication (LC15) Derived from the 14 Definitions

	Task 1 (n = 32)		Task 2 (n = 22)		Task 3 (n = 28)		Task 4 (n = 30)		Task 5 (n = 24)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
J101	0.43	0.08	0.51	0.09	0.52	0.07	0.47	0.11	0.43	0.09
J201	0.24	0.09	0.36	0.09	0.39	0.10	0.42	0.11	0.35	0.11
J301	0.19	0.07	0.29	0.08	0.32	0.09	0.39	0.11	0.30	0.11
J401	0.16	0.06	0.21	0.08	0.29	0.09	0.32	0.09	0.20	0.09
J501	0.14	0.06	0.12	0.07	0.29	0.09	0.29	0.09	0.19	0.08
J601	0.11	0.06	0.10	0.06	0.27	0.08	0.25	0.07	0.18	0.08
J701	0.10	0.05	0.07	0.04	0.25	0.08	0.24	0.07	0.15	0.08
J801	0.09	0.05	0.07	0.04	0.18	0.07	0.19	0.06	0.14	0.06
J901	0.08	0.05	0.07	0.04	0.17	0.08	0.19	0.06	0.13	0.06
O101	0.52	0.09	0.57	0.07	0.57	0.07	0.58	0.12	0.43	0.11
O201	0.23	0.07	0.42	0.08	0.36	0.10	0.43	0.12	0.33	0.10
O301	0.17	0.08	0.30	0.07	0.30	0.09	0.33	0.12	0.31	0.11
O401	0.14	0.07	0.28	0.07	0.29	0.08	0.28	0.10	0.27	0.09
O501	0.12	0.07	0.23	0.09	0.26	0.08	0.24	0.09	0.19	0.10

Note. Underlined = The criterion that had the largest *SD*. See Table 10.4 for the codes (e.g., J101).

Next, correlations were examined between LC15 derived from the definition of J101 and other definitions of sophisticated words with the largest standard deviations (see Table 10.10). It was found that all the correlations with J101 were strong (r = .52 to .90; Cohen, 1988; see section 2.6.4.1), which suggests that J101 works sufficiently as a substitute for definitions of sophisticated words with the largest standard deviations and that using only J101 is sufficient (an answer to the Research Question: Which criterion of sophisticated words is sufficient for analyzing speaking performance of novice learners of English?). Therefore, J101 was used for Study 2D (the next chapter) in addition to J1001.

Table 10.10
Intercorrelations Between the Measures of Lexical Sophistication (LC15) Derived from the Definition of J101 and Other Definitions with the Largest Standard Deviations

Task 1 (n = 32)	Task 2 (n = 22)	Task 3 (n = 28)	Task 4 (n = 30)	Task 5 (n = 24)
J201 (.81**)	J201 (.77**)	J201 (.81**)	O101 (.77**)	J201 (.90**)
O101 (.64**)	O501 (.52*)	O201 (.51**)	O201 (.86**)	J301 (.88**)
			O301 (.55**)	O101 (.86**)
				O301 (.65**)

Note. J201 (.81) = r between J101 and J201 was .81.
*p < .05. **p < .01.

Study 1C shows that the number of types is correlated very closely with the number of lemmas and tokens and it was decided to use the criterion of considering 101 to 8,000 words (J101) in the JACET8000 as sophisticated words. Additionally, in order to compare results with Study 1 (1B and 1C), the criterion of 1,001 to 8,000 words was also maintained.

10.4 Summary of Chapter 10

This chapter examines (a) whether types can be substituted for tokens and lemmas, and (b) which criteria of sophisticated words should be used in analyzing speaking performance of novice learners of English. It is shown that relationships between the number of types and tokens and between the number of types and lemmas are very strong. Therefore, the results of Studies 1 (1B and 1C) and 2 (2C to 2E) using different definitions of types can be considered to be comparable and the measures using tokens can be substituted by the measures using types. The results also indicate that the criterion of considering 101 to 8,000 words (J101) in the JACET8000 as sophisticated words should be used.

Chapter 11 Study 2D: Validation of the Speaking Performance Measures[11.1]

11.1 Purpose and Hypotheses

The purpose of this chapter is to examine the qualities of the various Speaking Performance Measures, with a specific focus on Japanese beginning learners of English. Specifically, this study investigates (a) whether the AS-unit length (the number of tokens per AS-unit) indicates fluency or syntactic complexity and (b) whether the Speaking Performance Measures of fluency, accuracy, syntactic complexity, and lexical complexity have positive evidence for validity. The Speaking Performance Measures are used in the main examination of relationships between productive vocabulary knowledge and speaking performance in Chapter 12 (Study 2E). Figure 11.1 shows where the present chapter (i.e., Chapter 11 Study 2D: Validation of the Speaking Performance Measures) is located in the overall research.

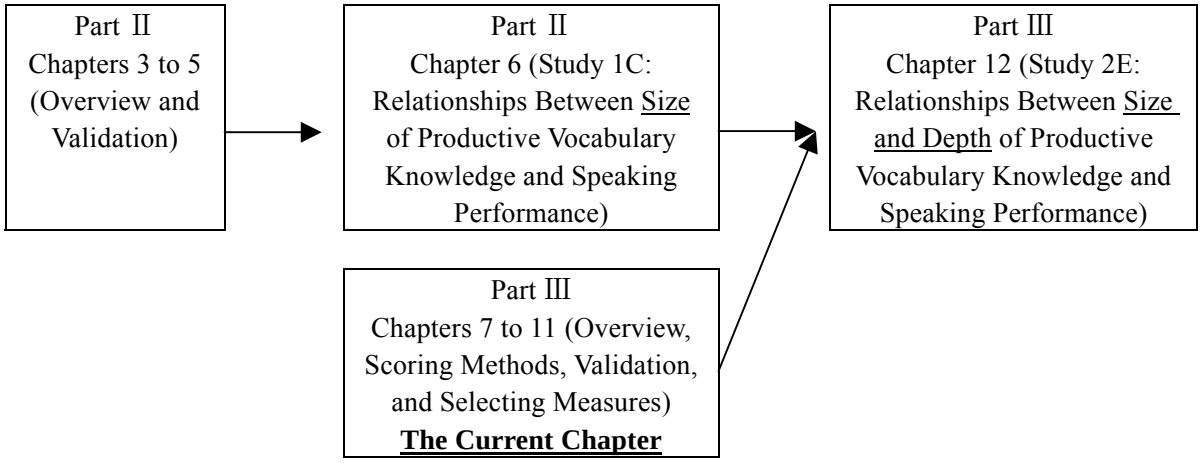


Figure 11.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

Two Main Hypotheses were posed in relation to the interpretation and validity of the Speaking Performance Measures. In order to examine the Main Hypothesis 2, seven Hypotheses for Validity were made. Table 11.1 illustrates the organization of the

current study (Study 2D).

Main Hypothesis 1: The AS-unit length (the number of tokens per AS-unit) loads on components that syntactic complexity measures load on and not on components that fluency measures load on. In other words, the AS-unit length is a syntactic complexity measure rather than a fluency measure.

Main Hypothesis 2: The Speaking Performance Measures have positive evidence for validity.

Hypothesis for Validity 1 (the substantive aspect): There is a positive response from test takers concerning the quality of the Speaking Test.

Hypothesis for Validity 2 (the consequential aspect): There is a positive response from test takers concerning the effect of the Speaking Test on learning English.

Hypothesis for Validity 3 (the structural aspect): The Speaking Performance Measures of each construct (i.e., fluency, accuracy, syntactic complexity, and lexical complex) in the same task load on the same component.

Hypothesis for Validity 4 (the generalizability aspect): Each measure across tasks loads on the same component.

Hypothesis for Validity 5 (the external aspect): Correlations between components of the same constructs are higher than those between components of the different constructs (if there is more than one component in the analyses of the Hypotheses for Validity 3 and 4).

Hypothesis for Validity 6 (the content aspect): Each measure has positive evidence for the content aspect.

Hypothesis for Validity 7 (the generalizability aspect): The reliability of raters is

high on the Speaking Performance Measures.

Table 11.1
Organization of Study 2D: Validation of the Speaking Performance Measures

	Validity aspect	Method to be used	Section reported
Main Hypothesis 1			11.3.3
Main Hypothesis 2			11.3.5
Hypothesis for Validity 6	Content	Expert professional judgment	11.3.6
Hypothesis for Validity 1	Substantive	Questionnaire	11.3.2
Hypothesis for Validity 3	Structural	Principal components analysis (PCA)	11.3.4
Hypothesis for Validity 4	Generalizability	・Reliability in showing consistency of performance across raters ・PCA	11.3.4 & 11.3.6
Hypotheses for Validity 5 and 7	External	Correlations	11.3.5 & 11.3.6
Hypothesis for Validity 2	Consequential	Questionnaire	11.3.2

The Main Hypothesis 1 was based on the previous studies (see section 2.3.1). The Hypothesis for Validity 5 was based on the principle of internal construction validation (Henning, 1987) that correlations between similar aspects are stronger than those between different aspects. The Hypotheses for Validity 1 and 2 were about the validity regarding the Speaking Test, whereas the Hypotheses for Validity 3 to 7 were about the validity regarding the Speaking Performance Measures.

Although examination of the same Main Hypotheses 1 and 2 was conducted in Study 1B, this study (Study 2D) investigated them again for the following four reasons. First, the Main Hypothesis 1 was supported in Study 1B but was tested again here in order to cross validate the results. Second, Study 2D increased the number of the Speaking Performance Measures used and some of them are new measures that need validation. Third, the validation of interpretation and uses based on scores and values is always needed, especially when the range of target test takers varies, which is the

case in Study 2D. Fourth, since Study 2D involves a greater number of participants than Study 1B, multivariate analyses can be used to combine speaking performance measures in order to enhance interpretation.

11.2 Method

11.2.1 Participants

The participants in this study were 244 students studying English in Japan (see Table 11.2). This group contained 141 (both public and national) junior high school

Table 11.2
Participants at Each School

School	School type	School year	*n* who took the Speaking Test	*n* for Study 2D	*n* excluded (Reasons for the exclusion)
JHS A	Public	3rd	14	7	7 (Problem 3, Name 2, No clauses 2)
JHS B	Public	3rd	57	34	23 (Problem 3, Name 10, No clauses 9, Playful 1)
JHS C	Public	3rd	26	14	12 (Problem 1, Name 5, No clauses 6)
JHS D	Public	3rd	27	6	21 (Interview 7, Problem 2, Name 1, No clauses 9, Playful 2)
JHS G	National	3rd	104	80	24 (Problem 15, Name 7, No clauses 2)
JHS Total			228	141	87 (Interview 7, Problem 24, Name 25, No clauses 28, Playful 3)
SHS A	Public	3rd	28	23	5 (No clauses 5)
SHS B	Public	1st to 3rd	22	17	5 (Problem 1, No clauses 4)
SHS C	Public	1st	36	33	3 (No clauses 3)
SHS E	Public	2nd to 3rd	15	9	6 (Problem 1, No clauses 5)
SHS G	National	1st	23	21	2 (Problem 1, Name 1)
SHS Total			124	103	21 (Problem 3, Name 1, No clauses 17)
Total			352	244	108 (Interview 7, Problem 27, Name 26, No clauses 45, Playful 3)

Note. JHS = Junior high school; SHS = Senior high school; No clauses = A student who did not produce any clauses in at least one task; Playful = A student who did not behave seriously (e.g., who sang a song during a task); Problem = A student who had a mechanical or procedural problem (e.g., a student who had a Version A tape and a Version B test booklet; see section 7.2.3.1); Interview = A student who took the face-to-face version of the Speaking Test first (see section 8.2.1 for the face-to-face version); Name = A student whose tape caused a matching problem between the tape and the name of the speaker.

students and 103 (both public and national) senior high school students, including 104 males (42.62%) and 140 females (57.38%). Their first language (L1) was Japanese. They were aged from 14 to 18, and all had studied English as a foreign language for approximately two to five years. Therefore, a wide range of learners in terms of English ability participated. In fact, the range in Study 2D was much wider than that in Study 1B. They were considered to be novice Japanese learners of English in this research. All of them were part of those who were analyzed in Chapter 9.

Originally, 352 students took the Speaking Test, but 108 students were not used for analysis (see Table 11.2 *Note*. for all the reasons). As seen in Table 11.3, out of the 108, 45 test takers who did not produce any clauses (including at least a subject and a verb) in at least one task were not used. In other words, the results cannot be generalized to junior and senior high school students who do not say at least one clause or more. Students who were totally silent during a task were excluded because they may not have understood what to do in the task. The reason for using a clause as the minimal unit was that clauses were used for the denominator in computing the Speaking Performance Measures (see section 5.2.1 for more explanation).

Table 11.3
Excluded Participants According to Each Reason

School	*n* of ST	*n* for Study 2D	Inter-view	Prob-lem	Name	No clauses	Play-ful	*n* ex-cluded
JHS	228	141	7 (3.07)	24 (10.53)	25 (10.96)	28 (12.28)	3 (1.32)	87 (38.16)
SHS	124	103	0 (0.00)	3 (2.42)	1 (0.81)	17 (13.71)	0 (0.00)	21 (16.94)
Total	352	244	7 (1.99)	27 (7.67)	26 (7.39)	45 (12.78)	3 (0.85)	108 (30.68)

Note. See Table 11.2 for the codes. ( ) = %. *n* of ST = *n* who took the Speaking Test.

Among the excluded participants, most were excluded because they did not

produce clauses in at least one task, which accounted for over 10% for both junior and senior high school students (i.e., 12.28% and 13.71%). Most of them produced fragmented utterances, such as “Horse” “Open the window” and “Notebook on the desk.” The abundance of those who did not utter clauses may indicate that the tasks used in this test were rather difficult for a certain percentage of students, probably because they were not accustomed to speaking without preparation or they did not have enough knowledge and ability to accomplish the tasks.

For junior high school students, a high percentage was also obtained for (a) those who had administrative problems such as recording difficulties (see the column of Problem in Table 11.3), and (b) those whose tape caused a matching problem between the tape and the name of the speaker (see the column of Name). The second problem was that some test takers did not say their names or said the name so softly that it was difficult to identify. As for the administrative problems, it seems difficult to reduce them because the pilot study was conducted and the main study was well planned in advance. However, the matching problem between the tape and the name of the speaker could have been avoided if test takers had been asked to write their names on the seals of the tape. Although lack of time made this impossible at some schools, at other schools it could have been done. This could be improved in a future study.

11.2.2 Instruments

The Speaking Test and the Speaking Test Questionnaire were used in this chapter. The Speaking Test used a tape-mediated format, which included five tasks. The details were described in sections 7.2.2.2 and 7.2.2.3.

11.2.3 Procedures

11.2.3.1 Test Administration Procedures

The details were described in section 7.2.3.1.

11.2.3.2 Coding Procedures

The procedures of transcription were mostly the same as those used in Study 1B (see section 5.2.3.2). As in Table 11.4, 30 measures (six for fluency, two for accuracy, four for syntactic complexity, and 18 for lexical complexity) were used in this study (see sections 5.2.3.2, 10.2.3.2, and 10.3, for rationales for the selection). Out of the 30, F2 and SC4 (the number of words per AS-unit) were the same measure, so there were actually 29 measures. Since five tasks were analyzed, and values of the measures were computed for each task, there were 145 variables in total (i.e., 29 measures * 5 tasks).

In order to compute 29 Speaking Performance Measures, coding was carried out for the following aspects: the number of AS-units, clauses, subordinate clauses, dysfluency markers, error-free clauses, error-free AS-units, tokens, word types, lexical word types, sophisticated word types. The definition was the same as that used in Study 1 unless noted below. Before raters began independent coding, they practiced coding based on the definitions and examples. Examples of coding and computing measures are seen in Appendix 11.1 (see also Appendix 11.2).

As for the number of AS-units, clauses, subordinate clauses, and dysfluency markers, more than one-third (42.62%; 104/244) was judged by two raters (the author and another rater majoring in language testing; Raters A and E). The correlations of the two raters were very high (see Table 11.5). When there were disagreements, the raters discussed the differences until they reached an accord, and the result was used as the final coding. There was no inconsistent coding regarding clauses and dysfluency

Table 11.4
Summary of 30 Speaking Performance Measures Used in Study 2D

Construct	Code	Old code	Measure of Study 2D
Fluency	F1	F1	No. of tokens per minute
	F2	F2	No. of words per AS-unit
	F3	F3	No. of clauses per minute
	F4+	F4	No. of dysfluency markers per minute
	F5+	F5	No. of dysfluency markers per token
	F6+	F6	No. of dysfluency markers per AS-unit
Accuracy	A1	A1	No. of error-free clauses per clause
	A2	A2	No. of error-free AS-units per AS-unit
Syntactic complexity	SC1	SC1	No. of clauses per AS-unit
	SC2	SC2	No. of subordinate clauses per AS-unit
	SC3		No. of subordinate clauses per clause
	SC4	SC3	No. of tokens per AS-unit
Lexical complexity			(Lexical diversity)
	LC1	LC1	Guiraud index: No. of types divided by the square root of No. of tokens
	LC2	LCN2	No. of types per AS-unit
	LC3	LCN5	No. of types per clause
			(Lexical density)
	LC4	LCN9	No. of lexical word types per token
	LC5	LCN14	No. of lexical word types divided by the square root of No. of tokens
	LC6	LCN3	No. of lexical word types per AS-unit
	LC7	LCN6	No. of lexical word types per clause
	LC8	LCN10	No. of lexical word types per type
			(Lexical sophistication using J1001)
	LC11	LC4	No. of sophisticated word types per token
	LC12	LCN12	No. of sophisticated word types divided by the square root of No. of tokens
	LC13	LCN4	No. of sophisticated word types per AS-unit
	LC14	LCN7	No. of sophisticated word types per clause
	LC15	LCN1	No. of sophisticated word types divided by No. of types
			(Lexical sophistication using J101)
	LC11	LC4	No. of sophisticated word types per token
	LC12	LCN12	No. of sophisticated word types divided by the square root of No. of tokens
	LC13	LCN4	No. of sophisticated word types per AS-unit
	LC14	LCN7	No. of sophisticated word types per clause
	LC15	LCN1	No. of sophisticated word types divided by No. of types

Note. + = Measure that has a negative meaning when it has a higher value. F2 and SC3 are the same measures. The old codes were used in Studies 1B, 1C, or 2C. J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

markers between the author and the other rater. Since the definitions and coding systems were considered to be established, the rest of the utterances were judged by

the author. As for the judgment of subordinate clauses, correlations were overall a little lower than others because the definition of subordinate clauses was a little different from the usual ones. For instance, in the utterance "I like reading books," *reading books* was considered a subordinate clause in this study (based on Foster et al., 2000; see section 5.2.3.2.2), and one rater overlooked

Table 11.5
Interrater Reliability (r) of the Number of AS-Units, Clauses, Subordinate Clauses, and Dysfluency Markers

Task	AS-unit	Clauses	Subordinate clauses	Dysfluency markers
1	.99**	.95**	.87**	.99**
2	.96**	.95**	1.00**	.86**
3	.98**	.92**	.73**	.94**
4	.97**	.96**	.84**	.94**
5	.95**	.94**	.92**	.91**

Note. $n = 104$.
**$p < .01$.

such subordinate clauses. After the definition was checked again, there was no diversion of judgment based on the definition.

In order to count the number of error-free clauses and error-free AS-units, evaluation of accuracy was done as follows. The criteria of errors were modified from Polio (1997) and McCarthy & Carter (2001). There were three main differences in the criteria in Study 2D from those used in Study 1B: (a) use of spoken grammar, (b) use of the criteria of expert users, and (c) treatment of discourse errors. First, Study 1C evaluated errors with the criteria of written grammar. However, differences between written grammar and spoken grammar have been found in research (e.g., Carter, 1998; McCarthy, 1998; McCarthy & Carter, 2001) using the CANCODE (Cambridge and Nottingham Corpus of Discourse in English) spoken corpus. It was judged that the criteria of spoken grammar should be introduced to the rating of spoken language in this study. The CANCODE corpus consists of "five million words of transcribed

conversations (in a variety of settings including private homes, shops, offices and other public places, and educational institutions [though informal settings] across the islands of Britain and Ireland, with a wide demographic spread" (McCarthy & Carter, 2001, p. 52). According to Carter (1998), the corpus "illustrates standard British spoken English (p. 43). Therefore, Study 2D used spoken grammar in error judgments. The definition and examples of spoken grammar were presented to raters before the rating. They made their judgment based on the criteria of whether the transcribed utterances were accurate and appropriate in terms of the tasks when they were produced orally. The second difference between Studies 1B and 2D was that Study 1B used the criteria of native speakers but Study 2D used those of expert users because there was doubt whether the criteria of native speakers should be used. In Study 2D, the perspective of world Englishes (e.g., Bolton, 2004) and expert users (McCarthy & Carter, 2001, p. 71) were considered important. The expert users included native speakers and proficient learners in the current study, and the expert users of English were used as raters. The third difference was that Study 1B did not consider discourse errors as errors, but in Study 2D, when there were discourse errors in terms of cohesion and coherence, the utterances were judged as having errors.

After the criteria were made, errors in the transcripts from all the five tasks were judged as follows. First, one third of the participants (n = 84) were evaluated by four raters. Raters A and B were proficient learners, whereas Raters C and D were native speakers of English (see Appendix 11.4 for their description and 11.5 for the criteria). They underlined the places where they found errors on the transcripts.

As seen in Table 11.6, the reliability was high overall and the ratings between the raters did not differ substantially either using four raters or two raters overall. The exception was Rater B, and when he and other raters were compared, the reliability

was low in some tasks. One of the reasons for the low reliability may be the task effects and Rater B's leniency, but the reasons need to be explored further. The results suggest that when the perspective of expert users and spoken grammar is taken into account, the ratings are similar overall but that in some cases, the reliability is lower, and proficient learners and native speakers may evaluate errors rather differently. However, this conclusion may be limited to the case of evaluation of error-free AS-units and clauses, the criteria and raters used in Study 2D.

Table 11.6
*Interrater Reliability (Cronbach's Alpha) of the Number of Error-Free AS-Units and Error-Free Clauses (*n = *84)*

	Task 1	Task 2	Task 3	Task 4	Task 5
Four Raters					
Error-free AS-units	.91	.85	.92	.85	.91
Error-free clauses	.91	.88	.93	.86	.93
Raters A & B					
Error-free AS-units	.89	.55	.85	.71	.86
Error-free clauses	.89	.64	.86	.69	.84
Raters A & C					
Error-free AS-units	.83	.86	.90	.83	.88
Error-free clauses	.83	.90	.91	.82	.89
Raters A & D					
Error-free AS-units	.89	.87	.88	.83	.85
Error-free clauses	.88	.90	.90	.84	.88
Raters B & C					
Error-free AS-units	.87	.46	.78	.74	.88
Error-free clauses	.86	.58	.79	.74	.86
Raters B & D					
Error-free AS-units	.76	.59	.79	.60	.77
Error-free clauses	.76	.64	.82	.62	.82
Raters C & D					
Error-free AS-units	.76	.91	.89	.74	.81
Error-free clauses	.76	.93	.90	.74	.87

Since the internal consistency was high for the two raters except for some tasks in which Rater B was involved, when two out of the four raters decided that there were errors in AS-units or clauses, these were judged as containing errors. Since the results suggest that the rating criteria of native speakers and proficient learners of English are

not very different, for the rest (two-thirds) of the participants, evaluation was made by two raters (Raters A and C), whose reliability was high enough in the rating of one-third of the participants, and the decision that the two agreed on (i.e., either existence or non-existence of errors) was used for the analysis. The reliability was very high between the two raters, as can be seen in Table 11.7.

Table 11.7
*Interrater Reliability (Cronbach's Alpha) of the Number of Error-Free AS-Units and Error-Free Clauses (*n = *159)*

	Task 1	Task 2	Task 3	Task 4	Task 5
Raters A & B					
Error-free AS-units	.94	.98	.95	.91	.95
Error-free clauses	.94	.98	.95	.92	.94

When the two raters disagreed, two steps were taken. First, the parts of the utterances about which they disagreed were analyzed, and if the phrases contained the same aspects as those rated in the previous rating (i.e., those that existed among the utterances in one-third of the participants), the decision of the previous rating was used for the judgment. Second, when one-third of the participants did not include the same aspects that two-thirds had, another rater (Rater B) was asked to rate. When two out of the three agreed, their decision was considered to be final. Although it was preferable to ask Rater D, it was difficult because of the time constraints.

The judgment of lexical word types was conducted as follows. First, a list of all the types used by all the participants was made. Second, all the types were judged by the author and another rater (who were both majoring in language testing) as to whether they belonged to (a) lexical words, (b) grammatical words, or (c) words that can belong to either lexical or grammatical words, depending on the context. There was a high agreement ratio (99%). Third, the two raters judged the words that can

belong to either lexical or grammatical words, depending on the context (2,071 types from 244 test takers). When the words to be judged were the last words and lacked sufficient context, and there was a difficulty in judging whether they were lexical or grammatical, the words that were expected to come next were imagined and judged. As a result, the agreement ratio was high (91%). When there were disagreements, the raters discussed them and came to an agreement, which was utilized as the final coding (see Appendix 11.3).

The evaluation of the number of tokens, the number of word types, the number of word lemmas, the number of lexical word types, and the number of sophisticated word types was done using the same criteria as those used in section 10.2.3.2.

After the coding, the values of speaking performance measures were computed.

11.2.3.3 Analysis Procedures

In order to investigate the Main Hypothesis 1 and the Hypotheses for Validity 3 and 4, principal components analysis (PCA; see section 2.6.4.2 for the details) was used. The Hypotheses for Validity 1 and 2 were examined through questionnaire items. In order to investigate the Hypothesis for Validity 5, simple correlations were used. The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1). When correlation coefficients were compared and one coefficient was higher than the other in a predicted direction, hypotheses were considered to be partially met. When the effect sizes (q) of differences between the correlation coefficients were medium (q = .30 to .49) or strong (q = .50 or more; Cohen, 1988; see section 2.6.4.1), the hypotheses were considered to be confirmed. The Hypothesis for Validity 6 was analyzed based on a professional judgment. The Hypothesis for Validity 7 was examined using simple

correlations or Cronbach's alpha. The Main Hypothesis 2 was tested using results of seven Hypotheses for Validity.

11.3 Results and Discussion

11.3.1 Assumption Check of Principal Components Analysis and Descriptive Statistics of the Speaking Performance Measures

In this section, assumptions of conducting PCA were examined and then descriptive statistics of the Speaking Performance Measures are reported. There are seven assumptions to be checked before PCA (Tabachnick & Fidell, 2001, pp. 588-590, 628-233): (a) sample size and missing data, (b) normality, (c) linearity, (d) absence of outliers among cases, (e) absence of singularity, (f) factorability of correlation matrices, and (g) absence of outliers among variables. Before the analyses with regard to the assumptions, two measures (i.e., Task 2 SC2 and Task 2 SC3) were excluded due to extreme kurtosis (both 244.00). The extreme kurtosis occurred because no test takers except for one produced subordinate clauses in Task 2, and all the values of SC2 and SC3 in Task 2 were 0.00 except for one person.

First, one assumption of PCA is that sample size (i.e., the number of participants used) is sufficiently large for analysis to be done. According to Comrey and Lee (1992), a sample size of 244 was a fair number (50 as very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1,000 as excellent; p. 217). However, the later analysis found "several high loading marker variables (> .80)," which "are highly correlated with one and only one factor, and load on it regardless of extraction or rotation technique" (Tabachnick & Fidell, p. 587). Since, if there are marker variables, the analyses "do not require such large sample sizes (about 150 cases should be sufficient) as solutions with lower loadings and/or fewer marker variables" (Guadagnoli & Velicer,

1988, as cited in Tabachnick & Fidell, p. 588), 244 were considered sufficient. Another assumption of PCA is that there should be no missing data. The data in Study 2D did not have any missing data regarding the 244 test takers.

Second, another assumption of PCA is multivariate normality and that "each variable and all linear combinations of the variables" should be normally distributed (Tabachnick & Fidell, 2001, p. 72). Multivariate normality is needed when "statistical inference is used to determine the number of factors" (Tabachnick & Fidell, p. 588). Since this study did not use statistical tests in deciding on the number of components, this assumption was not necessary.

Third, PCA assumes linearity and if linearity is satisfied, "there is a straight-line relationship between two variables" (Tabachnick & Fidell, 2001, p. 77). Since there were many pairs of variables, only pairs of either variable with strong skewness were examined through scatterplots, based on Tabachnick and Fidell's (p. 79) procedure. As a result, there was no evidence of curvilinearity, and the assumption of linearity was met.

Fourth, concerning the absence of outliers among cases, both univariate outliers and multivariate outliers were checked. As for the univariate outliers, extreme values, histograms, and boxplots were obtained using the "SPSS Explore" function. When the following two conditions were met, test takers were considered outliers. Firstly, test takers whose values belonged to areas in "more than 1.5 box-lengths from 25th or 75th percentile" in the boxplots (SPSS, 1999, p. 41) were detected. Second, the histograms show that out of those who satisfied the first condition, those who were "unattached to the rest of the distribution" (Tabachnick & Fidell, 2001, p. 68) were detected. Such outliers could also be detected with "extreme values" in the SPSS output. In order to detect multivariate outliers, Mahalanobis Distance was used. It should be noted that

Mahalanobis Distance has been found to be "not perfectly reliable" but that "alternative methods are computationally challenging and currently unavailable in statistical packages" (Tabachnick & Fidell, p. 69). Therefore, it was decided not to exclude all multivariate outliers detected through Mahalanobis Distance but to exclude only very extreme multivariate outliers. These "very extreme" multivariate outliers were decided based on the same method as that used in searching for the univariate outliers. As a result, 19 univariate and multivariate outliers were found in total, and 225 test takers (i.e., 244 – 19) were used for the subsequent analysis.

Fifth, PCA assumes the absence of singularity and if singularity is satisfied, "one of the variables is" not "a combination of two or more of the other variables" Tabachnick & Fidell, 2001, p. 83). In other words, redundant variables should not exist in PCA. In the data used in Study 2D, there was no singularity in the data.

Sixth, factorability of correlation matrices is the assumption that variables can be combined into factors, and as for Kaiser's measure of sampling adequacy ("Kaiser-Meyer-Olkin measure of sampling adequacy" in the SPSS output), values of .6 and above are required for good PCA (Tabachnick & Fidell, 2001, p. 589). This assumption was checked after each PCA.

Seventh, the last assumption regarding the absence of outliers among variables is that there are no "variables that are unrelated to others" (Tabachnick & Fidell, 2001, p. 589). In Study 2D, a component defined by only one variable was decided not to be interpreted.

After the assumptions were checked, 225 students were retained for the analysis. Table 11.8 shows means and standard deviations for the basic data and the Speaking Performance Measures for each task. The results indicate that the number of subordinate clauses and sophisticated word types with the criterion of 1,001 to 8,000

Table 11.8

Descriptive Statistics of the Basic Data and the Speaking Performance Measures

	Task 1		*Task 2*		*Task 3*		*Task 4*		*Task 5*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Tokens	37.24	13.68	29.80	12.99	25.49	11.06	28.23	11.58	30.23	11.99
AS-unit	6.40	2.17	4.84	1.64	4.15	1.70	4.92	1.87	4.42	1.59
Clauses	7.07	2.65	4.35	1.91	4.42	1.93	4.85	2.24	4.06	1.78
Dysfluency markers	3.76	3.72	4.62	4.32	4.10	4.15	3.51	3.64	4.83	4.24
Error-free clauses	5.35	2.61	1.11	1.11	2.66	1.69	2.80	1.97	1.48	1.40
Error-free AS-units	4.68	2.14	1.24	1.09	2.37	1.49	2.73	1.73	1.48	1.22
Subordinate clauses	0.72	1.04	0.00	0.00	0.35	0.61	0.24	0.58	0.19	0.57
Types	25.48	8.11	16.06	4.71	17.47	6.21	19.25	6.14	18.64	5.73
Lexical word types	16.04	5.20	9.18	2.72	9.60	3.55	10.79	3.81	9.48	3.48
Sophisticated word types (J1001)	1.71	1.41	1.15	0.86	2.24	1.21	2.38	1.05	2.17	1.21
Sophisticated word types (J101)	11.99	4.26	9.77	3.02	8.92	3.26	9.68	3.42	9.09	3.30
F1	49.65	18.25	39.73	17.32	33.99	14.74	37.64	15.44	40.30	15.99
F3	9.42	3.54	5.80	2.54	5.90	2.57	6.47	2.99	5.42	2.38
F4	5.01	4.97	6.16	5.76	5.46	5.54	4.68	4.85	6.44	5.65
F5	0.11	0.12	0.16	0.14	0.17	0.17	0.12	0.13	0.16	0.14
F6	0.62	0.68	0.97	0.91	1.02	0.98	0.75	0.81	1.17	1.11
A1	0.75	0.20	0.26	0.25	0.59	0.27	0.54	0.26	0.34	0.28
A2	0.73	0.20	0.26	0.21	0.57	0.27	0.54	0.24	0.34	0.24
SC1	1.10	0.16	0.89	0.21	1.07	0.21	0.98	0.18	0.93	0.28
SC2	0.11	0.15	0.00	0.00	0.09	0.18	0.04	0.10	0.05	0.17
SC3	0.08	0.11	0.00	0.00	0.06	0.11	0.03	0.08	0.03	0.10
SC4	5.84	1.14	6.15	1.74	6.26	1.44	5.84	1.38	6.93	1.81
LC1	4.15	0.65	2.98	0.41	3.45	0.56	3.63	0.53	3.40	0.53
LC2	4.10	0.91	3.47	0.92	4.48	1.28	4.15	1.12	4.42	1.22
LC3	3.77	0.87	4.35	2.38	4.28	1.40	4.46	1.81	5.27	2.68
LC4	0.44	0.07	0.34	0.11	0.40	0.10	0.40	0.08	0.33	0.10
LC5	2.62	0.47	1.72	0.35	1.91	0.41	2.03	0.41	1.73	0.45
LC6	2.58	0.60	2.00	0.60	2.48	0.82	2.30	0.64	2.22	0.71
LC7	2.38	0.59	2.53	1.50	2.38	0.94	2.51	1.28	2.69	1.62
LC8	0.63	0.08	0.58	0.10	0.56	0.10	0.56	0.08	0.51	0.11
J1001										
LC11	0.05	0.04	0.04	0.04	0.10	0.05	0.09	0.05	0.08	0.04
LC12	0.28	0.22	0.21	0.16	0.45	0.21	0.46	0.19	0.40	0.20
LC13	0.27	0.23	0.24	0.17	0.59	0.36	0.53	0.27	0.51	0.28
LC14	0.25	0.21	0.30	0.28	0.57	0.36	0.57	0.33	0.61	0.41
LC15	0.07	0.05	0.07	0.05	0.13	0.06	0.13	0.06	0.12	0.06
J101										
LC11	0.33	0.06	0.36	0.11	0.37	0.10	0.36	0.08	0.32	0.09
LC12	1.95	0.45	1.83	0.38	1.78	0.38	1.82	0.38	1.66	0.42
LC13	1.93	0.59	2.12	0.63	2.31	0.79	2.06	0.58	2.13	0.62
LC14	1.78	0.55	2.69	1.66	2.22	0.87	2.26	1.16	2.63	1.74
LC15	0.47	0.08	0.61	0.10	0.52	0.10	0.50	0.08	0.49	0.11

Note. $n = 225$. Task 2 SC2 and Task 2 SC3 were deleted for the subsequent analysis. See Appendix 11.6.

words in the JACET8000 (J1001; JACET Basic Word Revision Committee, 2003) was small (M = 0.00 to 0.72 for subordinate clauses; M = 1.15 to 2.38 for sophisticated word types with J1001) as the participants were beginning level learners.

When means were compared across tasks, differences between Task 1 and the other four tasks could be observed. In Task 1, all the basic data (e.g., the number of tokens) had the largest values except for the number of dysfluency markers and sophisticated word types (J1001). Moreover, values of the Speaking Performance Measures tended to be higher or lower in Task 1 than in the four other tasks. A peculiar tendency was seen in F1, F3, A1, A2, SC2, SC3, LC3, LC4, LC5, LC6, LC8, LC14J1001, LC12J101, and LC15J101. The result was that Task 1 elicited utterances with more fluency, more accuracy, more syntactic complexity, and more lexical complexity in terms of lexical diversity (from LC1 to LC4) and lexical density (from LC5 to LC8). As for lexical sophistication (from LC11 to LC15), the results were different between the measures and the definition of sophisticated words used. That is, the means of Task 1 LC14J1001 and Task 1 LC15J101 were higher than the same measures in other tasks, and the mean of Task 1 LC12J101 was lower than the same measures in other tasks. Since Task 1 elicited a greater number of words and better performance overall (except lexical sophistication), it can be considered easier than the other four tasks. This may be because it elicited more use of formulaic expressions.[11.2]

11.3.2 Investigating the Hypotheses for Validity 1 and 2

Analyses were conducted in this section using the Speaking Test Questionnaire. Table 11.9 shows test takers' reactions toward the Productive Vocabulary Knowledge Test (Productive VKT), after removal of the results of test takers who chose the option

"I don't know" in the questionnaire (see Table 11.10 for the three items).

Table 11.9
Test Takers' Responses Concerning the Speaking Test

	Q5	Q6	Q7	Q8	Q10	Q17
	(*n* =217)	(*n* = 238)	(*n* = 235)	(*n* = 216)	(*n* = 237)	(*n* = 223)
M (*SD*)	3.90 (1.24)	4.28 (1.04)	3.69 (1.21)	3.94 (1.27)	3.10 (1.50)	3.63 (1.33)

Table 11.10
Items on the Speaking Test Questionnaire

Part 2. Please choose the number that describes your feelings the most with regard to the test you have just taken.

---

0. I don't know. 1. No, I don't think so at all.
2. No, I don't think so very much. 3. Neither yes nor no.
4. Yes, I think so a little. 5. Yes, I think so very much.

(2-5) Do you think that this test measures English speaking ability?
(2-6) Was the tape sound clear?
(2-7) Was the test instruction easy to understand?
(2-8) Was operating the recorder easy?
(2-10) Were the sound and people nearby distracting?
(2-17) Do you want to study English more from now on?

Regarding questions about face validity (Q5), clarity of tape sound (Q6), clarity of instructions (Q7), ease of operating the recorder (Q8), and effects of the Speaking Test on future English learning (Q17), means were above 3.5 on the 1 to 5 Likert scale, with 5 indicating a positive reaction. Regarding a question about whether sound and people nearby were distracting (Q10), means were about the average (3.10) on the 1 to 5 Likert scale, with 5 indicating that the sound and people nearby were distracting. This suggests that test takers perceived the test as good to some degree in terms of face validity, testing clarity, ease of test procedures, and potential washback effects. Therefore, the Hypotheses for Validity 1 and 2 (There is a positive response from test takers concerning the quality of the Speaking Test and the effect of the Speaking Test on learning.) were confirmed. As for the distracting nature of sound and people nearby,

the results were neutral, but obtaining such weakly positive results might indicate that some improvements may be necessary for future research, although the distracting nature of sound and people nearby seems to be a problem that always occurs when tape-mediated speaking tests are administered in a group.

11.3.3 Investigating the Main Hypothesis 1: Meaning of the AS-Unit Length

In the Main Hypothesis 1, it was hypothesized that if the AS-unit length (i.e., the number of tokens per AS-unit, or SC4) were a syntactic complexity measure rather than a fluency measure, it would load on components that syntactic complexity measures loaded on and not on components that fluency measures loaded on.

PCA was performed in order to examine relationships between the AS-unit length and the Speaking Performance Measures of fluency and syntactic complexity. The value of Kaiser-Meyer-Olkin measure of sampling adequacy was high enough (0.67), which indicates that the assumption of factorability of correlation matrices was met (see section 11.3.1). Twelve components were extracted, and 11 components were determined to be of use (see Tables 11.11 and 11.12), which accounted for 82.78% of the total variance. They were named as follows: (C1) Speaking speed, (C2) Task 1 Syntactic complexity (on which Task 1 SC4 loaded), (C3) Task 5 Syntactic complexity (on which Task 5 SC4 loaded), (C4) Task 3 Syntactic complexity, (C5) Task 1 Dysfluency, (C6) Task 3 Dysfluency, (C7) Task 4 Dysfluency, (C8) Task 2 Dysfluency, (C9) Task 5 Dysfluency, (C10) Task 4 Syntactic complexity, and (C11) AS-unit length (on which Task 1 SC4, Task 2 SC4, Task 3 SC4, and Task 4 SC4 loaded).

Table 11.11 illustrates that SC4 did not load on any fluency components. SC4 in two tasks (Tasks 1 and 5) did load on the syntactic complexity components (Components 2 and 3), and SC4 in four tasks (Tasks 1, 2, 3, and 4) loaded on the

Table 11.11

Factor Loadings in the Pattern Matrix From Principal Components Analysis of Fluency and Syntactic Complexity Measures

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	Com
T1F1	0.91	0.09	-0.06	-0.03	-0.15	0.08	0.01	-0.04	-0.02	0.04	0.01	0.79
T1F3	0.81	0.29	-0.05	-0.01	-0.19	0.07	0.10	-0.11	-0.03	0.06	-0.13	0.76
T1F4	0.24	-0.03	-0.01	-0.04	0.84	0.07	0.05	-0.01	0.01	0.03	0.00	0.92
T1F5	-0.14	0.01	-0.01	0.01	1.01	-0.01	0.00	-0.02	-0.03	0.01	-0.05	0.96
T1F6	-0.10	0.04	0.01	-0.02	1.00	0.00	0.02	0.01	-0.03	0.01	0.04	0.98
T1SC1	-0.01	0.98	0.04	0.00	0.01	0.00	0.03	0.00	-0.03	0.01	0.04	0.95
T1SC2	-0.01	0.99	0.02	0.01	0.01	0.00	0.02	0.00	-0.01	-0.02	0.02	0.97
T1SC3	-0.02	0.99	0.02	0.00	0.01	0.00	0.01	0.00	-0.01	0.00	0.01	0.97
T1SC4	0.21	0.31	-0.02	-0.02	0.06	0.03	-0.16	0.12	-0.03	-0.05	0.38	0.38
T2F1	0.80	-0.05	-0.04	0.00	0.05	0.11	0.00	-0.19	0.00	-0.05	0.18	0.83
T2F3	0.62	-0.02	0.06	0.03	0.05	0.07	0.01	-0.15	-0.03	-0.09	-0.17	0.88
T2F4	0.22	-0.01	0.00	0.00	0.02	0.08	0.07	0.82	-0.02	-0.02	-0.01	0.93
T2F5	-0.12	-0.02	0.03	0.02	-0.02	0.00	0.00	1.02	0.02	-0.01	-0.14	0.95
T2F6	-0.04	0.03	-0.04	-0.03	-0.01	-0.01	-0.02	0.96	0.03	0.00	0.11	0.96
T2SC1	-0.10	0.13	-0.06	0.04	0.03	-0.06	-0.12	0.10	0.09	0.06	0.06	0.83
T2SC4	0.23	0.04	-0.24	-0.04	0.02	0.02	-0.10	0.02	0.09	0.12	0.64	0.59
T3F1	0.91	-0.03	-0.04	0.03	0.04	-0.31	0.10	0.07	0.12	-0.16	0.08	0.82
T3F3	0.91	-0.02	-0.04	0.24	0.02	-0.23	0.06	0.08	0.08	-0.07	-0.26	0.83
T3F4	0.21	0.00	-0.01	0.03	0.04	0.79	0.08	0.04	0.08	-0.06	-0.06	0.89
T3F5	-0.15	0.01	-0.02	0.01	0.00	1.01	-0.03	0.00	0.04	-0.02	-0.14	0.96
T3F6	-0.15	-0.01	-0.01	0.08	0.01	0.98	0.02	0.01	0.03	-0.04	0.04	0.96
T3SC1	0.01	-0.02	-0.01	0.95	-0.01	0.00	-0.01	0.04	-0.03	0.02	0.04	0.91
T3SC2	-0.03	0.00	-0.01	0.96	-0.02	0.05	-0.01	-0.01	-0.01	0.04	0.06	0.97
T3SC3	0.06	0.02	-0.01	0.95	-0.01	0.07	-0.01	-0.03	-0.01	0.03	0.03	0.95
T3SC4	-0.05	0.00	0.09	0.29	0.03	-0.15	0.06	-0.05	0.00	-0.14	0.78	0.75
T4F1	0.89	-0.02	0.02	-0.09	0.08	-0.07	-0.14	-0.03	0.12	-0.04	0.14	0.84
T4F3	0.80	-0.01	0.04	-0.05	0.02	-0.05	-0.22	0.03	0.15	0.24	-0.11	0.80
T4F4	0.25	0.01	0.05	0.00	0.10	0.02	0.78	0.06	0.06	0.02	0.02	0.92
T4F5	-0.09	0.02	-0.03	0.00	0.01	0.01	0.97	0.01	0.02	0.04	-0.02	0.93
T4F6	-0.09	0.02	-0.02	-0.03	-0.01	0.01	0.98	-0.01	0.02	0.04	0.10	0.97
T4SC1	-0.04	-0.06	0.04	0.02	-0.18	-0.04	0.07	0.06	0.07	0.75	0.07	0.67
T4SC2	-0.01	0.01	-0.01	0.05	0.08	-0.04	0.03	-0.03	-0.04	0.97	-0.04	0.93
T4SC3	0.02	0.01	-0.01	0.03	0.10	-0.03	0.02	-0.04	-0.05	0.96	-0.03	0.92
T4SC4	-0.07	-0.02	0.04	-0.12	-0.10	-0.05	0.28	-0.08	-0.01	0.04	0.60	0.57
T5F1	0.79	-0.16	-0.01	-0.02	-0.02	0.11	0.02	0.08	-0.27	0.04	0.19	0.79
T5F3	0.64	-0.08	0.23	-0.06	-0.06	0.09	0.01	0.13	-0.38	0.03	0.08	0.71
T5F4	0.26	-0.04	0.02	-0.07	-0.01	0.12	0.05	0.03	0.72	0.05	0.04	0.87
T5F5	-0.10	-0.01	-0.06	-0.03	-0.06	0.00	0.05	0.01	0.99	-0.04	0.01	0.93
T5F6	-0.06	-0.01	0.13	0.02	0.01	0.07	-0.01	0.01	0.92	-0.01	0.03	0.96
T5SC1	-0.01	0.03	0.83	-0.04	-0.04	0.04	-0.10	0.05	-0.01	0.04	0.05	0.77
T5SC2	-0.03	0.04	1.00	-0.01	0.01	-0.06	0.03	-0.02	0.02	-0.02	-0.09	0.92
T5SC3	-0.01	0.04	0.99	-0.01	0.02	-0.05	0.06	-0.01	0.00	-0.02	-0.09	0.90
T5SC4	0.10	-0.06	0.48	0.08	0.01	0.11	-0.11	-0.04	0.19	0.08	0.25	0.61
Cumu	25.12	37.29	44.66	51.45	57.47	62.75	67.70	72.14	76.47	79.71	82.78	
RSSL	9.29	3.52	4.40	3.25	4.28	5.06	4.36	4.57	4.86	4.28	3.82	

Note. Extraction Method: Principal Components Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 7 iterations. C = Component; Com = Communalities; RSSL = Rotation Sums of Squared Loadings; Cumu = Cumulative %.

Table 11.12
Correlations Between Components of Fluency and Syntactic Complexity in the Component Correlation Matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
C1	--	.17*	.31**	.08	.19**	.27**	.21**	.30**	.26**	.41**	.34**
C2		--	.00	.09	-.11	-.04	-.07	.04	.03	.16*	.00
C3			--	.03	.15*	.18**	.13*	.16*	.20**	.17**	.31**
C4				--	-.08	-.02	.02	-.06	.06	.12	.19**
C5					--	.30**	.30**	.27**	.29**	.05	.13*
C6						--	.36**	.29**	.44**	.08	.17*
C7							--	.23**	.28**	.10	.15*
C8								--	.30**	.08	.26**
C9									--	.19**	.14*
C10										--	.15*
C11											--

Note. $n = 225$. C = Component.
*$p < .05$. **$p < .01$.

AS-unit length component (Component 11). This result suggests that AS-unit length (SC4) is not a fluency measure and that, depending on tasks, it is related more to syntactic complexity measures in some cases, but in others, it is not related to these measures and instead, SC4 across tasks accumulates and creates one component. Therefore, if it is necessary to decide on whether AS-unit length reflects fluency or syntactic complexity, it can be interpreted that it belongs to syntactic complexity, not fluency, which supports the Main Hypothesis 1. This interpretation of AS-unit length as syntactic complexity (SC4) is used and F2 was excluded in the subsequent analysis. One thing to be mentioned is that even when SC4 did load on the syntactic complexity components in Tasks 1 and 5, loading values of SC4 (the number of tokens per AS-unit) were rather lower than other syntactic complexity measures (i.e., SC1, SC2, and SC3), which are the measures using the number of clauses and subordinate clauses in the numerators. Thus, SC4 and the other syntactic complexity measures tend to assess rather different elements of syntactic complexity.

These results were congruent with the results of Study 1B (see section 5.3.2) and

the empirical results from the previous studies (Mehnert, 1998; Takiguchi, 2003). The interpretation of the AS-unit length being a speaking performance measure of syntactic complexity seems to be tenable at least to learners at the novice level in Takiguchi (2003) and in this study, and probably also at the intermediate level in Mehnert (1998). In contrast, it seems that the results of the AS-unit length from Kimura (1995), Niwa (2000, as cited in Robinson, 2005), and Robinson (2001) need to be reinterpreted from fluency to syntactic complexity because the participants seem to belong to the novice and intermediate levels. One caution should be made that the present result only concerns the interpretation of the AS-unit length and probably this result may be extended to the T-unit length or the c-unit length because these units are very similar to the AS-units. However, this result may not be applicable to the length of other units, especially units mainly related to pauses or dysfluency markers, such as pausally defined units (e.g., Lennon, 1990b). There is a possibility that "the number of words per pausally defined unit" reflects fluency because in Mehnert (1998), "the number of syllables per pausally defined unit" loaded on the general fluency factor. In addition, the meaning of the T-unit length in writing as in Wolfe-Quintero et al. (1998) also needs to be investigated in future research. What should be emphasized at this point is that the argument by Wolfe-Quintero et al. (p. 15) that a measure belongs to a certain construct that is reflected in the numerator of the measure and that the denominator "indicates only the context for the measure …, not the type of measure it is" is not tenable at least for beginning level learners' speaking. The interpretation of the AS-unit length (i.e., "the number of tokens" divided by "the number of AS-units") as syntactic complexity implies that the denominator "the number of AS-units" affects the interpretation of syntactic complexity more than the numerator "the number of tokens."

11.3.4 Investigating the Hypotheses for Validity 3 and 4

In order to examine the internal structures and generalizability of the Speaking Performance Measures, PCA was conducted. After the analyses were done on each construct (i.e., fluency, accuracy, syntactic complexity, and lexical complex), the Hypotheses for Validity 3 and 4 were tested.[11.3]

First, for fluency measures, PCA was performed in order to examine relationships between fluency measures in five tasks. The value of Kaiser-Meyer-Olkin measure of sampling adequacy was high enough (0.81), which indicates that the assumption of factorability of correlation matrices was met (see section 11.3.1). As can be seen in Table 11.13, six components were extracted, and all the 6 components were determined to be of use, which accounted for 84.83% of the total variance. Component scores were saved and used for the regression analyses in the next chapter (Study 2E). They were named as follows: (PF1) Speaking speed, (PF2) Task 1 Dysfluency, (PF3) Task 2 Dysfluency, (PF4) Task 4 Dysfluency, (PF5) Task 3 Dysfluency, and (PF6) Task 5 Dysfluency. The present results of deriving two types of fluency, one for speaking speed and the other for dysfluency, are consistent with previous studies in which fluency measures were divided into two categories: “temporal” measures and measures of “a degree of freedom from various dysfluency markers” (Lennon, 1990b, p. 403).

As seen in Table 11.13, six components appeared with F1 and F3 loading on the PF1, and F4, F5, and F6 loading on the other components (PF2 to PF6). Even when the measures were derived from the same tasks, they did not always belong to the same component. The five fluency measures were divided into two types (i.e., F1 and F3 vs. F4 to F6) and interpreted as “speaking speed” and “dysfluency.” Therefore, the Hypothesis for Validity 3 (The Speaking Performance Measures of fluency on the

Table 11.13

Factor Loadings in the Pattern Matrix From Principal Components Analysis of Fluency Measures

	PF1	PF2	PF3	PF4	PF5	PF6	Com
T1F1	0.86	-0.22	-0.02	0.03	0.05	0.02	0.74
T1F3	0.80	-0.29	-0.09	0.06	0.05	0.02	0.63
T1F4	0.21	0.84	0.00	0.04	0.05	0.03	0.90
T1F5	-0.14	1.02	-0.03	-0.02	-0.03	-0.01	0.96
T1F6	-0.09	1.00	0.02	0.01	-0.01	-0.01	0.97
T2F1	0.88	0.09	-0.17	0.00	0.12	-0.02	0.78
T2F3	0.79	0.14	-0.22	-0.05	0.12	-0.06	0.61
T2F4	0.20	0.03	0.83	0.05	0.08	-0.01	0.93
T2F5	-0.17	-0.01	1.01	-0.04	0.00	0.03	0.94
T2F6	-0.05	-0.01	1.00	-0.02	-0.02	0.03	0.96
T3F1	0.81	0.05	0.08	0.09	-0.28	0.10	0.78
T3F3	0.82	0.01	0.00	0.06	-0.23	0.10	0.71
T3F4	0.21	0.03	0.04	0.07	0.78	0.08	0.89
T3F5	-0.13	-0.01	-0.01	-0.05	0.99	0.05	0.95
T3F6	-0.09	-0.01	0.02	0.01	0.97	0.04	0.96
T4F1	0.87	0.09	0.03	-0.14	-0.07	0.13	0.81
T4F3	0.80	-0.01	0.01	-0.18	-0.10	0.20	0.69
T4F4	0.22	0.07	0.04	0.82	0.00	0.05	0.93
T4F5	-0.13	-0.02	-0.04	1.02	-0.01	0.00	0.97
T4F6	-0.08	-0.01	-0.02	1.01	0.00	-0.01	0.97
T5F1	0.83	0.02	0.13	0.03	0.10	-0.28	0.75
T5F3	0.77	0.00	0.16	-0.01	0.09	-0.34	0.64
T5F4	0.26	0.01	0.05	0.03	0.12	0.73	0.88
T5F5	-0.14	-0.03	0.01	0.02	0.02	0.99	0.93
T5F6	-0.03	0.03	0.02	-0.01	0.07	0.93	0.94
Cumu	36.37	54.86	63.55	71.80	79.06	84.83	
RSSL	7.89	4.18	4.47	4.53	4.51	4.64	

Note. Extraction Method: Principal Components Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 6 iterations. C = Component; Com = Communalities; RSSL = Rotation Sums of Squared Loadings; Cumu = Cumulative %.

same task load on the same component) was only partially supported. Furthermore, F1 and F3 of all the five tasks loaded on the PF1 (Speaking speed), which indicates that F1, F3, and "speaking speed" tend to have generalizability across tasks. On the other hand, each component of F4, F5, and F6 (the proportion of dysfluency markers) appeared for each task, so it was found that F4, F5, and F6, and "dysfluency" tend to lack strong generalizability across tasks and that their results are likely to change, depending on tasks. Therefore, the Hypothesis for Validity 4 (Each measure of fluency

across tasks load on the same component) was supported for F1 and F3, but not supported for F4, F5, and F6.

Second, for accuracy measures, PCA was performed in order to examine relationships between accuracy measures in the five tasks. The value of Kaiser-Meyer-Olkin measure of sampling adequacy was high enough (0.61), which indicates that the assumption of factorability of correlation matrices was met (see section 11.3.1). Five components were extracted, and all the five components were determined to be of use, which accounted for 95.56% of the total variance. Component scores were saved and used for the regression analyses in the next chapter (Study 2E). They were named as follows: (PA1) Task 1 Accuracy, (PA2) Task 3 Accuracy, (PA3) Task 4 Accuracy, (PA4) Task 5 Accuracy, and (PA5) Task 2 Accuracy.

As seen in Table 11.14, five components appeared with A1 and A2 loading on components for each task. Therefore, the Hypothesis for Validity 3 (The Speaking Performance Measures of accuracy on the same task load on the same component) was

Table 11.14
Factor Loadings in the Pattern Matrix From Principal Components Analysis of Accuracy Measures

	PA1	PA2	PA3	PA4	PA5	Com
T1A1	0.99	0.02	-0.01	0.01	0.00	37.51
T1A2	0.99	-0.02	0.02	-0.01	0.00	55.97
T2A1	0.01	0.01	-0.02	-0.01	0.96	70.25
T2A2	0.00	0.00	0.02	0.01	0.96	83.64
T3A1	0.03	0.99	-0.01	0.00	-0.02	95.56
T3A2	-0.03	0.99	0.01	0.00	0.03	37.51
T4A1	0.01	0.02	0.97	0.02	-0.04	55.97
T4A2	-0.01	-0.02	0.98	-0.02	0.04	70.25
T5A1	-0.02	-0.04	0.02	0.98	0.00	83.64
T5A2	0.02	0.04	-0.02	0.96	0.00	95.56
Cumu	37.51	55.97	70.25	83.64	95.56	
RSSL	2.45	2.61	2.45	2.22	2.14	

Note. Extraction Method: Principal Components Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 5 iterations. C = Component; Com = Communalities; RSSL = Rotation Sums of Squared Loadings; Cumu = Cumulative %.

supported. Furthermore, each component of A1 and A2 appeared for each task, so A1 and A2 and "accuracy" tend to lack strong generalizability across tasks and their results are likely to change, depending on tasks. Therefore, the Hypothesis for Validity 4 (Each measure of accuracy across tasks load on the same component) was not supported for A1 and A2.

Third, for syntactic complexity measures, PCA was performed in order to examine relationships between syntactic complexity measures in the five tasks. The value of Kaiser-Meyer-Olkin measure of sampling adequacy was high enough (0.72), which indicates that the assumption of factorability of correlation matrices was met (see section 11.3.1). Five components were extracted, and all the 6 components were determined to be of use, which accounted for 75.39% of the total variance. Component scores were saved and used for the regression analyses in the next chapter (Study 2E). They were named as follows: (PSC1) Task 1 Syntactic complexity, (PSC2) Task 3 Syntactic complexity, (PSC3) Task 5 Syntactic complexity, (PSC4) Task 4 Syntactic complexity, (PSC5) AS-unit length and Task 2 Syntactic complexity.

As seen in Table 11.15, five components appeared with SC1, SC2, SC3, and SC4 loading on each component for Tasks 1, 3, and 5 (PSC1 to PSC3) and with SC1, SC2, and SC3 loading on PSC4 for Task 4, and with SC4 of all the five tasks and SC1 of Task 2 loading on PSC5. Therefore, the Hypothesis for Validity 3 (The Speaking Performance Measures of syntactic complexity on the same task load on the same component) was supported for Tasks 1, 2, 3, and 5 but partially supported for Task 4. Each component of SC1, SC2, and SC3 appeared for each task, so SC1, SC2, and SC3 tend to lack strong generalizability across tasks and their results are likely to change, depending on tasks. On the other hand, SC4 of all the five tasks loaded on the component of PSC5, which indicates that SC4 tends to have generalizability across

tasks. Therefore, the Hypothesis for Validity 4 (Each measure of syntactic complexity across tasks load on the same component) was supported for SC4 but not supported for SC1, SC2, and SC3.

Table 11.15
Factor Loadings in the Pattern Matrix From Principal Components Analysis of Syntactic Complexity Measures

	PSC1	PSC2	PSC3	PSC4	PSC5	Com
T1SC1	0.97	0.01	0.02	0.01	0.03	0.95
T1SC2	0.99	0.02	0.00	0.00	-0.01	0.97
T1SC3	0.98	0.01	0.00	0.00	-0.02	0.97
T1SC4	0.36	-0.03	0.05	-0.06	0.44	0.35
T2SC1	0.05	-0.10	-0.13	0.05	0.67	0.42
T2SC3	0.03	0.00	-0.12	0.09	0.76	0.57
T3SC1	0.00	0.96	-0.04	0.02	-0.06	0.90
T3SC2	0.01	0.99	-0.04	0.03	-0.06	0.96
T3SC3	0.03	0.97	-0.02	0.06	-0.05	0.94
T3SC4	-0.05	0.45	0.16	-0.25	0.40	0.51
T4SC1	-0.05	0.00	0.04	0.72	0.22	0.63
T4SC2	0.02	0.04	0.01	0.97	-0.04	0.94
T4SC3	0.02	0.02	0.01	0.96	-0.03	0.92
T4SC4	-0.13	-0.04	0.09	0.00	0.61	0.40
T5SC1	0.02	-0.07	0.80	0.02	0.13	0.72
T5SC2	0.02	-0.02	1.00	0.00	-0.18	0.93
T5SC3	0.03	-0.02	0.99	0.00	-0.16	0.91
T5SC4	-0.07	0.09	0.57	0.10	0.31	0.59
Cumu	24.31	40.38	55.62	67.20	75.39	
RSSL	3.16	3.30	3.24	2.79	2.55	

Note. Extraction Method: Principal Components Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 6 iterations. C = Component; Com = Communalities; RSSL = Rotation Sums of Squared Loadings; Cumu = Cumulative %.

Fourth, for lexical complexity measures, PCA was performed in order to examine relationships between syntactic complexity measures in the five tasks. The value of Kaiser-Meyer-Olkin measure of sampling adequacy was high enough (0.70), which indicates that the assumption of factorability of correlation matrices was met (see section 11.3.1). Twenty-one components were extracted, and the 12 components were determined to be of use, which accounted for 74.78% of the total variance. Component scores were saved and used for the regression analyses in the next chapter

(Study 2E).

As seen in Table 11.16, 12 components had complicated structures. For example, for Task 1, LC1, LC5, and LC12J101 loaded on the PLC1, and LC2, LC3, LC6, LC7, LC13J101, and LC14J101 loaded on the PLC2, and LC11J1001, LC12J1001, LC13J1001, LC14J1001, and LC15J1001 loaded on the PLC3. For Task 2, LC1 and LC4 loaded on the PLC1, and LC2, LC4, LC5, LC8, LC11J101, LC12J101, and LC15J101 loaded on the PLC6. Since some measures of the same task gathered but the other measures did not, the Hypothesis for Validity 3 (The Speaking Performance Measures of syntactic complexity on the same task load on the same component) was only partially supported for all the five tasks. Table 11.17 summarizes the names of the components of lexical complexity. Although Task 2 LC4 and Task 4 LC6 loaded on the first component, they were not included in the interpretation of the component due to the low loadings.

As for subcategories of lexical complexity, LC11J1001, LC12J1001, LC13J1001, LC14J1001, and LC15J1001 of the same task loaded on the same component and they did not mix with other measures except for Task 3 LC13J1001 and Task 3 LC14J1001 (PLC4). These results suggest that these measures and “Lexical sophistication of J1001” tend to assess rather different elements from “Lexical sophistication of J101,” whose criterion of sophisticated words is different.

Table 11.16
Factor Loadings in the Pattern Matrix From Principal Components Analysis of Lexical Complexity Measures

	PLC1	PLC2	PLC3	PLC4	PLC5	PLC6
T1LC1	0.55	0.21	0.05	0.11	0.03	0.13
T1LC2	0.07	0.98	0.05	0.03	0.04	-0.02
T1LC3	0.03	1.08	-0.08	-0.01	-0.02	0.03
T1LC4	-0.25	0.22	0.10	0.05	0.08	-0.08
T1LC5	0.40	0.03	0.00	0.05	-0.02	0.08
T1LC6	0.00	0.80	0.03	0.00	0.00	-0.03
T1LC7	-0.03	0.87	-0.09	-0.03	-0.05	0.01
T1LC8	-0.13	-0.22	-0.05	-0.05	-0.04	-0.05
T1LC11J1001	-0.01	-0.08	1.01	-0.01	0.00	0.01
T1LC12J1001	0.14	-0.10	0.98	0.00	-0.01	0.04
T1LC13J1001	0.05	0.14	0.95	-0.04	-0.02	0.01
T1LC14J1001	0.04	0.19	0.93	-0.05	-0.04	0.04
T1LC15J1001	0.04	-0.16	0.99	-0.03	-0.01	0.00
T1LC11J101	-0.18	0.26	0.14	0.06	0.09	-0.06
T1LC12J101	0.31	0.10	0.06	0.05	0.01	0.07
T1LC13J101	0.00	0.70	0.07	0.00	0.02	-0.02
T1LC14J101	-0.03	0.80	-0.03	-0.03	-0.02	0.02
T1LC15J101	-0.08	-0.07	0.03	-0.03	-0.01	-0.02
T2LC1	0.47	-0.07	0.02	-0.10	0.07	0.12
T2LC2	0.02	0.05	-0.07	0.00	-0.01	-0.32
T2LC3	0.06	0.00	0.05	0.02	0.02	-0.17
T2LC4	-0.32	0.03	0.03	0.04	0.06	0.61
T2LC5	0.14	0.00	0.03	-0.01	0.00	0.89
T2LC6	-0.12	0.05	-0.05	0.04	-0.05	0.29
T2LC7	0.00	0.00	0.05	0.05	-0.01	0.13
T2LC8	-0.20	0.07	0.02	0.07	-0.08	0.97
T2LC11J1001	-0.11	0.00	0.00	-0.03	0.05	0.00
T2LC12J1001	0.09	-0.04	-0.02	-0.01	0.00	0.09
T2LC13J1001	0.02	-0.04	-0.04	0.00	-0.03	-0.07
T2LC14J1001	0.04	0.02	-0.03	0.01	-0.03	-0.02
T2LC15J1001	0.01	-0.02	-0.03	0.01	-0.02	0.04
T2LC11J101	-0.27	-0.02	0.02	0.00	0.09	0.56
T2LC12J101	0.23	-0.07	0.02	-0.08	0.03	0.80
T2LC13J101	-0.05	0.01	-0.06	-0.01	-0.02	0.23
T2LC14J101	0.04	-0.02	0.05	0.02	0.01	0.09
T2LC15J101	-0.11	-0.02	0.01	-0.02	-0.04	0.96
T3LC1	0.83	-0.08	0.19	0.17	0.03	-0.06
T3LC2	0.17	-0.03	-0.02	0.90	-0.09	0.00
T3LC3	0.06	-0.03	0.05	1.09	0.09	-0.01
T3LC4	-0.08	-0.04	0.05	0.27	-0.04	-0.01
T3LC5	0.63	0.03	0.06	0.09	0.00	-0.01
T3LC6	0.12	0.02	-0.08	0.79	-0.10	0.01
T3LC7	0.04	0.02	-0.02	0.93	0.06	0.01
T3LC8	-0.05	0.09	-0.09	-0.06	-0.04	0.03
T3LC11J1001	-0.10	-0.03	0.04	-0.05	0.01	-0.01
T3LC12J1001	0.25	0.01	0.06	-0.17	0.03	-0.01
T3LC13J1001	0.00	-0.02	-0.02	0.31	-0.04	0.03
T3LC14J1001	-0.03	-0.01	0.01	0.41	0.03	0.00
T3LC15J1001	-0.08	0.02	-0.02	-0.24	0.00	0.01

(Table 11.16 continues)

(Table 11.16 continued)

	PLC1	PLC2	PLC3	PLC4	PLC5	PLC6
T3LC11J101	-0.24	-0.09	0.05	0.16	0.00	-0.02
T3LC12J101	0.42	-0.03	0.07	-0.03	0.06	-0.03
T3LC13J101	0.00	-0.01	-0.08	0.69	-0.06	0.00
T3LC14J101	-0.08	0.00	-0.02	0.86	0.08	0.00
T3LC15J101	-0.27	0.01	-0.07	-0.18	0.02	0.02
T4LC1	0.85	-0.01	0.00	-0.01	-0.03	-0.13
T4LC2	0.13	0.00	0.01	0.04	-0.01	0.03
T4LC3	-0.18	0.03	0.06	0.07	0.00	-0.03
T4LC4	0.10	0.07	0.01	-0.01	0.04	0.05
T4LC5	0.83	0.05	-0.10	-0.04	-0.05	-0.03
T4LC6	0.31	0.03	-0.07	-0.03	-0.01	0.08
T4LC7	-0.08	0.03	0.01	0.04	-0.01	-0.02
T4LC8	0.30	0.08	-0.12	-0.03	-0.02	0.11
T4LC11J1001	-0.10	-0.01	0.02	0.02	0.02	0.00
T4LC12J1001	0.24	0.01	-0.01	-0.01	-0.01	-0.04
T4LC13J1001	0.01	-0.01	-0.01	-0.02	-0.03	0.01
T4LC14J1001	-0.14	0.01	0.02	0.02	-0.02	-0.01
T4LC15J1001	-0.05	0.01	-0.04	-0.01	-0.01	0.00
T4LC11J101	-0.11	-0.08	0.08	-0.02	0.05	-0.01
T4LC12J101	0.60	-0.09	-0.01	-0.02	-0.03	-0.10
T4LC13J101	0.17	-0.07	-0.02	-0.02	0.00	0.03
T4LC14J101	-0.15	-0.02	0.04	0.05	0.00	-0.03
T4LC15J101	0.03	-0.11	0.00	0.00	-0.03	0.01
T5LC1	0.61	0.02	-0.03	-0.01	0.05	0.06
T5LC2	0.05	0.03	0.01	0.01	-0.37	0.02
T5LC3	0.04	0.03	0.02	-0.02	-0.17	0.00
T5LC4	-0.10	0.07	0.09	0.02	0.83	-0.11
T5LC5	0.34	0.03	-0.02	0.00	0.78	-0.05
T5LC6	0.04	0.02	0.00	0.00	0.30	-0.08
T5LC7	0.06	0.04	0.02	0.01	0.20	-0.04
T5LC8	-0.03	0.03	0.00	0.01	0.98	-0.13
T5LC11J1	-0.11	0.03	0.06	-0.01	0.17	-0.02
T5LC12J1	0.12	-0.01	0.00	0.01	0.07	0.02
T5LC13J1	-0.01	-0.02	0.02	0.02	-0.13	0.01
T5LC14J1	0.00	-0.03	0.00	0.03	-0.13	-0.01
T5LC15J1	-0.07	-0.02	0.03	0.00	0.08	-0.03
T5LC11J101	-0.15	-0.01	0.02	0.01	0.79	0.06
T5LC12J101	0.29	-0.05	-0.10	-0.01	0.74	0.15
T5LC13J101	-0.02	-0.05	-0.06	-0.01	0.29	0.10
T5LC14J101	0.03	0.00	0.00	-0.01	0.15	0.04
T5LC15J101	-0.09	-0.09	-0.09	0.00	0.86	0.12
Cumu	12.75	22.53	31.28	38.50	44.66	50.50
RSSL	6.63	6.62	6.55	7.37	7.05	6.72

(Table 11.16 continues)

(Table 11.16 continued)

	PLC7	PLC8	PLC9	PLC10	PLC11	PLC12	Com
T1LC1	0.07	-0.01	0.00	-0.02	-0.04	0.05	0.77
T1LC2	0.00	0.03	-0.01	-0.03	-0.02	0.04	0.91
T1LC3	-0.02	-0.02	0.00	0.00	0.02	-0.06	0.92
T1LC4	0.07	0.01	0.01	-0.07	-0.03	0.07	0.84
T1LC5	-0.01	0.00	0.01	0.00	0.00	0.01	0.94
T1LC6	-0.05	0.04	-0.01	-0.01	0.01	0.01	0.90
T1LC7	-0.07	-0.02	0.00	0.01	0.04	-0.08	0.90
T1LC8	-0.10	0.02	0.00	0.03	0.04	-0.08	0.89
T1LC11J1	-0.01	0.02	-0.01	0.00	-0.04	0.00	0.98
T1LC12J1	-0.02	0.00	0.00	0.01	-0.02	0.01	0.98
T1LC13J1	-0.02	0.03	0.00	0.03	0.00	0.00	0.98
T1LC14J1	-0.03	0.01	0.00	0.03	0.02	-0.03	0.96
T1LC15J1	-0.03	0.01	-0.02	0.03	-0.01	-0.04	0.98
T1LC11J101	0.13	-0.03	0.01	-0.05	-0.04	0.11	0.87
T1LC12J101	0.07	-0.04	0.00	-0.01	-0.01	0.04	0.92
T1LC13J101	0.02	0.00	0.01	-0.02	0.00	0.03	0.95
T1LC14J101	0.00	-0.04	0.01	0.00	0.04	-0.06	0.94
T1LC15J101	0.03	-0.04	0.00	0.03	0.02	-0.01	0.96
T2LC1	0.10	-0.09	-0.01	-0.05	-0.02	-0.09	0.74
T2LC2	-0.05	0.04	0.01	0.07	-0.01	-0.01	0.95
T2LC3	-0.02	-0.01	0.01	-0.01	0.03	-0.06	0.97
T2LC4	0.13	-0.03	0.03	-0.03	0.02	0.06	0.84
T2LC5	0.11	-0.05	-0.02	0.02	0.00	-0.03	0.90
T2LC6	-0.05	0.04	0.01	0.09	0.01	0.03	0.88
T2LC7	-0.01	-0.02	0.01	0.00	0.04	-0.03	0.97
T2LC8	0.01	0.02	-0.02	0.09	0.03	0.05	0.86
T2LC11J1001	0.95	0.02	0.00	-0.02	0.02	-0.03	0.95
T2LC12J1001	0.97	0.03	0.00	0.03	-0.02	-0.03	0.98
T2LC13J1001	0.99	0.05	0.00	0.03	0.00	-0.02	0.95
T2LC14J1001	0.80	-0.07	0.00	-0.01	0.04	-0.01	0.91
T2LC15J1001	0.96	0.04	0.00	0.02	-0.02	-0.01	0.97
T2LC11J101	0.07	-0.01	0.03	-0.09	0.00	0.06	0.83
T2LC12J101	0.00	-0.02	-0.01	-0.07	-0.02	-0.03	0.88
T2LC13J101	-0.11	0.06	0.01	0.02	-0.01	0.02	0.91
T2LC14J101	-0.05	0.00	0.00	-0.03	0.03	-0.03	0.98
T2LC15J101	-0.12	0.06	-0.01	-0.03	-0.01	0.03	0.85
T3LC1	-0.01	-0.06	0.05	-0.06	0.06	-0.06	0.80
T3LC2	-0.01	-0.01	-0.02	-0.05	-0.33	-0.04	0.87
T3LC3	0.02	-0.08	0.05	0.04	-0.27	0.03	0.90
T3LC4	0.04	-0.02	0.04	-0.02	0.65	0.14	0.83
T3LC5	0.02	-0.04	0.03	0.05	0.77	-0.06	0.88
T3LC6	-0.01	0.01	-0.04	-0.01	0.19	-0.07	0.89
T3LC7	0.02	-0.06	0.02	0.07	0.19	0.01	0.89
T3LC8	0.04	0.00	0.01	0.10	0.91	-0.01	0.84
T3LC11J1001	0.02	0.94	0.02	-0.02	0.03	0.03	0.96
T3LC12J1001	0.04	0.98	0.01	0.00	0.10	-0.04	0.96
T3LC13J1001	0.00	0.86	-0.02	-0.03	-0.11	-0.03	0.96
T3LC14J1001	-0.01	0.81	0.01	0.01	-0.10	0.00	0.96
T3LC15J1001	0.04	1.02	0.00	0.02	0.07	-0.04	0.98
T3LC11J101	-0.01	0.03	0.00	-0.09	0.66	0.15	0.85

(Table 11.16 continues)

(Table 11.16 continued)

	PLC7	PLC8	PLC9	PLC10	PLC11	PLC12	Com
T3LC12J101	-0.05	0.03	-0.01	-0.06	0.82	-0.03	0.83
T3LC13J101	-0.05	0.07	-0.06	-0.06	0.22	-0.06	0.85
T3LC14J101	-0.01	0.00	-0.01	0.02	0.22	0.01	0.90
T3LC15J101	-0.03	0.06	-0.04	-0.01	0.86	0.02	0.87
T4LC1	0.02	0.04	-0.04	0.03	-0.10	0.08	0.75
T4LC2	0.02	-0.04	-0.02	0.03	0.02	-0.31	0.98
T4LC3	0.00	0.00	-0.06	0.00	0.01	-0.10	0.96
T4LC4	-0.09	-0.03	0.01	0.08	-0.02	0.74	0.90
T4LC5	-0.02	0.07	-0.02	0.07	-0.07	0.57	0.88
T4LC6	-0.02	0.02	-0.01	0.05	0.00	0.12	0.90
T4LC7	-0.02	0.02	-0.06	0.03	-0.02	0.17	0.97
T4LC8	-0.06	0.04	-0.01	0.06	0.01	0.74	0.78
T4LC11J1001	-0.03	-0.02	0.93	0.00	0.01	0.10	0.97
T4LC12J1001	0.03	0.03	1.02	-0.01	-0.02	0.06	0.97
T4LC13J1001	0.00	0.01	0.86	0.01	-0.03	-0.13	0.97
T4LC14J1001	-0.02	0.00	0.74	-0.01	0.00	-0.05	0.96
T4LC15J1001	0.01	0.00	1.00	0.00	0.03	0.04	0.98
T4LC11J101	-0.02	-0.07	0.04	-0.03	0.04	0.83	0.84
T4LC12J101	0.05	0.01	0.00	-0.05	-0.01	0.66	0.89
T4LC13J101	0.04	-0.02	0.01	-0.03	0.05	0.20	0.92
T4LC14J101	0.01	0.00	-0.05	-0.01	0.02	0.21	0.98
T4LC15J101	0.04	-0.04	0.03	-0.09	0.08	0.84	0.85
T5LC1	-0.05	0.08	0.05	-0.03	-0.02	-0.13	0.74
T5LC2	-0.01	0.00	-0.01	0.02	0.00	0.01	0.98
T5LC3	0.04	0.01	0.00	-0.07	0.00	0.02	0.98
T5LC4	0.08	0.05	-0.02	0.00	0.00	0.13	0.86
T5LC5	0.00	0.02	0.00	-0.02	0.01	-0.09	0.88
T5LC6	0.01	-0.03	-0.03	0.03	0.02	0.01	0.92
T5LC7	0.04	-0.02	-0.02	-0.09	0.00	0.03	0.96
T5LC8	0.04	-0.05	-0.04	0.01	0.02	-0.02	0.83
T5LC11J1001	0.03	0.03	-0.01	0.87	0.01	0.06	0.95
T5LC12J1001	0.00	0.02	0.00	0.97	0.00	-0.06	0.97
T5LC13J1001	-0.01	-0.03	0.01	0.96	0.02	0.03	0.96
T5LC14J1001	0.04	-0.01	0.02	0.73	0.01	0.00	0.92
T5LC15J1001	0.02	-0.01	-0.01	0.94	0.01	-0.03	0.97
T5LC11J101	0.00	0.05	0.01	0.04	-0.02	0.13	0.85
T5LC12J101	-0.09	0.03	0.03	0.02	-0.02	-0.11	0.89
T5LC13J101	-0.08	-0.03	0.01	0.07	0.00	0.01	0.85
T5LC14J101	0.00	-0.01	0.01	-0.08	-0.02	0.01	0.98
T5LC15J101	-0.08	-0.02	0.01	0.06	0.00	-0.04	0.83
Cumu	55.64	59.94	64.11	67.92	71.46	74.78	
RSSL	5.09	5.81	5.15	5.10	6.07	5.41	

Note. Extraction Method: Principal Components Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 12 iterations. C = Component; Com = Communalities; RSSL = Rotation Sums of Squared Loadings; Cumu = Cumulative %.

Table 11.17
Summary of Interpretations of the Components of Lexical Complexity

	Lexical density and sophistication	Lexical sophistication	Lexical complexity
G	PLC1: Lexical complexity (Derived divided by the square root of No. of tokens)		
T1		PLC3: Task 1 Lexical sophistication (using J1001)	PLC2: Task 1 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
T2	PLC6: Task 2 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	PLC7: Task 2 Lexical sophistication (using J1001)	
T3	PLC11: Task 3 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	PLC8: Task 3 Lexical sophistication (using J1001)	PLC4: Task 3 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
T4	PLC12: Task 4 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	PLC9: Task 4 Lexical sophistication (using J1001)	
T5	PLC5: Task 5 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	PLC10: Task 5 Lexical sophistication (using J1001)	

Note. G = General component, which measures of all the tasks loaded on; T = Task. J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Furthermore, each measure appeared at different components of each task except for the following: (a) LC1, which loaded on the PLC1 across all the tasks, (b) LC5, which loaded on the PLC1 across Tasks 1, 3, 4, and 5, and (c) LC12J101, which loaded on PLC1 across Tasks 1, 3, and 4. Thus, most lexical complexity measures tend to lack strong generalizability across tasks and their results are likely to change, depending on tasks. Therefore, the Hypothesis for Validity 4 (Each measure of lexical complexity

across tasks loads on the same component) was not supported for most measures and supported for the three exceptional measures (i.e., LC1, LC5, and LC12J101), which tend to have relatively strong generalizability across tasks. It should be noted that LC1, LC5, and LC12J101 are measures with the square root of the number of tokens in the denominators.

In addition, most measures loaded on a component, but as seen in Table 11.18, there were 17 exceptions: four measures in Task 1 (LC4, LC8, LC11J101, and LC15J101), five measures in Task 2 (LC3, LC6, LC7, LC13J101, and LC14J101), four measures in Task 4 (LC3, LC7, LC13J101, and LC14J101), and four measures in Task 5 (LC3, LC7, LC13J101, and LC14J101). Although there were no measures that did not load any components across all the five tasks, LC3, LC7, LC13J101, and LC14J101 did not load any components across three tasks (i.e., Tasks 1, 2, and 4). The characteristics and validity of these measures need to be examined further.

Table 11.18
Speaking Performance Measures That Did Not Load on Any Components

Task 1	Task 2	Task 3	Task 4	Task 5
LC4	LC3	None	LC3	LC3
LC8	LC6		LC7	LC7
LC11J101	LC7		LC13J101	LC13J101
LC15J101	LC13J101		LC14J101	LC14J101
	LC14J101			

In summary, the Hypothesis for Validity 3 (The Speaking Performance Measures of lexical complexity on the same task load on the same component) was supported for accuracy and partially supported for fluency, syntactic complexity, and lexical complexity. The Hypothesis for Validity 4 (Each measure across tasks load on the same component) was partially supported for fluency, syntactic complexity, and lexical complexity, and not supported for accuracy. These results suggest that the measures of

the same construct are not always related strongly to each other and that most values of each measure are likely to be affected by tasks, as are most elements of speaking performance, as assessed by each measure. This further implies that most elements of speaking performance may need to be defined in terms of tasks as well as knowledge and ability in defining constructs before test construction and to be interpreted with both tasks and traits, which supports Chapelle's (1998) interactionalist position. In addition, the reason for the distinct results regarding accuracy compared to the other three constructs may be that the number of accuracy measures was limited and the elements of accuracy assessed were limited.

Table 11.19 summarizes the results obtained from Studies 1B and 2D. There were 13 measures comparable across the two studies (i.e., F1, F3 to F6, A1, A2, SC1, SC2, SC4, LC1, LC4 (LC2 in Study 1B) and LC11J1. When the results that found strong generalizability across all the five tasks were focused on, there were differences between the two studies (Studies 1B and 2D). In Study 1B, no measures were strongly generalizable across all the tasks, whereas in Study 2D, three measures were strongly generalizable across all the tasks. There may be two reasons for the diversions. First, the method to measure time was different for F1 and F3. In Study 1B, all test takers had different speaking times, while in Study 2D, they had the same speaking time of 45 seconds. This result indicates that the method that Study 2D used was better in terms of generalizability across tasks. The second reason for obtaining different results in Table 11.19 may be due to using a different method of analysis. In Study 1B, simple correlations were used, whereas Study 2D used principal components analysis (PCA). However, in Study 1B, the correlation coefficients of SC3 were .47, .48, .60 across tasks (see section 5.3.3), and they were very close to the criterion of strong correlations (r = .50 or above). Thus, it could be said that SC3 was strongly generalizable across

the tasks both in Studies 1B and 2D, and this is used as a conclusion of this study.

Table 11.19
Comparisons Between Studies 1B and 2D: More Generalizable Measures

Study 1B that were generalizable across the three tasks (see section 6.3)	Study 2D that were generalizable across all the five tasks
None	F1, F3, SC3

Table 11.20 shows the summary of interpretations of each component. Overall, there were 28 components and each was considered to be an element of speaking performance.

Table 11.20
Summary of Interpretations of Each Component for Each Aspect of Speaking Performance

Construct	Code	Interpretation
Fluency	PF1	Speaking speed
	PF2	Task 1 Dysfluency
	PF3	Task 2 Dysfluency
	PF4	Task 4 Dysfluency
	PF5	Task 3 Dysfluency
	PF6	Task 5 Dysfluency
Accuracy	PA1	Task 1 Accuracy
	PA2	Task 3 Accuracy
	PA3	Task 4 Accuracy
	PA4	Task 5 Accuracy
	PA5	Task 2 Accuracy
Syntactic complexity	PSC1	Task 1 Syntactic complexity
	PSC2	Task 3 Syntactic complexity
	PSC3	Task 5 Syntactic complexity
	PSC4	Task 4 Syntactic complexity
	PSC5	AS-unit length and Task 2 Syntactic complexity
Lexical complexity	PLC1	Lexical complexity (Derived divided by the square root of No. of tokens)
	PLC2	Task 1 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
	PLC3	Task 1 Lexical sophistication (Using J1001)
	PLC4	Task 3 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
	PLC5	Task 5 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC6	Task 2 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC7	Task 2 Lexical sophistication (Using J1001)
	PLC8	Task 3 Lexical sophistication (Using J1001)
	PLC9	Task 4 Lexical sophistication (Using J1001)
	PLC10	Task 5 Lexical sophistication (Using J1001)
	PLC11	Task 3 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC12	Task 4 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)

Note. J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

11.3.5 Investigating the Hypothesis for Validity 5

Tables 11.21 to 11.24 show the correlations between the components of the same

construct. As for fluency, relationships between the six fluency components were weak or moderate (r = .19 to .42). With regard to accuracy, there were almost zero, weak, and moderate correlations (r = .03 to .32) between "Accuracy" for the five tasks. Regarding syntactic complexity, there were almost zero and low correlations between syntactic complexity components (r = .04 to .29). Concerning lexical complexity, there were almost zero and low relationships between lexical complexity components (r = -.22 to .36). Although it was expected that positive correlations would be obtained between lexical complexity components because they all assess some elements of lexical complexity, negative correlations were found. However, since the degree of negative correlations was small or almost zero (r = -.22 to -.02), these components were not excluded for the subsequent analysis.

Table 11.21
Correlations Between the Components of Fluency in the Component Correlation Matrix

	PF1	PF2	PF3	PF4	PF5	PF6
PF1	--	.20**	.32**	.25**	.19**	.28**
PF2		--	.28**	.33**	.30**	.28**
PF3			--	.30**	.28**	.31**
PF4				--	.39**	.33**
PF5					--	.42**

Note. n = 225.
*p < .05. **p < .01.

Table 11.22
Correlations Between the Components of Accuracy in the Component Correlation Matrix

	PA1	PA2	PA3	PA4	PA5
PA1	--	.30**	.28**	.17*	.23**
PA2		--	.32**	.29**	.25**
PA3			--	.25**	.18**
PA4				--	.03

Note. n = 225.
*p < .05. **p < .01.

Table 11.23
Correlations Between the Components of Syntactic Complexity in the Component Correlation Matrix

	PSC1	PSC2	PSC3	PSC4	PSC5
PSC1	--	.09	.04	.14*	.10
PSC2		--	.10	.14*	.26**
PSC3			--	.14*	.29**
PSC4				--	.20**

Note. $n = 225$.
*$p < .05$. **$p < .01$.

Table 11.24
Correlations Between the Components of Lexical Complexity in the Component Correlation Matrix

	PLC2	PLC3	PLC4	PLC5	PLC6	PLC7	PLC8	PLC9	PLC10	PLC11	PLC12
PLC1	-.05	-.07	-.22**	.04	-.08	-.07	-.15*	-.22**	-.08	-.13	-.14
PLC2	--	.29**	.19**	.08	.04	-.01	-.03	.02	-.06	.08	.12
PLC3		--	.00	.16*	.09	.07	.02	.01	.11	.09	.05
PLC4			--	-.03	.00	-.10	.35**	.14*	-.11	.36**	.03
PLC5				--	.29**	.01	.07	-.02	.29**	.17*	.19**
PLC6					--	.18**	.06	.06	.06	.08	.14*
PLC7						--	.04	.01	.08	-.02	.08
PLC8							--	.12	.02	.29**	.11
PLC9								--	.05	.06	.15*
PLC10									--	.02	.08
PLC11										--	.19**

Note. $n = 225$.
*$p < .05$. **$p < .01$.

Tables 11.25 to 11.28 demonstrate the correlations between components of the different constructs. The degree of correlation varied from almost zero relationships to strong ones and from negative to positive ones. The strong correlations were found between "Speaking speed (PF1)" and "AS-unit length and Task 2 Syntactic complexity (PSC5)" ($r = .51$), and between "Speaking speed (PF1)" and "Lexical complexity (Derived divided by the square root of the number of tokens: PLC1)" ($r = .80$). The results suggest that those who have faster speaking speed tend to produce longer AS-unit length and to have a greater degree of lexical complexity. This tendency may be related to the number of words that test takers produced because the measures that loaded on the three components tended to have the number of tokens, types, lexical types, or sophisticated types in the numerators. In other words, those who produce

more tokens tend to have faster speaking speed, produce longer AS-unit length, and produce more types, lexical types, and sophisticated types, which leads to a greater degree of lexical complexity (see section 12.3.5).

Table 11.25
Correlations Between the Components of Fluency and the Other Three Aspects

	PA1	PA2	PA3	PA4	PA5	PSC1	PSC2	PSC3	PSC4	PSC5	PLC1	PLC2
PF1	.16*	.18**	.13*	.21**	-.14*	.19**	.13*	.32**	.39**	.51**	.80**	-.16**
PF2	-.05	-.13*	-.06	.04	.04	-.11	-.07	.16*	.06	.18**	.15*	.03
PF3	.07	.09	.12	.16*	.03	.06	-.02	.19**	.06	.30**	.22**	.07
PF4	-.05	-.08	-.15*	.02	-.14*	-.03	.06	.15*	.14*	.18**	.18**	-.09
PF5	-.09	-.05	-.12	.02	-.06	-.05	.02	.15*	.02	.14*	.10	-.06
PF6	.11	.08	.15*	.14*	-.06	.04	.06	.22**	.19*	.22**	.20**	-.06

Note. $n = 225$.
*$p < .05$. **$p < .01$.

	PLC3	PLC4	PLC5	PLC6	PLC7	PLC8	PLC9	PLC10	PLC11	PLC12
PF1	-.10	-.38**	-.06	-.19**	-.12	-.19**	-.31**	-.08	-0.19**	-0.30**
PF2	-.10	.03	-.04	-.28**	-.06	-.01	-.16*	-.01	0.11	-0.01
PF3	-.01	-.10	-.10	-.26**	-.22**	-.14*	-.13*	.01	-0.08	-0.12
PF4	-.01	-.05	.01	-.15*	-.12	-.03	-.12	.14*	0.07	-0.17*
PF5	-.07	.01	-.12	-.17*	-.10	-.07	-.22**	-.04	0.06	-0.09
PF6	-.01	-.14*	-.19*	-.08	-.03	-.10	-.22**	-.07	-0.03	-0.16*

Table 11.26
Correlations Between the Components of Accuracy and the Other Three Aspects

	PF1	PF2	PF3	PF4	PF5	PF6	PSC1	PSC2	PSC3	PSC4	PSC5	PLC1
PA1	.16*	-.05	.07	-.05	-.09	.11	.11	.09	.01	.06	.27**	.14*
PA2	.18*	-.13*	.09	-.08	-.05	.08	.14*	.08	.00	.11	.19**	.16*
PA3	.13*	-.06	.12	-.15*	-.12	.15*	.10	.04	.01	.08	.25**	.11
PA4	.21**	.04	.16*	.02	.02	.14*	.16*	.10	.23**	.27**	.17*	.15*
PA5	-.14*	.04	.03	-.14*	-.06	-.06	.04	.00	.02	-.09	-.01	-.05

Note. $n = 225$.
*$p < .05$. **$p < .01$.

	PLC2	PLC3	PLC4	PLC5	PLC6	PLC7	PLC8	PLC9	PLC10	PLC11	PLC12
PA1	-.08	-.05	-.10	-.21**	-.20**	.09	-.03	-.05	-.09	-.24**	-.08
PA2	-.09	.01	-.17*	-.18**	-.16*	.08	-.04	-.09	-.02	-.38**	-.11
PA3	.04	-.03	-.05	-.24**	-.09	-.09	-.06	-.03	-.09	-.14*	-.30**
PA4	-.10	.01	-.10	-.24**	-.24**	-.07	-.03	-.20**	-.06	-.13*	-.02
PA5	.01	-.06	.07	-.07	-.16*	.14*	.00	.06	-.09	-.01	-.04

Table 11.27
Correlations Between the Components of Syntactic Complexity and the Other Three Aspects

	PF1	PF2	PF3	PF4	PF5	PF6	PA1	PA2	PA3	PA4	PA5
PSC1	.19**	-.11	.06	-.03	-.05	.04	.11	.14*	.10	.16*	.04
PSC2	.13*	-.07	-.02	.06	.02	.06	.09	.08	.04	.10	.00
PSC3	.32**	.16*	.19**	.15*	.15*	.22**	.01	.00	.01	.23**	.02
PSC4	.39**	.06	.06	.14*	.02	.19**	.06	.11	.08	.27**	-.09
PSC5	.51**	.18**	.30**	.18**	.14*	.22**	.27**	.19**	.25**	.17*	-.01

Note. $n = 225$.
*$p < .05$. **$p < .01$.

	PLC1	PLC2	PLC3	PLC4	PLC5	PLC6	PLC7	PLC8	PLC9	PLC10
PSC1	.10	.01	.21**	-.09	.07	-.02	.04	.06	-.08	-.03
PSC2	.21**	.02	-.05	.06	-.16*	.02	-.08	-.09	-.09	-.14*
PSC3	.24**	.03	-.06*	.06	-.15*	-.11	-.08	-.11	-.11	-.15*
PSC4	.26**	-.18**	-.13*	-.19**	-.14*	-.12	-.02	-.03	-.18**	.06
PSC5	.41**	.18**	-.14*	.10	-.29**	-.31**	-.27**	.02	-.11	-.09

	PLC11	PLC12
PSC1	-.05	.08
PSC2	-.11	-.11
PSC3	-.17*	-.16*
PSC4	-.05	-.20**
PSC5	-.17*	-.35**

Table 11.28
Correlations Between the Components of Lexical Complexity and the Other Three Aspects

	PF1	PF2	PF3	PF4	PF5	PF6	PA1	PA2	PA3
PLC1	.80**	.15*	.22**	.18**	.10	.20**	.14*	.16*	.11
PLC2	-.16*	.03	.07	-.09	-.06	-.06	-.08	-.09	.04
PLC3	-.10	-.10	-.01	-.01	-.07	-.01	-.05	.01	-.03
PLC4	-.38**	.03	-.10	-.05	.01	-.14*	-.10	-.17*	-.05
PLC5	-.06	-.04	-.10	.01	-.12	-.19**	-.21**	-.18**	-.24**
PLC6	-.19**	-.28**	-.26**	-.15*	-.17*	-.08	-.20**	-.16*	-.09
PLC7	-.12	-.06	-.22**	-.12	-.10	-.03	.09	.08	-.09
PLC8	-.19**	-.01	-.14*	-.03	-.07	-.10	-.03	-.04	-.06
PLC9	-.31**	-.16*	-.13*	-.12	-.22**	-.22**	-.05	-.09	-.03
PLC10	-.08	-.01	.01	.14*	-.04	-.07	-.09	-.02	-.09
PLC11	-.19**	.11	-.08	.07	.06	-.03	-.24**	-.38**	-.14*
PLC12	-.30**	-.01	-.12	-.17*	-.09	-.16*	-.08	-.11	-.30**

Note. $n = 225$.
*$p < .05$. **$p < .01$.

	PA4	PA5	PSC1	PSC2	PSC3	PSC4	PSC5
PLC1	.15*	-.05	.10	.21**	.24**	.26**	.41**
PLC2	-.10	.01	.01	.02	.03	-.18**	.18**
PLC3	.01	-.06	.21**	-.05	-.06	-.13*	-.14*
PLC4	-.10	.07	-.09	.06	.06	-.19**	.10
PLC5	-.24**	-.07	.07	-.16*	-.15*	-.14*	-.29**
PLC6	-.24**	-.16*	-.02	.02	-.11	-.12	-.31**
PLC7	-.07	.14*	.04	-.08	-.08	-.02	-.27**
PLC8	-.03	.00	.06	-.09	-.11	-.03	.02
PLC9	-.20**	.06	-.08	-.09	-.11	-.18*	-.11
PLC10	-.06	-.09	-.03	-.14*	-.15*	.06	-.09
PLC11	-.13*	-.01	-.05	-.11	-.17*	-.05	-.17**
PLC12	-.02	-.04	.08	-.11	-.16*	-.20**	-.35**

In the Hypothesis for Validity 5, it was hypothesized that the correlation between components of the same constructs would be higher than those between components of different constructs. As can been seen in Tables 11.21 to 11.28, some components satisfied the hypothesis whereas others did not. For example, a correlation between PF1 and PF5 (r = .19) was higher than the one between PF1 and PA1 (r = .16) but lower than the one between PF1 and PSC3 (r = .32), and the former combination satisfied the hypothesis, whereas the latter did not. In order to examine the results overall, the correlations were averaged using Fisher's z transformation. Table 11.29 shows that for fluency, the averaged correlations between the component scores of the same construct were higher than the averaged ones between the component scores of the different constructs in all cases (i.e., .30 > .02; .30 > .13; .30 > -.06), as was also true for accuracy, syntactic complexity, and lexical complexity. The effect sizes of differences between the averaged correlation coefficients were almost zero, small, or medium (q = .02 to .37; see Table 11.30). Therefore, the Hypothesis for Validity 5 was partially supported.

Table 11.29
Averaged Correlations Between the Components of the Four Aspects

	Fluency	Accuracy	Syntactic complexity	Lexical complexity
Fluency	.30**	.02	.13*	-.06
Accuracy	--	.23**	.10	-.07
Syntactic complexity		--	.15*	-.05
Lexical complexity			--	.07

Table 11.30
*Differences Between Averaged Correlation Coefficients (*q*) of the Components of the Four Aspects*

$z_F - z_A$	$z_F - z_{SC}$	$z_F - z_{LC}$	$z_A - z_F$	$z_A - z_{SC}$	$z_A - z_{LC}$
0.29	0.18	0.37	0.21	0.13	0.30

Note. z_F = Transformed correlation coefficient of averaged fluency component.

$z_{SC} - z_F$	$z_F - z_A$	$z_F - z_{SC}$	$z_{LC} - z_F$	$z_{LC} - z_A$	$z_{LC} - z_{SC}$
0.02	0.05	0.20	0.13	0.14	0.12

11.3.6 Investigating the Hypotheses for Validity 6 and 7 and the Main Hypothesis 2: Validity Argument of the Speaking Performance Measures

As seen in sections 11.3.2 to 11.3.5, it was found that the Hypotheses for Validity 1 and 2 were supported, while the Hypotheses for Validity 3, 4, and 5 were partially supported.

As for the Hypothesis for Validity 6 regarding the content aspects, the Speaking Performance Measures used in this study were mainly derived from the measures that met all the following criteria: ratio measures used in two or more empirical L2 speaking studies published since 1990, and measures that use the category of "fluency," "accuracy," or "complexity" except for "lexical complexity." Except for the AS-unit length (the number of tokens per AS-unit), of which the meaning was explored in section 11.3.3, the measures had one single meaning and there was no disagreement among experts. In addition, some measures were added, but there were reasons for including them. Therefore, it can be concluded that this hypothesis was mostly satisfied.

As for the Hypothesis for Validity 7 (The reliability of raters is high on the Speaking Performance Measures), as seen in section 12.2.3.2, all the reliability of raters in calculating the measures was high on the Speaking Performance Measures, which leads to the support of this hypothesis.

So far, the validity was examined from seven perspectives. Results that did not agree with the hypotheses were negative evidence for validity, and the reasons for diversions need to be examined, which may lead to the further revision or exploration of the meaning of the Speaking Performance Measures if necessary. However, much positive evidence for validity was observed. Therefore, the Main Hypothesis 2 (The Speaking Performance Measures have positive evidence for validity.) was confirmed

and there is an acceptable level of validity of inferences and uses based on the values of the Speaking Performance Measures for novice Japanese learners of English. Thus, it is concluded that using the Speaking Performance Measures in this research is meaningful and appropriate.

11.4 Summary of Chapter 11

This chapter examines (a) whether the AS-unit length (the number of tokens per AS-unit) indicates fluency or syntactic complexity and (b) whether the Speaking Performance Measures of fluency, accuracy, syntactic complexity, and lexical complexity have positive evidence for validity. The results suggest that the AS-unit length belongs to syntactic complexity, not fluency and that there is an acceptable level of validity of inferences and uses based on the values of the Speaking Performance Measures for novice Japanese learners of English.

Chapter 12 Study 2E: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance[12.1]

12.1 Purpose, Hypotheses, and a Research Question

The purpose of this chapter is to examine relationships between size and depth of productive vocabulary knowledge and speaking performance, which is the main focus in the current dissertation. This study focuses on monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level. In order to ensure this research is of a high standard, seven studies were conducted in the previous chapters (Chapters 4 and 5, 7 to 11). Figure 12.1 shows where the present chapter (i.e., Chapter 12 Study 2E: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance) is located in the overall research. Four Hypotheses and one Research Question were addressed for this purpose.

Hypothesis 1: There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types in speaking performance across tasks.

Hypothesis 2: There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to have better speaking performance related to vocabulary across tasks.

Hypothesis 3: There is a moderate or strong tendency for those who produce a greater number of tokens and types in speaking performance to have better speaking performance related to vocabulary across tasks.

Hypothesis 4: Size of productive vocabulary knowledge is related more to speaking performance than depth of productive vocabulary knowledge is.

Research Question: Are there any individual factors other than productive vocabulary knowledge related to speaking performance to a moderate or strong degree?

The Hypotheses 1 to 3 were based on Study 1C (see section 6.3), which suggests that those who have more productive vocabulary knowledge may produce more tokens and types (Hypothesis 1) and may excel in vocabulary aspects of speaking performance (Hypothesis 2). The Hypothesis 3 was posed to check the results of the Hypotheses 1 and 2. The Hypothesis 4 was based on Meara (1996; see section 2.2.1). The Research Question was posed in order to explore effects of individual factors other than productive vocabulary knowledge on speaking performance.

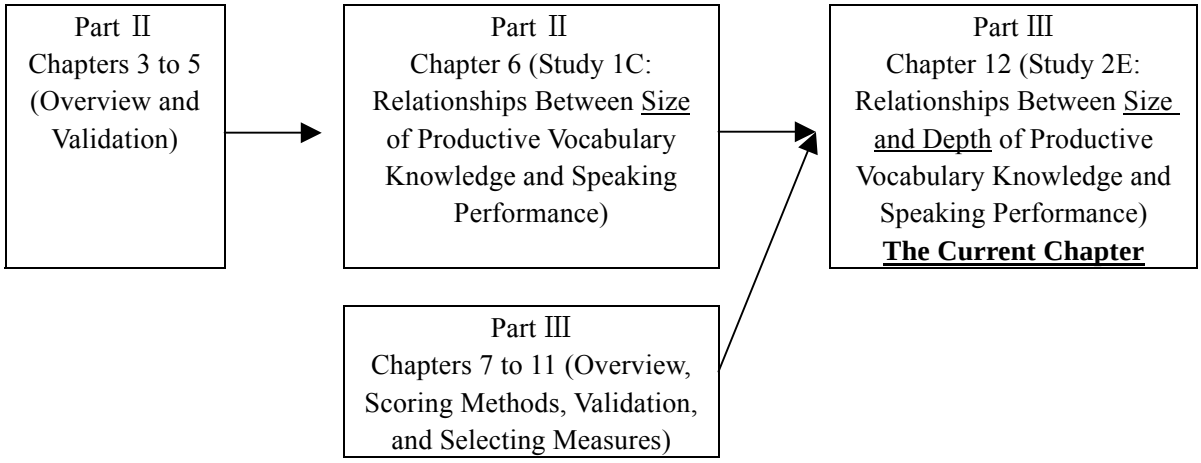


Figure 12.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

12.2 Method

12.2.1 Participants

The participants in this study were 225 students studying English in Japan (see Table 12.1). This group contained 128 (both public and national) junior high school

students and 97 (both public and national) senior high school students. All of them were the same as those who were analyzed in Study 2D (n = 225). A wide range of learners in terms of English ability participated. In fact, the range in Study 2E was much wider than that in Study 1C. They were considered to be novice Japanese learners of English in this research.[12.2]

Table 12.1
Participants at Each School

School	School type	School year	*n* who took the Speaking Test	*n* for Study 2E	*n* excluded (Reasons for the exclusion)
JHS A	Public	3rd	14	4	10 (Problem 3, Name 2, No clauses 2, Outlier 3)
JHS B	Public	3rd	57	32	25 (Problem 3, Name 10, No clauses 9, Playful 1, Outlier 2)
JHS C	Public	3rd	26	13	13 (Problem 1, Name 5, No clauses 6, Outlier 1)
JHS D	Public	3rd	27	5	22 (Interview 7, Problem 2, Name 1, No clauses 9, Playful 2, Outlier 1)
JHS G	National	3rd	104	74	30 (Problem 15, Name 7, No clauses 2, Outlier 6)
JHS Total			228	128	100 (Interview 7, Problem 24, Name 25, No clauses 28, Playful 3, Outlier 13)
SHS A	Public	3rd	28	22	6 (No clauses 5, Outlier 1)
SHS B	Public	1st to 3rd	22	16	6 (Problem 1, No clauses 4, Outlier 1)
SHS C	Public	1st	36	33	3 (No clauses 3)
SHS E	Public	2nd to 3rd	15	8	7 (Problem 1, No clauses 5, Outlier 1)
SHS G	National	1st	23	18	5 (Problem 1, Name 1, Outlier 3)
SHS Total			124	97	27 (Problem 3, Name 1, No clauses 17, Outlier 6)
Total			352	225	108 (Interview 7, Problem 27, Name 26, No clauses 45, Playful 3, Outlier 19)

Note. JHS = Junior high school; SHS = Senior high school; No clauses = A student who did not produce any clauses in at least one task; Playful = A student who did not behave seriously (e.g., who sang a song during a task); Problem = A student who had a mechanical or procedural problem (e.g., a student who had a Version A tape and a Version B test booklet; see section 7.2.3.1); Interview = A student who took the face-to-face version of the Speaking Test first; Name = A student whose tape caused a matching problem between the tape and the name of the speaker. Outlier = A student who belonged to the outlier group (see section 11.3.1).

12.2.2 Instruments

The two tests and one questionnaire were used in this chapter: (a) the Productive Vocabulary Knowledge Test (Productive VKT), (b) the Speaking Test, and (c) the Speaking Test Questionnaire. The Productive VKT consisted of two sections: the Size Section and the Depth Section. The Depth Section had three subsections: the Derivation Subsection, Antonym Subsection, and Collocation Subsection. The Speaking Test used a tape-mediated format, which included five tasks. The details were described in sections 7.2.2.1 to 7.2.2.3

12.2.3 Procedures

12.2.3.1 Test Administration Procedures

The details were described in section 7.2.3.1.

12.2.3.2 Scoring and Coding Procedures

The details were described in sections 9.2.3.2 and 11.2.3.2 (see Table 12.2 for the 28 components derived from principal components analysis; see also Table 12.3 for the 29 Speaking Performance Measures used).

Table 12.2
Summary of 28 Component Scores Used in Study 2E (Derived from Study 2D)

Construct	Code	Interpretation
Fluency	PF1	Speaking speed
	PF2	Task 1 Dysfluency
	PF3	Task 2 Dysfluency
	PF4	Task 4 Dysfluency
	PF5	Task 3 Dysfluency
	PF6	Task 5 Dysfluency
Accuracy	PA1	Task 1 Accuracy
	PA2	Task 3 Accuracy
	PA3	Task 4 Accuracy
	PA4	Task 5 Accuracy
	PA5	Task 2 Accuracy
Syntactic complexity	PSC1	Task 1 Syntactic complexity
	PSC2	Task 3 Syntactic complexity
	PSC3	Task 5 Syntactic complexity
	PSC4	Task 4 Syntactic complexity
	PSC5	AS-unit length and Task 2 Syntactic complexity
Lexical complexity	PLC1	Lexical complexity (Derived divided by the square root of No. of tokens)
	PLC2	Task 1 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
	PLC3	Task 1 Lexical sophistication (Using J1001)
	PLC4	Task 3 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)
	PLC5	Task 5 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC6	Task 2 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC7	Task 2 Lexical sophistication (Using J1001)
	PLC8	Task 3 Lexical sophistication (Using J1001)
	PLC9	Task 4 Lexical sophistication (Using J1001)
	PLC10	Task 5 Lexical sophistication (Using J1001)
	PLC11	Task 3 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)
	PLC12	Task 4 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)

Note. J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated. The same table as Table 11.20.

Table 12.3
Summary of 29 Speaking Performance Measures Used in Study 2E

Construct	Code	Old code	Measure of Study 2E
Fluency	F1	F1	No. of tokens per minute
	F3	F3	No. of clauses per minute
	F4+	F4	No. of dysfluency markers per minute
	F5+	F5	No. of dysfluency markers per token
	F6+	F6	No. of dysfluency markers per AS-unit
Accuracy	A1	A1	No. of error-free clauses per clause
	A2	A2	No. of error-free AS-units per AS-unit
Syntactic complexity	SC1	SC1	No. of clauses per AS-unit
	SC2	SC2	No. of subordinate clauses per AS-unit
	SC3		No. of subordinate clauses per clause
	SC4	SC3	No. of tokens per AS-unit
Lexical complexity			(Lexical diversity)
	LC1	LC1	Guiraud index: No. of types divided by the square root of No. of tokens
	LC2	LCN2	No. of types per AS-unit
	LC3	LCN5	No. of types per clause
			(Lexical density)
	LC4	LCN9	No. of lexical word types per token
	LC5	LCN14	No. of lexical word types divided by the square root of No. of tokens
	LC6	LCN3	No. of lexical word types per AS-unit
	LC7	LCN6	No. of lexical word types per clause
	LC8	LCN10	No. of lexical word types per type
			(Lexical sophistication using J1001)
	LC11	LC4	No. of sophisticated word types per token
	LC12	LCN12	No. of sophisticated word types divided by the square root of No. of tokens
	LC13	LCN4	No. of sophisticated word types per AS-unit
	LC14	LCN7	No. of sophisticated word types per clause
	LC15	LCN1	No. of sophisticated word types divided by No. of types
			(Lexical sophistication using J101)
	LC11	LC4	No. of sophisticated word types per token
	LC12	LCN12	No. of sophisticated word types divided by the square root of No. of tokens
	LC13	LCN4	No. of sophisticated word types per AS-unit
	LC14	LCN7	No. of sophisticated word types per clause
	LC15	LCN1	No. of sophisticated word types divided by No. of types

Note. + = Measure that has a negative meaning when it has a higher value. F2 and SC3 are the same measures. The old codes were used in Studies 1B, 1C, or 2C. J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated. F2 were deleted because No. of tokens per AS-unit was found to mean syntactic complexity in Study 2D.

12.2.3.3 Analysis Procedures

Six analyses were conducted in this chapter (Study 2E; see Table 12.4 for a

summary; see also Table 12.5 for description of independent variables [IVs] and dependent variables [DVs]).

Table 12.4
Summary of Analysis Procedures in Study 2E

	Method	Independent variable		Dependent variables			H or RQ
		Size and depth	Questionnaire	(a) Tokens, and types	(b) Component scores	(c) Values	
1	MRA	*		*			H1
2	MRA	*			*		H2
3	MRA	*				*	H2
4	Simple correlation (*r*)			*	*		H3
5	Relative Pratt index	*		*	*	*	H4
6	MRA	*	*		*		RQ

Note. MRA = Multiple regression analysis; H = Hypothesis; RQ = Research Question.

Table 12.5
Summary of Independent and Dependent Variables in Study 2E

Independent variables (IVs)	Size: Estimated by the formula: The number of correct/78*3000
	Depth: total scores of correct answers for each subsection Derivations Antonyms Collocations
	11 questionnaire items (see Table 12.17)
Dependent variables (DVs)	(a) The number of tokens The number of types
	(b) Component scores: They were combined scores derived from principal components analysis (PCA) using (c), which was conducted in Study 2D (see section 11.3.4) Fluency: 6 component scores (PF1 to PF6) Accuracy: 5 component scores (PA1 to PA5) Syntactic complexity: 5 component scores (PSC1 to PSC5) Lexical complexity: 12 component scores (PLC1 to PLC 12) 28 component scores in total
	(c) Values of the Speaking Performance Measures for each task (e.g., F1 in Task 1) Fluency: 5 measures * 5 tasks = 25 Accuracy: 2 measures * 5 tasks = 10 Syntactic complexity: 4 measures * 5 tasks – 2 measures (i.e., SC2 and SC3 in Task 2; deleted due to extreme kurtosis; see section 11.3.1) = 18 Lexical complexity: 18 measures * 5 tasks = 90 143 variables in total

In order to investigate the Hypotheses 1 and 2, multiple regression analysis

(MRA) was conducted. In the next section (section 12.3.1), assumptions when using MRA were checked, followed by MRA from three perspectives: using "size and depth of productive vocabulary knowledge" as IVs and (1) "the number of tokens and types" as DVs, (2) "component scores of speaking performance" as DVs, and (3) "values of the Speaking Performance Measures for each task" as DVs.

In order to investigate the Hypothesis 3, (4) simple correlations were used. The interpretation of $|r|$ was as follows: Lower than small < .10; .10 ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large (Cohen, 1988; see section 2.6.4.1). In order to examine the Hypothesis 4 and consider the relative importance of independent variables, (5) the relative Pratt index (*d*; Thomas, 1992; Thomas, Hughes, & Zumbo, 1998; Zumbo, 2005) was used. Standardized regression coefficients (β) were presented for two reasons: (a) to examine the direction of relationships (i.e., positive or negative) and (b) to use standardized regression coefficients to complement the relative Pratt index.

In addition, in order to investigate the Research Question, another MRA was conducted using (6) "size and depth of productive vocabulary knowledge and responses toward questionnaire items" as IVs and "component scores of speaking performance" as DVs.

The second and third analyses (i.e., using "size and depth of productive vocabulary knowledge" as IVs and (2) "component scores of speaking performance" as DVs, and (3) "values of the Speaking Performance Measures for each task" as DVs) were similar, but they were conducted because each method has its own advantages. The second analysis was done because integrated interpretation is possible by using component scores derived from values of the Speaking Performance Measures for each task. The third analysis had the advantages that results of Studies 1C and 2E can be compared using the same measures and that the consistency of the interpretation can

be checked. However, some of the Speaking Performance Measures do not have sufficient variation (see Table 11.8). If only the Speaking Performance Measures with small variations are used, it is rather difficult to conclude whether results showing lack of relationships are due to almost zero relationships or due to problems with the selection of the measures. By using the component scores, these problems can be avoided, and this is another advantage of the second method. Since there are more advantages in the second method, the results of the third method were only used in a complementary way.

Since there were multiple dependent variables, MRA was repeatedly conducted. Of the three types of MRA, standard regression was mainly used, and sequential regression was used in a complementary way.

As for using component scores of lexical complexity, some may wonder if component scores that have different types of denominators but similar interpretation (e.g., PLC4 [Task 3 Lexical complexity with denominators of the number of AS-units or clauses] and PLC11 [Task 3 Lexical density and sophistication with denominators of the number of tokens, √tokens, or types]) may not be very meaningful to be included in the equation. However, there was insufficient evidence that they were less valid than other components in Study 2D. Thus, all the possible component scores were retained and used for the analyses since this study intended to include as many elements of speaking performance as possible.

The interpretation of adjusted R^2 (the proportion explained) was based on Cohen (1988, pp. 413-414): Lower than small < .0196; .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large. Only medium or large effect sizes were interpreted. The two analyses were conducted (a) with all the participants included (n = 225) and (b) with a smaller number of participants after outliers, if any, were excluded. If there were

moderate or strong effect sizes in either analysis, the results were interpreted. This is because the first analysis has an advantage that the same number of participants can be kept constant and that results across tasks and measures are more comparable, whereas the second analysis, which excludes outliers, is better in terms of meeting the assumption of MRA.

12.3 Results and Discussion of Study 2E

12.3.1 Assumption Check of Multiple Regression Analysis and Descriptive Statistics of the Productive Vocabulary Knowledge Test

In this section, assumptions of conducting MRA were examined and then descriptive statistics of the Productive VKT were reported.

There are six assumptions to be checked in order to perform MRA (Tabachnick & Fidell, 2001, pp. 116-122): (a) the ratio of cases to IV, (b) absence of outliers among the IVs and on the DV, (c) absence of multicollinearity and singularity, (d) normality, linearity, and homoscedasticity of residuals, (e) independence of errors, and (f) outliers in the solution.

First, one assumption of MRA is that the ratio of cases to IV is sufficiently large for analysis to be done. According to Tabachnick and Fidell (2001, p. 117), when the critical ratio is set at α = .05 and the ratio of Type 2 error is specified at β = .20, a sample size of more than "50 + 8*m" (m = the number of IVs) is necessary for testing multiple correlations, whereas a sample size of more than "104 + m" is required for testing individual predictors. Thus, in the analyses of the present study, 82 (i.e., 50 + 8m = 50 + 8*4 = 50 + 32) or 108 (i.e., 104 + m = 104 + 4) test takers are needed. Study 2E had 225 participants, which was beyond the required number of test takers.

Second, another assumption of MRA is that there are no outliers among the IVs

or on the DV. In the analysis of examining the assumption of principal components analysis (PCA), both univariate and multivariate outliers were already checked (see section 11.3.1), so this assumption was met.

Third, MRA assumes the absence of multicollinearity and singularity. "With multicollinearity, the variables are very highly correlated [say, .90 and above]. With singularity, … one of the variables is a combination of two or more of the other variables" (Tabachnick & Fidell, 2001, pp. 82-83). Regarding singularity, there was no singularity in the data used in Study 2E. As for multicollinearity, correlations between the Size Section and the depth subsections were high but not very high (r = .62 to .81). Tolerance (i.e., 1 – SMS [squared multiple correlations]; Small values are considered to be a problem.; Tabachnick & Fidell, p. 84) and Condition Index (i.e., "a measure of tightness or dependency of one variable on the others"; High values are considered to be a problem.; Tabachnick & Fidell, p. 85) were also checked using the SPSS output, and the results showed no evidence for multicollinearity.

Fourth, MRA assumes that residuals are normally distributed and that they have linearity and homoscedasticity. If linearity is satisfied, "there is a straight-line relationship between two variables" (Tabachnick & Fidell, 2001, p. 77). If homoscedasticity is met, "the variability in scores for one continuous variable is roughly the same at all values of another continuous variable" (Tabachnick & Fidell, p. 79). When MRA was conducted using the 225 participants, examination of residual scatterplots derived in the SPSS output showed that there was no evidence of violation of normality, linearity, and homoscedasticity.

Fifth, concerning independence of errors, MRA assumes that "errors of prediction are independent of one another," and significant autocorrelation of errors indicates non-independence of errors (Tabachnick & Fidell, 2001, p. 121). The current

study used Durbin-Watson statistics, which "range in value from 0 to 4" (SPSS, 1999, p. 401). According to SPSS, "a value near 2 indicates non-autocorrelation," and "a value toward 0 indicates positive autocorrelation" and "a value toward 4 indicates negative autocorrelation" (p. 401). In this study, all values of the Durbin-Watson statistics were near to 2, so there was no evidence of non-independence of errors.

Sixth, the last assumption is that there are no outliers in the solution, which can be detected examining large residuals in the SPSS output. If standardized residuals were higher than |±3.3|, those with such high residuals were judged as outliers (Tabachnick & Fidell, 2001, p. 122). Thirty-four outliers were found in total. Since the sample size decreased when all the outliers were excluded, these outliers were not excluded. However, the results when the outliers were excluded for each analysis were also presented. When the interpretation was different between the two analyses, the results with moderate or strong effect sizes were interpreted.

Table 12.6 shows means and standard deviations for the Productive VKT (see Table 11.8 for other descriptive statistics).

Table 12.6
Descriptive Statistics of the Size and Depth Sections on the Productive Vocabulary Knowledge Test

	M	*SD*	Minimum	Maximum	Skewness	Kurtosis	α
Size (78 items)	1118.80	398.36	38.00	2077.00	0.16	-0.41	.92
Derivation (20 items)	7.88	3.70	0.00	19.00	0.05	-0.51	.80
Antonym (17 items)	6.70	2.94	0.00	14.00	0.13	-0.47	.75
Collocation (18 items)	10.60	2.87	2.00	18.00	-0.37	0.53	.73

Note. n = 225. All the means and *SD*s for the component scores (e.g., PF1) were 0.00 and 1.00 because they were standardized scores. See Table 11.8 for the basic data and the Speaking Performance Measures. Size was estimated using the following formula: [the number of correct words]/78*3,000]. See Appendixes 12.1 to 12.25 for all the detailed results.

12.3.2 Investigating the Hypothesis 1

The first analysis was conducted using "size and depth of productive vocabulary

knowledge" as IVs and "the number of tokens and types" as DVs. As seen in Table 12.7, some values of the Speaking Performance Measures were predicted by size and depth of the Productive VKT scores to some degree, whereas the other values were not. For example, 17% of the number of tokens in Task 2 was predicted by size and depth of productive vocabulary knowledge. On the other hand, 8% of the number of tokens in Task 1 was predicted by size and depth.

When adjusted R^2 was examined, medium or large effects (R^2 = 0.13 or more) of size and depth of productive vocabulary knowledge were observed both in the number of tokens and the number of types in four tasks (Tasks 2 to 5). Therefore, the Hypothesis 1 (There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types in speaking performance across tasks.) was partially confirmed because it was supported only in four tasks out of the five. It is concluded that there is a moderate or strong tendency in some tasks for those (Japanese learners of English at the novice level) with larger and deeper productive vocabulary knowledge (than others of this level) to produce a greater number of tokens and types when they produce monologic description and comparison without pre-task planning time. The proportions explained in Task 1 were less than 13% but close to it (8 to 10%; see section 12.3.4 for the explanation of differences between Task 1 and other tasks).

Table 12.7
Proportions Explained (Adjusted R^2) of the Number of Tokens and Types by the Productive Vocabulary Knowledge Test

	Adjusted R^2 [n = 225]					Adjusted R^2 [Outliers excluded]				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
Tokens	.08	.17	.21	.26	.17	.08 (0)	.17 (0)	.25 (1)	.26 (0)	.17 (0)
Types	.10	.14	.20	.27	.19	.10 (0)	.14 (1)	.26 (2)	.31 (1)	.19 (0)

Note. T = Task; ( ) = The number of outliers excluded for the analysis. (1) = 224 (i.e., 225-1) students were analyzed. Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large.

12.3.3 Investigating the Hypothesis 2

The second analysis was conducted using "size and depth of productive vocabulary knowledge" as IVs and "component scores of speaking performance" as DVs. As seen in Table 12.8, some values of the component scores of the Speaking Performance Measures were predicted by size and depth of the Productive VKT scores to some degree, whereas the other values were not. For example, 13% of PA1 was predicted by size and depth of productive vocabulary knowledge. On the other hand, only 5% of PF3 was predicted by size and depth.

When adjusted R^2 was examined, medium effects (R^2 = 0.13 to 0.25) of size and depth of productive vocabulary knowledge were observed in PF1 (Speaking speed, 23%), PA1 (Task 1 Accuracy, 13%), PSC5 (AS-unit length and Task 2 syntactic complexity, 23%), and PLC1 (Lexical complexity [Derived divided by the square root of the number of tokens], 19%). Among the four, PF1, PSC5, and PLC1 (19% to 23%), whose interpretation was not limited to a certain task, had higher percentages explained by size and depth than PA1 (13%). PF1, PSC5, and PLC1 can be considered vocabulary-related aspects of speaking performance because the main measures that loaded on PF1, PSC5, and PLC1 had "the number of tokens, types, lexical word types, sophisticated word types (using J101), and clauses" in the numerators (see Table 12.9). This suggests that the Hypothesis 2 (There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to have better speaking performance related to vocabulary across tasks.) was supported. In other words, if a person has larger and deeper productive vocabulary knowledge, vocabulary aspects in speaking performance tend to excel probably because productive vocabulary knowledge can be used for vocabulary-related aspects of speaking performance. It

Table 12.8
Proportions Explained (Adjusted R^2) of the Component Scores by the Productive Vocabulary Knowledge Test

Code	Interpretation	Adjusted R^2 [n = 225]	Adjusted R^2 [Outliers excluded]
PF1	Speaking speed	.23	.26 (1)
PF2	Task 1 Dysfluency	-.01	.04 (3)
PF3	Task 2 Dysfluency	.05	.04 (1)
PF4	Task 4 Dysfluency	.00	.01 (3)
PF5	Task 3 Dysfluency	-.01	-.01 (1)
PF6	Task 5 Dysfluency	.02	.02 (1)
PA1	Task 1 Accuracy	.13	.13 (1)
PA2	Task 3 Accuracy	.10	.10 (0)
PA3	Task 4 Accuracy	.08	.08 (0)
PA4	Task 5 Accuracy	.10	.10 (0)
PA5	Task 2 Accuracy	.00	.00 (0)
PSC1	Task 1 Syntactic complexity	.02	.02 (0)
PSC2	Task 3 Syntactic complexity	.01	.03 (3)
PSC3	Task 5 Syntactic complexity	.01	.07 (7)
PSC4	Task 4 Syntactic complexity	.03	.06 (1)
PSC5	AS-unit length and Task 2 Syntactic complexity	.23	.23 (0)
PLC1	Lexical complexity (Derived divided by the square root of No. of tokens)	.19	.22 (1)
PLC2	Task 1 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)	.00	-.01 (1)
PLC3	Task 1 Lexical sophistication (Using J1001)	.00	.00 (1)
PLC4	Task 3 Lexical complexity (Derived divided by No. of AS-units or clauses; Lexical sophistication using J101)	.03	.03 (4)
PLC5	Task 5 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	.03	.03 (0)
PLC6	Task 2 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	.07	.07 (0)
PLC7	Task 2 Lexical sophistication (Using J1001)	.01	.01 (0)
PLC8	Task 3 Lexical sophistication (Using J1001)	.02	.02 (1)
PLC9	Task 4 Lexical sophistication (Using J1001)	.02	.04 (2)
PLC10	Task 5 Lexical sophistication (Using J1001)	.00	.00 (2)
PLC11	Task 3 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	.05	.05 (0)
PLC12	Task 4 Lexical density and sophistication (Derived divided by No. of tokens, the square root of No. of words, No. of types; Lexical sophistication using J101)	.04	.05 (1)

Note. ( ) = The number of outliers excluded for the analysis. (1) = 224 (i.e., 225-1) students were analyzed. Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large.

Table 12.9
Variables That Loaded on the Components Explained by the Productive Vocabulary Knowledge Test to a Moderate or Strong Degree

Construct	Code	Interpretation	Measures	Numerator	Denominator
Fluency (23 to 26%)	PF1	Speaking speed	F1, F2 for five tasks	No. of tokens and clauses	
Accuracy (13%)	PA1	Task 1 Accuracy	T1A1, T1A2	No. of error-free clauses and error-free AS-units	No. of clauses and AS-units
Syntactic complexity (23%)	PSC5	AS-unit length and Task 2 Syntactic complexity	T1SC4, T2SC1, T2SC4, T3SC4, T4SC4, T5SC4	No. of tokens and clauses	No. of AS-units
Lexical complexity (19 to 22%)	PLC1	Lexical complexity (Derived divided by the square root of No. of tokens)	T1LC1, T1LC5, T1LC12J101, T2LC1, T2LC4, T3LC1, T3LC5, T3LC12J101, T4LC1, T4LC5, T4LC6, T4LC8, T4LC12J101, T5LC1, T5LC5	No. of types, lexical word types and sophisticated word types (J101)	No. of tokens, AS-units, and types

Note. Measures = Measures that loaded on the component (.30 or above). ( ) = Percentage explained by productive vocabulary knowledge. T = Task.

should be noted that PLC2 to PLC12, which are also vocabulary-related aspects of speaking performance, were not moderately or strongly related to productive vocabulary knowledge. Thus, strictly speaking, there were moderate or strong relationships between productive vocabulary knowledge and main vocabulary aspects when Japanese learners of English at the novice level produced monologic description and comparison without pre-task planning time. PLC2 to PLC12 can be considered to be rather minor because they appeared later than PLC1 in the PCA.

In addition, PA1 (Task 1 Accuracy) was also explained by size and depth of productive vocabulary knowledge to a moderate extent. The reasons are explored in section 12.3.4. The proportions of PA2 to PA4 explained were rather close to that of PA1, but that of PA5 (Task 2) was smaller.

Moreover, two points should be noted. First, all four aspects of speaking performance (i.e., fluency, accuracy, syntactic complexity, and lexical complexity) were related to productive vocabulary knowledge, although accuracy was limited to a certain task.

Second, only one element for each aspect was found to be related. To be specific, elements related to productive vocabulary knowledge are (a) one element of fluency (speaking speed), (b) accuracy in a task (i.e., Task 1), (c) one element of syntactic complexity, as mainly measured by the AS-unit length, and (d) one element of lexical complexity, as mainly measured by the Guiraud index (LC1), the number of lexical word types divided by the square root of the number of tokens (LC5), and the number of sophisticated word types (J101) divided by the square root of the number of tokens (LC12J101). Productive vocabulary knowledge is not associated with other elements of speaking performance to a moderate degree. The other elements include dysfluency for each task (one element of fluency), proportions of clauses and subordinate clauses (one element of syntactic complexity), and lexical complexity apart from PLC1, and they are not related to productive vocabulary knowledge across tasks. The reason for lack of relationships with dysfluency may be that dysfluency markers consist of three elements: repetitions, self-corrections, and false starts. They were considered to be something that lowers fluency in this study, but they could be interpreted differently. Some dysfluency markers are necessary and beneficial for effective speaking performance because they tend to reduce pauses and make their messages more accurate, so that misunderstanding may be reduced and listeners' comprehension may be enhanced (Buck, 2001, p. 41). Thus, measures that use dysfluency markers may not be very valid as a measure of fluency. The reason for lack of relationships with proportions of clauses and subordinate clauses may be that the number of clauses and

subordinate clauses is rather limited (see Table 11.8).

12.3.4 Investigating the Speaking Performance Measures for Each Task

The third analysis was conducted using "size and depth of productive vocabulary knowledge" as IDs and "values of the Speaking Performance Measures for each task" as DVs. As seen in Table 12.10, some values of the Speaking Performance Measures were predicted by size and depth of the Productive VKT scores to some degree, whereas the other values were not. For example, 17% of Task 2 F1 was predicted by size and depth of productive vocabulary knowledge. On the other hand, only 2% of Task 1 F4 was predicted by size and depth.

When adjusted R^2 was examined, medium or large effects (R^2 = .13 or more) of size and depth of productive vocabulary knowledge were observed in two measures in Task 1 (A1 and A2), four measures in Task 2 (F1, SC4, LC4, and LC11J101), five measures in Task 3 (F1, F3, LC1, LC4, and LC11J101), three measures in Task 4 (F1, F3, and LC1), two measures in Task 5 (F1 and F3). When adjusted R^2 in the analysis of outliers excluded was examined, medium effects were found in Task 2 SC1 in addition to the variables reported above. These results were generally in line with the results when the component scores were analyzed (see section 12.3.3). Although some differences in the two analyses were observed, it is not surprising that some measures in some tasks were not predicted by size and depth to a moderate degree, whereas similar component scores were predicted. For instance, PF1 (23%; see Table 12.8), which F1 and F3 loaded on, was moderately explained, whereas Task 1 F1 (8%), Task 1 F3 (4%), and Task 2 F3 (9%) were not explained to a moderate degree. This is because the component scores were weighted by measures with higher loading values. However, one noticeable large difference was that PLC6 (7%), which Task 2

Table 12.10

Proportions Explained (Adjusted R^2) of the Speaking Performance Measures for Each Task by the Productive Vocabulary Knowledge Test

	Adjusted R^2 [n = 225]					Adjusted R^2 [Outliers excluded]				
Task	1	2	3	4	5	1	2	3	4	5
F1	.08	.17	.21	.26	.17	.08 (0)	.17 (0)	.25 (1)	.27 (1)	.17 (0)
F3	.04	.09	.15	.21	.14	.05 (1)	.09 (0)	.15 (1)	.22 (1)	.14 (0)
F4	.02	.09	.02	.06	.07	.03 (2)	.09 (1)	.02 (1)	.04 (2)	.08 (1)
F5	-.02	.00	-.01	-.01	.00	.02 (4)	.00 (0)	.00 (3)	-.01 (4)	.01 (2)
F6	-.02	.05	-.01	.00	.01	.01 (4)	.04 (2)	-.01 (1)	.00 (3)	.05 (3)
A1	.13	.00	.11	.09	.09	.15 (2)	.00 (0)	.11 (0)	.09 (0)	.09 (0)
A2	.13	.01	.09	.06	.09	.13 (0)	.01 (1)	.09 (0)	.06 (0)	.09 (0)
SC1	.02	.12	.01	.06	.03	.02 (0)	.14 (1)	.03 (4)	.05 (4)	.07 (3)
SC2	.02	--	.00	.02	.00	.01 (1)	--	.01 (4)	.04 (3)	.02 (7)
SC3	.02	--	.00	.03	.00	.02 (0)	--	.01 (3)	.05 (3)	.02 (7)
SC4	.06	.15	.03	.06	.06	.05 (1)	.15 (0)	.06 (2)	.06 (2)	.08 (2)
LC1	.11	.02	.13	.18	.12	.11 (0)	.02 (0)	.17 (1)	.22 (1)	.12 (0)
LC2	.04	.01	-.01	.02	.02	.03 (1)	.03 (1)	.00 (4)	.02 (2)	.03 (3)
LC3	.01	.05	.01	.04	.00	.00 (2)	.05 (4)	.01 (4)	.04 (4)	.03 (5)
LC4	.09	.16	.13	.08	.04	.09 (0)	.17 (1)	.18 (2)	.08 (0)	.03 (2)
LC5	-.01	.02	.00	.06	.01	-.01 (0)	.02 (0)	.00 (1)	.06 (1)	.01 (0)
LC6	.00	.00	.01	.01	-.01	-.01 (1)	-.01 (2)	-.01 (3)	.02 (3)	.00 (3)
LC7	.01	.07	.04	.03	.02	.01 (1)	.06 (6)	.02 (4)	.01 (5)	.03 (4)
LC8	.10	.10	.05	-.01	.02	.10 (0)	.11 (1)	.08 (2)	-.01 (0)	.02 (0)
J1001										
LC11	.00	.03	.04	.05	.02	.00 (1)	.02 (1)	.04 (3)	.07 (2)	.02 (1)
LC12	.00	.02	.01	.02	.00	.00 (1)	.02 (0)	.01 (0)	.02 (1)	.01 (2)
LC13	.00	.01	.01	-.01	-.01	.00 (1)	.01 (0)	.01 (0)	.00 (2)	-.01 (1)
LC14	.00	.00	.02	.03	.02	.00 (3)	.01 (2)	.00 (3)	.04 (3)	.04 (5)
LC15	.00	.03	.02	.04	.00	.00 (1)	.03 (0)	.02 (0)	.04 (1)	.00 (0)
J101										
LC11	.00	.14	.14	.07	.08	.00 (0)	.15 (1)	.16 (3)	.06 (1)	.08 (1)
LC12	.02	.00	-.01	.07	.00	.02 (0)	.00 (0)	.00 (1)	.07 (0)	.00 (0)
LC13	.00	-.01	.02	.01	.00	-.01 (2)	-.01 (2)	-.01 (4)	.03 (2)	.00 (2)
LC14	-.01	.05	.05	.03	.02	-.02 (2)	.07 (5)	.03 (4)	.02 (4)	.05 (4)
LC15	.00	.07	.06	.00	.04	.00 (0)	.07 (0)	.05 (1)	.00 (0)	.04 (0)

Note. ( ) = The number of outliers excluded for the analysis. (1) = 224 (i.e., 225-1) students were analyzed. Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large.

LC11J101 loaded on, and PLC11 (5%), which Task 3 LC11J101 loaded on, were not moderately explained, whereas Task 2 LC11J101 (14%) and Task 3 LC11J101 (14%) were predicted to a moderate degree. This may have been caused by lack of

relationships of other measures that loaded on PLC6 and PLC11.

There were the following three task differences in proportions of speaking performance explained (adjusted R^2) by productive vocabulary knowledge. First, in Task 1, the proportions of F1 (8%) and F3 (4%) were generally lower than those in other tasks (14 to 26% with 9% of Task 2 F3 being an exception). Second, in Task 2, the proportions of A1 (0%), A2 (1%), and LC1 (2%) were lower than those in other tasks (6 to 18%), whereas in Task 2, the proportions of SC1 (12%) and SC4 (15%) were generally higher than those in other tasks (1 to 6%). Third, in Tasks 2 and 3, the proportions of LC4 (13 to 16%) and LC11J101 (both 14%) were generally higher than those in other tasks (0 to 9%). While it is rather difficult to explain these diverging results by the task characteristics in detail, some attempts to explain were made here. Table 12.11 summarizes task characteristics of all the five tasks.

Table 12.11
Characteristics of Tasks in the Speaking Test

Task	Content	Specific content	Predicted difficulty	Familiarity	Language	Content
1	Description	Self-introduction	Easiest	Most familiar	More formulaic	Think by themselves
2	Picture comparison	Comparing pictures on the left and the right	Second easiest	Least familiar	Need to use discourse markers (which was not supplied)	Provided (in the pictures)
3	Picture description	Washing dishes	Most difficult	Second most familiar		Provided (in the picture)
4	Picture description	Riding bicycles	Most difficult	Second most familiar		Provided (in the picture)
5	Picture comparison	Comparing Taro's rooms of before and after	Third easiest	Least familiar	Need to use discourse markers (which was supplied)	Provided (in the pictures)

The first difference may be caused by two characteristics in Task 1: (a) the characteristic of eliciting more use of formulaic expressions, and (b) that of having students decide themselves what they will say.

●First, test takers engaging in the other four tasks needed to express the content of pictures and tended to retrieve words word by word. Thus, the process of retrieving productive vocabulary knowledge (in the Productive VKT and in everyday life) was similar to that of vocabulary-related speaking performance, so relationships became stronger between productive vocabulary knowledge and vocabulary-related speaking performance. In contrast, test takers conducting Task 1 tended to retrieve longer chunks of expressions they had in memory. Thus, this process is less similar, so relationships may have become weaker between productive vocabulary knowledge and vocabulary-related speaking performance. In other words, retrieval of a single word level of vocabulary may be less associated with speaking performance in tasks that elicits more formulaic expressions.

●Second, Task 1 (self-introduction) has students decide the content of the talk for themselves. In the other four tasks, since the content is provided in the pictures, students can concentrate on expressing the content, and the presented content may also work as an aid to enhance lexical retrieval. On the other hand, test takers engaging in Task 1 first need to decide what to talk about and explore expressions to convey the content, so the speaking performance may be affected more by factors other than productive vocabulary knowledge. Examples of such factors are the ability for effective assessment of what they want to express and what they can express with their knowledge and ability (i.e., strategic competence in Bachman & Palmer, 1996), and the ability to make quick decisions (i.e., personal characteristics or strategic

competence in Bachman & Palmer). As for the question of which situation is more likely in real life speaking (i.e., the case of the content is provided and that of speakers needing to decide on the content), both seem to occur in everyday life, for example giving the listener all the messages from another person and thinking of and giving important messages to the listener.

It seems that the two characteristics of Task 1 affect weaker relationships between productive vocabulary knowledge and speaking speed aspects, as measured by F1 and F3. In summary, the current study suggests that there are differences between Tasks 2 to 5 and Task 1, and that speaking performance elicited by Tasks 2 to 5 (and possibly their task types) tend to show stronger relationships between productive vocabulary knowledge and the main aspects of vocabulary-related speaking performance. It is speculated that two factors may explain the differences in these relationships. The first factor is a difference in the lexical retrieval process, that is, retrieving words word by word (i.e., a single word level of vocabulary) versus retrieving formulaic expressions. The second factor is the existence or absence of the content of the talk. If the two factors affect relationships between productive vocabulary knowledge and speaking performance, the following can be concluded. When the content of the talk is already decided and when fewer formulaic expressions are used in speaking performance, there is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types in speaking performance and better speaking performance mainly related to vocabulary.

The second difference between Task 2 and other tasks may be due to a characteristic in Task 2 of eliciting less accurate and less lexical diversity, which is

reflected in lower values of A1, A2, and LC1 (see Table 11.8). Accuracy was lower in Task 2 because, in order to compare two pictures successfully, it was necessary to use discourse markers (e.g., "in the left picture" and "in the right picture" in Task 2 and "before" and "after" in Task 5)[12.3] in order to indicate which picture test takers were talking about. When these markers were omitted, units (i.e., clauses and AS-units) that should have included the markers were judged as incorrect and the proportions of error-free units decreased. However, tendencies of Tasks 2 and 5 were different because in Task 5, the discourse markers that were required minimally (i.e., "before" and "after") were supplied along with the picture (see Appendix 7.3), so test takers used the discourse markers more often, which led to higher accuracy than in Task 2.[12.4] Lexical diversity in Task 2, as measured by LC1, was lower because the discourse markers tended to be repeated, which reduced the proportions of the number of types to the total number of tokens.

In this research, it is concluded that lack of relationships between productive vocabulary knowledge and "accuracy and LC1" in Task 2 is considered exceptional and that there is generally a moderate association between them unless tasks are very difficult. Then, a reason there is normally a moderate association needs to be sought. One explanation may be that those with less productive vocabulary knowledge may "use up available processing space in lexical searches" and cannot direct attention to accuracy (Ellis, 2005b, p. 8) and types of words they use in speaking performance. On the other hand, those with greater productive vocabulary knowledge tend to have some processing space to attend to form and variation of words that they produce in addition to searching for words that they need to use to accomplish tasks, which leads to greater accuracy and one element of lexical complexity (lexical diversity). One point to be mentioned is that productive vocabulary knowledge is less related to accuracy than to

elements of fluency, syntactic complexity, and lexical complexity, and that relationships between accuracy and attention to form, which may be made possible due to greater productive vocabulary knowledge, may be weaker than those between productive vocabulary knowledge and other vocabulary-related aspects.

As for the third difference (i.e., the higher proportions of LC4 and LC11J101 in Tasks 2 and 3 than in other tasks), the results are rather surprising because Tasks 3 and 4 were similar in difficulty and task structure (i.e., simple description of a picture). It was expected that Tasks 3 and 4 would have similar tendency. Further explanation is needed to explicate these second and third differences between tasks.

There are two other points to be noted. First, with regard to sophisticated words, there were two definitions of "sophisticated" used in this study: the criterion of considering 1,001 to 8,000 words as sophisticated (J1001) and the one of regarding 101 to 8,000 words as sophisticated (J101) in the JACET8000 (JACET Basic Words Revision Committee, 2003). Related to productive vocabulary knowledge was J101, so in relation to examining speaking performance, using the criterion of regarding 101 to 8,000 words as sophisticated may function better.

Second, the degrees of these relationships vary substantially according to the aspects of speaking performance examined and measures used, and the results were consistent with the finding seen in the previous studies (see section 2.4.2).

12.3.5 Investigating the Hypothesis 3

In this section, correlations are examined between the number of tokens and types and the component scores that were explained by productive vocabulary knowledge to a moderate or strong degree. Based on the Hypothesis 3, it was hypothesized that there would be strong correlations between the number of tokens and

types and vocabulary-related aspects of speaking performance (i.e., PF1, PSC5, and PLC1), and that there would not be strong correlations between the number of tokens and types and PA1 since the measures that loaded on PA1 were less related to vocabulary in the computing formula (i.e., A1: "the number of error-free clauses" divided by "the number of clauses").

As can be seen in Table 12.12, in relation to the number of tokens and types, there were moderate or strong correlations with PF1 (r = .79 to .88), PSC5 (r = .36 to .56), and PLC1 (r = .64 to .87) in all the five tasks, and there were not with PA1 (r = .03 to .20). Therefore, the Hypothesis 3 (There is a moderate or strong tendency for those who produce a greater number of tokens and types in speaking performance to have better speaking performance related to vocabulary across tasks.) was supported. Based on the Hypotheses 1 to 3, it is concluded that there is a moderate or strong tendency in some tasks for Japanese learners of English at the novice level who have larger and deeper productive vocabulary knowledge (than others of this level) to produce a greater number of tokens and types and better speaking performance related to vocabulary when they produce monologic description and comparison without

Table 12.12
Correlations Between the Main Component Scores and the Number of Tokens and Types

	T1 Tokens	T2 Tokens	T3 Tokens	T4 Tokens	T5 Tokens
PF1	.83**	.86**	.85**	.88**	.83**
PA1	.03	.14*	.17*	.19**	.07
PSC5	.36**	.56**	.49**	.50**	.49**
PLC1	.66**	.67**	.75**	.79**	.64**

Note. n = 225. C = Component. Underlined = .30 or more.
*p < .05. **p < .01.

	T1 Types	T2 Types	T3 Types	T4 Types	T5 Types
PF1	.79**	.80**	.83**	.84**	.81**
PA1	.09	.11	.16*	.20**	.08
PSC5	.40**	.48**	.50**	.49**	.47**
PLC1	.67**	.77**	.83**	.87**	.76**

pre-task planning time.

PF1 and PLC1 have constant strong correlations across tasks as compared to PSC5 because the formula of PF1 and PLC1 includes only the number of tokens or types but PSC5 has the number of AS-units in the denominator as well as the number of tokens in the numerator.

12.3.6 Investigating the Hypothesis 4

In this section, the relative importance of size and depth of productive vocabulary knowledge was examined using the relative Pratt index (see section 2.6.4.1). Only DVs (i.e., the number of tokens and types, component scores, and values of the Speaking Performance Measures), whose proportions were explained by productive vocabulary knowledge to a moderate or strong degree, were investigated. The criterion for considering variables unimportant was less than 0.13 (i.e., 1/2*p = 1/[2*4]; Zumbo, 2005, p. 11). Higher values of *d* were considered to show that independent variables were more related to DVs. For instance, in PF1 in Table 12.13, the value of the Antonym Subsection was the highest (.58), followed by the Size Section (.28), which demonstrates that speaking speed (PF1) was affected more by knowledge of antonyms than by vocabulary size. In addition, the positive standardized regression coefficients (β) of antonyms (.29) suggest that there are positive relationships between knowledge of antonyms and speaking speed. In other words, those with more knowledge of antonyms tend to speak faster. Table 12.14 summarizes the relative importance of size and the three subsections of depth. The results were as follows. The number of tokens and types were related the most to antonyms in the five tasks. Among the component scores, PA1 and PSC5 were related the most to size, whereas PF1 and PLC1 were related the most to antonyms. Among the Speaking

Table 12.13
Relative Importance of Size and Depth in Speaking Performance

	Adjusted R^2	Relative Pratt index				β			
		Size	Depth			Size	Depth		
			Deri	Anto	Collo		Deri	Anto	Collo
T2 Tokens	.17	0.27	0.10	0.54	0.09	.13	.05	.24	.05
T2 Types	.14	0.14	0.06	0.78	0.03	.06	.03	.31	.02
T3 Tokens	.21	0.28	0.20	0.40	0.12	.14	.11	.20	.08
T3 Types	.20	0.30	0.22	0.38	0.09	.15	.12	.19	.06
T4 Tokens	.26	0.31	0.02	0.67	0.00	.18	.01	.35	.00
T4 Types	.27	0.43	0.05	0.57	-0.05	.24	.03	.31	-.04
T5 Tokens	.17	0.23	0.01	0.80	-0.04	.11	.01	.34	-.02
T5 Types	.19	0.14	0.19	0.70	-0.03	.07	.11	.32	-.02
PF1	.23	0.28	0.07	0.58	0.07	.15	.04	.29	.05
PA1	.13	1.25	-0.28	0.00	0.03	.51	-.20	.00	.02
PSC5	.23	0.69	0.16	0.25	-0.09	.35	.09	.14	-.07
PLC1	.19	0.20	0.18	0.62	0.00	.10	.10	.29	.00
T1A1	.13	1.26	-0.27	-0.03	0.04	.50	-.19	-.02	.03
T1A2	.13	1.24	-0.29	0.03	0.02	.50	-.22	.01	.01
T2F1	.17	0.27	0.10	0.54	0.09	.13	.05	.24	.05
T2SC1	.12 (.14)	0.83	0.22	-0.23	0.18	.32	.09	-.13	.09
T2SC4	.15	0.21	0.13	0.60	0.06	.09	.06	.25	.03
T2LC4	.16	0.53	-0.17	0.50	0.15	-.24	.11	-.23	-.08
T2LC11J101	.14	0.55	-0.15	0.48	0.12	-.23	.08	-.20	-.07
T3F1	.21	0.28	0.20	0.40	0.12	.14	.11	.20	.08
T3F3	.15	0.13	0.29	0.40	0.18	.06	.13	.18	.09
T3LC1	.13	0.30	0.22	0.37	0.11	.12	.10	.15	.05
T3LC4	.13	0.47	0.05	0.48	0.00	-.19	-.03	-.19	.00
T3LC11J101	.14	0.25	0.10	0.72	-0.07	-.11	-.05	-.29	.04
T4F1	.26	0.31	0.02	0.67	0.00	.18	.01	.35	.00
T4F3	.21	0.19	0.05	0.66	0.11	.10	.03	.32	.07
T4LC1	.18	0.61	0.03	0.48	-0.11	.29	.02	.23	-.09
T5F1	.17	0.23	0.01	0.80	-0.04	.11	.01	.34	-.02
T5F3	.14	0.22	-0.10	0.91	-0.03	.10	-.06	.37	-.02

Note. n = 225. ( ) = The result when outliers were excluded. Deri = Derivation; Anto = Antonym, Collo = Collocation; Underlined = the most important variable.

Table 12.14
Order of the Importance Based on the Results of the Relative Pratt Index

Most important	Second most important	Third most important	Fourth most important	
Size				PA1 (Task 1 Accuracy); T1A1, T1A2,
Size	Antonyms			T2LC11J101, T4LC1
Size	Antonyms	Derivations		PSC5 (AS-unit length and Task 2 Syntactic complexity)
Size	Antonyms	Collocations		T2LC4
Size	Derivations	Collocations		T2SC1
Antonyms	Size			T2 Tokens, T2 Types, T4 Tokens, T4 Types, T5 Tokens; PF1 (Speaking speed); T2F1, T3LC4, T3LC11J101, T4F1, T4F3, T5F1, T5F3
Antonyms	Size	Derivations		T3 Tokens, T3 Types; PLC1 (Lexical complexity [Derived divided by the square root of No. of tokens]); T2SC4, T3F1, T3LC1
Antonyms	Derivations	Collocations	Size	T3F3
Antonyms	Derivations	Size		T5 Types

Note. From Table 12.13. T = Task.

Performance Measures for each task, Task 1 A1, Task 1 A2, Task 2 LC11J101, Task 4 LC1, Task 2 LC4, and Task 2 SC1 were related the most to size, whereas Task 2 F1, Task 3 LC4, Task 3 LC11J101, Task 4 F1, Task 4 F3, Task 5 F1, Task 5 F3, Task 2 SC4, Task 3 F1, Task 3 LC1, and Task 3 F3 were related the most to antonyms. Overall, among the three depth aspects, antonyms appeared as the first factor to influence the DVs, followed by derivations and collocations. The exception was Task 2 SC1 in which the relative importance was in the order of size, derivations, and collocations. The results of the standardized regression coefficients (β) were generally consistent with those using the relative Pratt index (d). The only difference between the two (β and d) was Task 3 LC4, in which the relative Pratt index suggests that antonyms are related more than size, whereas the standardized regression coefficient is the same

value for antonyms and size.

Overall, the results were that the number of types and tokens, speaking speed (PF1), and one element of lexical complexity (PLC1) were related more to knowledge of antonyms than to vocabulary size and knowledge of derivations or collocations, whereas Task 1 accuracy (PA1) and one element of syntactic complexity (PSC5) were associated more with vocabulary size than with vocabulary depth.

Additionally, the standardized regression coefficients (β) of the most important aspect of vocabulary knowledge were negative in Task 2 LC4 and Task 3 LC4 (the number of lexical word types per token) and Task 2 LC11J101 and Task 3 LC11J101 (the number of sophisticated word types per token [J101]). This indicates that lexical density and lexical sophistication (J101) were negatively related to productive vocabulary knowledge. In other words, those with more productive vocabulary knowledge tend to use more more grammatical words (i.e., function words) and non-sophisticated words (i.e., very basic words) that belong to the most frequent word group of up to 100 words. This suggests that in speaking, using more grammatical and basic words may be more important than trying to use more lexical and sophisticated words.

When results using the component scores and those using the Speaking Performance Measures for each task were compared, the general trend was the same. However, there were two main differences. First, the order of importance was size first and antonyms second in PSC5, whereas this was the opposite in Task 2 SC4, which mainly loaded on PSC5. Second, the order of importance was antonyms first and size second in PLC1, whereas this was the opposite in Task 4 LC1, which mainly loaded on PLC1. These differences may suggest that the relative importance varies to some degree according to the target of the investigation. Thus, as for elements of syntactic

complexity and lexical complexity, both size and antonyms may be important, and this conclusion is used for the interpretation in this study (2E). Thus, the interpretation of the results were that Task 1 accuracy (PA1) was related the most to vocabulary size, whereas the number of tokens and types and speaking speed (PF1) were associated the most with knowledge of antonyms. One element of lexical complexity and that of syntactic complexity were affected the most both by knowledge of antonyms as well as vocabulary size. Thus, the Hypothesis 4 (Size of productive vocabulary knowledge is related more to speaking performance than depth of productive vocabulary knowledge is.) was supported for accuracy in one task, and not supported for the number of tokens and types and speaking speed, and partially supported for one element of lexical complexity and that of syntactic complexity. It is concluded that size of productive vocabulary knowledge is related more to accuracy in a task, and that at least one aspect of depth (antonyms, or word connections) is related more to one element of fluency (speaking speed) than size is. It is also shown that both size and antonyms of productive vocabulary knowledge are related to one element of syntactic complexity and one element of lexical complexity. These relationships are likely to be found when Japanese learners of English at the novice level produce monologic description and comparison without pre-task planning time.

The results that antonyms of productive vocabulary knowledge are related more to one element of fluency than size in monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level suggest that at least one aspect of depth is more important than size in an element of speaking performance, which was contrary to Meara (1996), who stated that for those who have small vocabulary, size is the only and the most important dimension (p. 45). However, this study demonstrates the importance of antonyms at least for some elements of

speaking performance (speaking speed, one element of lexical complexity, and one element of syntactic complexity) of Japanese learners of English at the novice level.

It should be noted that the method of analysis (i.e., having three subsections of depth with different total marks for each) did not affect the results of depth having more importance. There were two reasons for this. First, the relative Pratt index is standardized because β and r, which were used for computing d, are both standardized. Second, the three subsections were not combined and analyzed and the three depth subsections were independently entered in the equation. In addition, the relative differences in reliability do not seem to affect the results because the reliability of the Size Section was higher (.92; see Table 12.6) than that of the three depth subsections (.73 to .80), and if the higher reliability of the Size Section affected the results, those of the correlations and MRA would show stronger relationships with size of productive vocabulary knowledge, but some of the results were opposite.

Some may wonder whether if size were entered first in the equation, followed by depth, using the sequential method of MRA, the results would be different because there were strong correlations between size and depth (r = .62 to .81; see Table 12.15) and especially a strong correlation was found between size and antonyms (r = .81).

Table 12.15
Correlations Between Size and Depth (Derivations, Antonyms, and Collocations)

	Size	Derivation	Antonym	Collocation
Size	--	.78**	.81**	.67**
Derivation		--	.69**	.64**
Antonym			--	.62**

Note. n = 225.
**p < .01.

Table 12.16 shows the results when this analysis was conducted. As a result,

Table 12.16
Proportions Explained (Adjusted R^2) of the Number of Tokens and Types, the Component Scores, the Speaking Performance Measures for Each Task by the Productive Vocabulary Knowledge Using Sequential Regression Method

	Adjusted R^2	Size first Adjusted R^2	Depth second Adjusted R^2	Depth first Adjusted R^2	Size second Adjusted R^2
T2 Tokens	.17	.15	.02	.17	.00
T2 Types	.14	.12	.02	.14	.00
T3 Tokens	.21	.19	.02	.21	.00
T3 Types	.20	.18	.02	.19	.00
T4 Tokens	.26	.22	.03	.25	.00
T4 Types	.27	.24	.02	.26	.01
T5 Tokens	.17	.14	.03	.17	.00
T5 Types	.19	.16	.03	.19	.00
PF1	.23	.20	.03	.23	.00
PA1	.13	.12	.00	.07	.06
PSC5	.23	.23	.00	.20	.03
PLC1	.19	.16	.03	.19	.00
T1A1	.13	.12	.00	.07	.06
T1A2	.13	.12	.01	.07	.06
T2F1	.17	.15	.02	.17	.00
T2SC1	.12 (.14)	.12	.00	.10	.02
T2SC4	.15	.13	.02	.15	.00
T2LC4	.16	.15	.01	.15	.01
T2LC11J101	.14	.13	.01	.13	.01
T3F1	.21	.19	.02	.21	.00
T3F3	.15	.13	.02	.15	.00
T3LC1	.13	.12	.01	.13	.00
T3LC4	.13	.13	.00	.12	.00
T3LC11J101	.14	.12	.02	.14	.00
T4F1	.26	.22	.03	.25	.00
T4F3	.21	.18	.03	.21	.00
T4LC1	.18	.17	.01	.17	.02
T5F1	.17	.14	.03	.17	.00
T5F3	.14	.11	.03	.14	.00

Note. n = 225. ( ) = The result when outliers were excluded.

when size was entered first, most of the proportions were explained by size and the rest were 0 to 3%. This suggests the small addition of depth and a small depth contribution. However, for the following two reasons, depth among beginners was considered essential. First, when depth was entered first in the equation, followed by size, using the sequential method of MRA, most of the proportions were explained by depth, and

the rest were 0 to 3% except for PA1, T1A1, and T1A2 (6%). The high proportions show the importance of depth itself. Second, the rationale of putting size first in the equation needs logical or theoretical justification (Tabachnick & Fidell, 2001, p. 132). However, although the greater importance of size over depth is suggested by Meara (1996), this is not a sufficient rationale for using the sequential method in this study because it has not been empirically tested. In addition, entering size first into the equation may make the adoption of the Hypothesis 4 easier, and testing the relative importance without assuming the order seemed more appropriate. Thus, the results in Table 12.13, rather than those in Table 12.16 were used for the main interpretation.

While it was expected that size would explain some elements of speaking performance, the reasons antonyms explained speaking speed more than size, derivations, and collocations were explored. Since Nation (2001) suggests the importance of collocations in speaking performance (pp. 317-318), it was expected that knowledge of collocations would affect speaking performance more than that of antonyms and derivations. However, antonyms turned out to be more important than collocations for two reasons. First, in Tasks 2 and 5, in which two pictures were compared, knowledge of antonyms was required, which may have increased the importance of antonyms. However, the importance of antonyms was seen in the other three tasks. There also appears to be another reason, as explained below.

Second, the definition of collocations used in this study could have caused these results. As explained in section 7.2.2.1.2.3, collocations were defined as words occurring together, and aspects of "varying degrees of exclusivity" (Schmitt, 2000, p. 77) were not included in the definition in this study. Accordingly, in making items, stimulus words were selected based on the definition, and in scoring responses from test takers, they were scored as correct unless they were wrong, unacceptable,

redundant, unusual, unlikely, or weird as a combination of words (see section 9.2.3.2.1). Thus, collocations in this study were wider in definition than normally considered, and the test scores derived from the Collocation Subsection may not have assessed the collocations in a technical sense. This may be why collocations did not affect speaking performance more than other elements (e.g., antonyms) did. Thus, this study may not be able to conclude that knowledge of collocations is less important than knowledge of antonyms in speaking performance. One point to be added is that making test items of collocations to assess knowledge of words occurring together with "varying degrees of exclusivity" (Schmitt, p. 77) may be rather difficult when test takers include junior high school students because words they have learned may not include many words with "varying degrees of exclusivity" (Schmitt, p. 77) that match the definition of collocations.

If the second reason is adopted in the interpretation, the interpretation in the current study is done using the results of size, derivations, and antonyms. If knowledge of antonyms is considered to be knowledge of word associations and knowledge of derivations is considered to be knowledge of word forms, this study (2E) seems to suggest that in speaking speed, word connections are more important than size and knowledge of word forms. In other words, in order to speak faster, one element of depth of productive vocabulary knowledge, that is, knowledge of connections between words, is needed more than vocabulary size and knowledge of word forms. This may also suggest that in order to speak faster, once a minimum level of vocabulary size is gained, increasing knowledge of semantic connections between words should be the main concern. Regarding the minimum level of vocabulary size, the effective criterion of 100 words first comes to mind, but this is the criterion of words that appear in speaking performance. Thus, the minimum requirement of vocabulary size that

underlies speaking performance seems to be much larger. Further examination is needed to examine this criterion.

One explanation of the importance of word links can be derived from Aitchison (2003), who states that in speaking performance, "it is normal for the mind to activate many more words than are likely to be used in the course of a conversation" (p. 219). After meaning and form are activated and suppressed, appropriate words are selected to be used for speaking (pp. 224-226). Those who tend to have less knowledge of word connections may not be able to activate words adequately. This may lead to failure in continuing talk or to slower speaking speed.

The last thing to be added is that it was found that although some proportions of some elements of speaking performance were explained by productive vocabulary knowledge, the maximum proportion possible was approximately 30% (see 31% of Task 4 Types Adjusted R^2 [Outliers excluded] in Table 12.7), which indicates that productive vocabulary knowledge is related to some elements of speaking performance but that contributions of factors other than productive vocabulary knowledge are larger, which supports Buck (2001), who stated that "in language use, declarative knowledge is of very limited value" (p. 14). It should be noted that these results do not mean that productive vocabulary knowledge is not very important in speaking performance. As it is impossible to speak without productive vocabulary knowledge, productive vocabulary knowledge is an essential requirement of speaking performance. However, this study indicates that when actual speaking is done, factors other than productive vocabulary knowledge affect speaking performance more.

12.3.7 Investigating the Research Question

In the previous section, the degree to which speaking performance can be

explained by size and depth of productive vocabulary knowledge was examined. It was found that the maximum proportion was approximately 30%, which indicates that effects of factors other than productive vocabulary knowledge are larger. The next question is what factors, other than productive vocabulary knowledge, explain the other 70%. This was explored using 11 questionnaire items requiring affective and cognitive individual responses. Table 12.17 shows the items used for this analysis.

Table 12.17
Items on the Speaking Test Questionnaire

Part 1. Please choose the number that describes your feelings the most.
0. I don't know. 1. No, I don't think so at all. 2. No, I don't think so very much.
3. Neither yes nor no. 4. Yes, I think so a little. 5. Yes, I think so very much.
Example: Do you often play computer games?
→When you do not play with computer games at all Example [1]
Q1 (2-1) Do you often speak with your family and friends in Japanese?
Q2 (2-2) Have you spoken English during classes and outside classes?
Q3 (2-3) Have you had much experience in taking tests in which you speak English, either at school or outside school?
Q4 (2-9) Were you worried or nervous during the test?
Q5 (2-10) Were sound and people nearby distracting during the test?
Q6 (2-11) Did you want to speak much during the test?
Q7 (2-12) During the test, did you want to use many grammatical items that you have studied so far?
Q8 (2-13) During the test, did you want to use many words that you have studied so far?
Q9 (2-14) Did you want to speak without making many errors during the test?
Q10 (2-15) Did you want to speak fluently during the test?
Q11 (2-16) Did you want to use memorized sentences during the test?

Analyses were conducted after removal of the results of test takers who chose the option "I don't know" in the questionnaire. As has been done in section 12.3.1, assumptions were checked and then MRA was done. This was because the number of participants changed (n = 180 in this analysis) since some test takers did not answer some questionnaire items. This analysis was conducted using "size and depth of productive vocabulary knowledge and responses toward questionnaire items" as IVs

and “component scores of speaking performance” as DVs.

There are six assumptions to be checked before MRA (Tabachnick & Fidell, 2001, pp. 116-122). The same procedures were taken as described in section 12.3.1. First, regarding the ratio of cases to IV, a sample size of 180 was sufficiently large for testing multiple correlations as well as testing individual predictors. Second, regarding outliers among the IVs and on the DV, there were 12 extreme univariate outliers. Thus, 168 participants were retained (i.e., 180 – 12). There were no extreme multivariate outliers. Third, concerning the absence of multicollinearity and singularity, the results showed no evidence of multicollinearity. Fourth, the examination of residual scatterplots showed that there was no evidence of violation of normality, linearity, and homoscedasticity. Fifth, with regard to independence of errors, there was no evidence of non-independence of errors. Sixth, there were 12 outliers in total. Thus, the two analyses were done by using 168 students and by excluding outliers for each MRA.

Table 12.18 shows (a) the total adjusted R^2 (proportions explained by productive vocabulary knowledge and questionnaire items), (b) adjusted R^2 explained by only productive vocabulary knowledge, and (c) adjusted R^2 with questionnaire items (i.e., (a) – (b)). In the analysis in this section, sequential MRA was used because Research Question (Are there any individual factors other than productive vocabulary knowledge related to speaking performance to a moderate or strong degree?) included the order of entering IVs. As a result, the proportions explained only by questionnaire items (i.e., (c)) were less than 13% (Cohen, 1988; see section 2.6.4.1), which suggests that there were no moderate or strong effects of individual factors that could be assessed by the Speaking Test Questionnaire even when 11 factors were put in the equation at the same time. Thus, concerning the Research Question, it is demonstrated that there are no individual factors other than productive vocabulary knowledge that

Table 12.18
Proportions Explained (Adjusted R^2) of the Component Scores by the Productive Vocabulary Knowledge Test and the Questionnaire Items

	Adjusted R^2 [n = 168]			Adjusted R^2 [Outliers excluded]		
	(a) Total	(b) PVK	(c) Question-naire	(a) Total	(b) PVK	(c) Question-naire
PF1	.30	.22	.08	.30 (0)	.22	.08
PF2	.07	-.01	.08	.08 (1)	-.02	.10
PF3	.10	.04	.07	.10 (0)	.04	.07
PF4	-.03	-.01	-.02	-.03 (1)	-.01	-.03
PF5	-.08	-.02	-.06	-.07 (1)	-.01	-.05
PF6	.01	.02	-.02	.01 (0)	.02	-.02
PA1	.12	.14	-.02	.12 (0)	.14	-.02
PA2	.16	.10	.06	.16 (0)	.10	.06
PA3	.15	.09	.06	.15 (0)	.09	.06
PA4	.22	.11	.10	.22 (0)	.11	.10
PA5	-.01	.00	-.01	-.01 (0)	.00	-.01
PSC1	.01	.02	-.01	.01 (0)	.02	-.01
PSC2	.04	.02	.02	.04 (0)	.02	.02
PSC3	.00	.02	-.02	.08 (4)	.06	.01
PSC4	.07	.03	.04	.07 (0)	.03	.04
PSC5	.23	.22	.01	.23 (0)	.22	.01
PLC1	.21	.18	.03	.21 (0)	.18	.03
PLC2	-.02	-.01	-.01	-.02 (1)	-.01	-.01
PLC3	.04	.00	.04	.04 (0)	.00	.04
PLC4	-.01	.04	-.05	.03 (2)	.03	.00
PLC5	.01	.03	-.02	.01 (0)	.03	-.02
PLC6	.10	.07	.03	.10 (0)	.07	.03
PLC7	.03	.02	.01	.03 (0)	.02	.01
PLC8	.05	.03	.02	.05 (0)	.03	.02
PLC9	.05	.03	.02	.07 (1)	.03	.03
PLC10	.03	.01	.02	.03 (0)	.01	.02
PLC11	.03	.06	-.03	.03 (0)	.06	-.03
PLC12	.02	.05	-.02	.02 (0)	.05	-.02

Note. ( ) = The number of outliers excluded for the analysis. PVK = productive vocabulary knowledge. Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large.

can be assessed by the questionnaire items and associated with speaking performance to a moderate or strong degree.

It is also shown that the maximum addition by the questionnaire items was 10% in PA4 (Task 5 Accuracy). As far as the questionnaire items could assess, the L1 and L2 speaking experience (Q1 and Q2) and experience of taking a speaking test (Q3) were not closely related.

The highest proportion explained in total was 30% in speaking speed, which is less than half of speaking speed. Therefore, what is related to the rest of the proportions (70% or more)? This is a question beyond this study, but possible factors were outlined for the future research based on Bachman and Palmer's (1996) framework (see sections 2.1.1 and 2.1.2 for the terms). It should be noted that measurement errors are always involved, as is true of any study.

First, the components of language ability are considered. In language knowledge, other aspects that affected speaking performance apart from the productive vocabulary knowledge included (a) aspects of vocabulary knowledge that are not included in the current study (e.g., grammatical functions and constraints on use, procedural knowledge), (b) organizational knowledge other than vocabulary knowledge (e.g., knowledge of syntax and textual knowledge), (c) pragmatic knowledge, and (d) strategic competence (see Dörnyei & Kormos, 1998 for strategies needed to encounter difficulty in speaking performance). Furthermore, L1 speaking ability may be related.

Second, regarding characteristics of test takers other than language ability, speaking performance may be affected by (e) personal characteristics (e.g., age, native language, and culture), (f) topical knowledge, and (g) affective schemata. Although also some aspects were assessed in the Speaking Test Questionnaire and every attempt was made to exclude these negative effects in the test construction stage, there may be other aspects. Other examples may include the experience of conducting similar tasks (especially picture description tasks and comparison tasks), which was not covered in the questionnaire, and creativity (Albert & Kormos, 2004), extraversion, openness (Verhoeven & Vermeer, 2002), "capacity and/or efficiency in the STM [short-term memory] and WM [working memory]," and anxiety (Dewaele, 2002, pp. 240-241)

Third, with regard to characteristics of language use tasks and test tasks, the

tasks used in this study may have affected the results, though such effects were described and discussed in section 12.3.4.

Among the possible factors, the crucial factors may be (a) aspects of vocabulary knowledge that are not included in the current study. In particular, this study dealt with declarative knowledge of vocabulary, but procedural knowledge of vocabulary that enables automatic speaking processing may affect speaking performance more. In addition, this study dealt with knowledge of individual words, but knowledge of multi-word units (see Moon, 1997, Read, 2000, 2005; Schmitt, 2004; Wray, 2002) may affect speaking performance more.

12.4 General Discussion

In this section, after the results of Studies 1C and 2E were compared, the results of the current study and the previous studies (see section 2.4.2) were compared. Then, relationships between productive vocabulary knowledge and speaking performance are discussed more generally than in section 12.3.

Table 12.19 shows results of 13 common measures used both in Studies 1C and 2E. In the two studies, Tasks 1 and 5 were basically the same (with minor differences; see section 7.2.2.2). The results of Studies 1C and 2E were very similar in that F5, F6, SC1, SC2, and LC4 were not explained by productive vocabulary knowledge to a moderate or strong degree. In the explanation, the names of the measures used in Study 1C were used and those used in Study 2E were put in the parenthesis [ ] if the names were different.

Table 12.19
Comparisons Between Proportions Explained (Adjusted R^2) of the Speaking Performance Measures by the Productive Vocabulary Knowledge Test

	Study 1C			Study 2E				
	Task 1	Task 4	Task 5	Task 1	Task 2	Task 3	Task 4	Task 5
F1 (No. of tokens per minute)	-.01	.00	.05	.08	.17	.21	.26	.17
F3 (No. of clauses per minute)	-.01	-.01	-.01	.04	.09	.15	.21	.14
F4 (No. of dysfluency markers per minute)	.17	.06	.11	.02	.09	.02	.06	.07
F5 (No. of dysfluency markers per token)	.03	-.01	-.01	-.02	.00	-.01	-.01	.00
F6 (No. of dysfluency markers per AS-unit)	.10	.09	.00	-.02	.05	-.01	.00	.01
A1 (No. of error-free clauses per clause)	.01	.11	.20	.13	.00	.11	.09	.09
A2 (No. of error-free AS-units per AS-unit)	.00	.01	.23	.13	.01	.09	.06	.09
SC1 (No. of clauses per AS-unit)	.04	.09	.06	.02	.12	.01	.06	.03
SC2 (No. of subordinate clauses per clause)	.01	.10	-.01	.02	--	.00	.02	.00
SC3 [SC4] (No. of tokens per AS-unit)	.33	.28	.41	.06	.15	.03	.06	.06
LC1 (No. of types divided by the square root of No. of tokens)	.26	.41	.21	.11	.02	.13	.18	.12
LC2 (No. of lexical tokens per token) [LC4 (No. of lexical word types per token)]	.14	.08	.12	.09	.16	.13	.08	.04
LC4 [LC11J1001] (No. of sophisticated word types per token)	-.01	-.01	.02	.00	.03	.04	.05	.02

Note. [ ] = the name of the measure used in Study 2E; -- = Task 2 SC2 was not used for the analysis (see section 11.3.1); Underlined = .13 or more. The interpretation of adjusted R^2 based on Cohen (1988, pp. 413-414): .0196 ≤ small < .13; .13 ≤ medium < .26; .26 ≤ large.

When medium or strong effects were examined, three differences were observed in Tasks 1 and 5 between Studies 1C and 2E. First, in Study 1C, there was not a moderate relationship between productive vocabulary knowledge and F1 (the number of tokens per minute) in Task 5, which was contrary to the results in Study 2E. In Study 1C, F3 (the number of clauses per minute) was not related to productive vocabulary knowledge in Task 5 to a moderate or strong degree, whereas in Study 2E, it was related to a moderate degree. In Study 1C, there was a moderate relationship

between speaking performance and F4 (the number of dysfluency markers per minute) in Task 1, but in Study 2E, F4 was not related in Task 1 to a moderate or strong degree. These differences may be attributed to the different method of measuring speaking time. In Study 1C, a task finished when test takers did not speak for 15 seconds for a pedagogical reason. Therefore, the speaking time varied substantially between test takers, so the denominators used for computing the measures were quite different from each other. However, in Study 2E, the same speaking time (45 seconds) was used for computing F1, F3, and F4. Since the results of 2E were more consistent with the Hypothesis 2 (There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to have better speaking performance related to vocabulary across tasks.). It is concluded that the measurement of speaking time in Study 2E is better than that in Study 1C.

Second, moderate or strong relationships (R^2 = .13 or more) appeared in different measures. In Study 1C, A1 and A2 were related to productive vocabulary knowledge in Task 5 to a moderate degree, whereas in Study 2E, they were related to productive vocabulary knowledge in Task 1 to a moderate degree. In Study 1C, SC3 [SC4 in Study 2E] was related to productive vocabulary knowledge in Tasks 1 and 5 to a strong degree, whereas in Study 2E, it was not related to productive vocabulary knowledge in Tasks 1 and 5 to a moderate or strong degree. In Study 1C, LC1 was related to productive vocabulary knowledge both in Tasks 1 and 5 to a moderate or strong degree, whereas in Study 2E, it was not related in Tasks 1 and 5 to a moderate or strong degree. In Study 1C, LC2 [LC4 in Study 2E] was related to productive vocabulary knowledge in Task 1 to a moderate degree, whereas in Study 2E, it was not in Task 1 to a moderate or strong degree. The differences of LC2 may not be very crucial because the proportions explained were rather similar, however. These

differences may have been derived from the difference in measuring speaking time because it may affect the number of tokens, AS-units, and clauses produced, which leads to different results of the measures.

Third, when the degree of proportions was examined, the proportions in Study 1C were somewhat higher than those in Study 2E in LC2 [LC4 in Study 2E] in Task 1, A1 and A2 in Task 5, and SC3 [SC4 in Study 2E] and LC1 in Tasks 1 and 5. However, the tendency was contrary to A1 and A2 in Task 1. In fact, the proportion of SC3 [SC4 in Study 2E] was 41% and the conclusion reached in sections 12.3.6 and 12.3.7 that the highest proportion explained in total was 30% needs to be modified. It is concluded that the highest proportion of speaking performance explained by productive vocabulary knowledge was approximately 40%. The tendency of Study 1C to have higher proportions may be explained by a difference in time requirement between Studies 1C and 2E. In Study 1C, 90 seconds were given for Tasks 1 and 5 and test takers with higher speaking ability could show their ability or productive vocabulary knowledge more fully than in Study 2E, in which 60 seconds were given for Tasks 1 and 5. In other words, tasks in Study 1C revealed more variations between test takers, which led to more proportions being explained, as described previously. In other words, if test takers speak for a longer time, their ability is more reflected in their utterances, and relationships between productive vocabulary knowledge and speaking performance tend to become stronger. Exceptions were seen in A1 and A2 in Task 1, in which values in Study 1C were lower than in Study 2E. This may be attributed to differences in (a) test formats (face-to-face format in Study 1C vs. tape-mediated format in Study 2E), and (b) task characteristics. The face-to-face format in Study 1C enabled an interviewer in front of a test taker to give non-verbal feedback (e.g., nodding). In addition, even when the interviewer does not give any feedback at all, test

takers may think that the interviewer's silence indicates comprehension of what they are saying. In contrast, in the tape-mediated format in Study 2E, a test taker may wonder if their utterances were sufficiently comprehensive and those who have more productive vocabulary knowledge and have some processing space in attending to forms and accuracy attempt to speak more accurately. This tendency of accuracy in the tape-mediated format being higher than the face-to-face format was consistent with Shohamy's (1994) finding that a tape-mediated format elicits more self-correction than a face-to-face version because test takers "pay more attention to linguistic accuracy" in the former (p. 115). This attention to accuracy may be possible only with tasks like Task 1 (see Table 12.11 for the characteristics of Task 1). Among characteristics of Task 1, the key factor may be the possibility of using more formulaic expressions because it may reduce the burden of the speaking process and provide processing space to attend to accuracy. It should be noted that this explanation has "the assumption of a limited capacity, that is, there are limits on the amount of information that human beings can process from input or for output" (Ellis, 2005b, p. 6). The results throughout the current study suggest that relationships between productive vocabulary knowledge and speaking performance tend to vary according to measures, tasks, and test formats (i.e., face-to-face vs. tape-mediated) used.

Next, the results of this study and the previous studies were compared, with special focus on Koizumi (2005d) and Ukrainetz and Blomquist (2002), which dealt with relationships between productive vocabulary knowledge and speaking performance and used the same or very similar measures as that used in the present study (see section 2.4.2).

Table 12.20 shows a summary of essential features and the results of the two studies above and the current study. When the proportions explained by productive

Table 12.20
Comparisons Between the Current Study and the Previous Studies

	Tasks to elicit productive vocabulary knowledge	Elements of speaking performance targeted	r (R^2)
The present study: Study 2E	Size Section (Size): The written translation format: in which a L1 meaning is presented to elicit an equivalent L2 word. The first letter of the word was provided. Depth Section (Depth): The written format to elicit forms of derivations, antonyms, collocations of a word.	No. of tokens	(.08 to .26)
Ukrainetz & Blomquist (2002)	A: The examiner pointed to a picture silently or with a word label. The child labelled the picture or provided a synonym for the word. Elicited was a single word (nouns, verbs, or adjectives; p. 65).	No. of tokens	.25 (.06)
Ukrainetz & Blomquist (2002)	B: The child labelled or provided a category name for the items represented. Elicited was a single word (p. 66).	No. of tokens	.32 (.10)
The present study: Study 2E	Size Depth	No. of types	(.10 to .31)
Koizumi (2005d)	The same as the Size Section in the present study	No. of types	.65 (.42)
Ukrainetz & Blomquist (2002)	A	No. of lemmas [types]	.48 (.23)
Ukrainetz & Blomquist (2002)	B	No. of lemmas [types]	.46 (.21)
The present study: Study 1C	Size	No. of tokens per AS-unit [SC4]	(.28 to .41)
The present study: Study 2E	Size Depth	No. of tokens per AS-unit [SC4]	(.03 to .15)
Ukrainetz & Blomquist (2002)	A	The mean length of utterance [SC4]	.29 (.08)
Ukrainetz & Blomquist (2002)	B	The mean length of utterance [SC4]	.36 (.13)
The present study: Study 1C	Size	No. of types/√(No. of tokens) [LC1]	(.21 to .41)
Koizumi (2005d)	The same as the Size Section in the present study	No. of types/√(No. of tokens) [LC1]	.62 (.38)
The present study: Study 2E	Size Depth	No. of types/√(No. of tokens) [LC1]	(.02 to .22)
The present study: Study 1C	Size	No. of lexical types/√(No. of tokens) [LC4]	(.08 to .14)
Koizumi (2005d)	The same as the Size Section in the present study	No. of lexical tokens /√(No. of tokens) [LC4]	-.46 (.21)
The present study: Study 2E	Size Depth	No. of lexical types/√(No. of tokens) [LC4]	(.03 to .18)

Note. Only results of comparable measures were presented. [ ] = The name of a measure in Study 2E in the current study; See Table 6.4 for R^2 in Study 1C and Table 12.10 for R^2 in Study 2E.

vocabulary knowledge were compared, three points needed to be mentioned. First, Study 1C and Koizumi (2005d) produced similar results. For example, in Study 1C, the proportions explained of LC1 (the Guiraud index: the number of types divided by the square root of the number of tokens) were .21 to .41 and the one in Koizumi belonged to that range (.38). In fact, this is natural because Koizumi used participants overlapping with the present study. Some of the participants in Koizumi were the same, but the participants in the two studies were different in that Koizumi included those who did not utter clauses as well. The similar results indicate that this difference did not affect the results to a large degree.

Second, the proportions explained in Study 1C (e.g., R^2 = .21 to .41) in LC1 (the Guiraud index: the number of types divided by the square root of the number of tokens) and Koizumi (2005d; R^2 = .38) were higher than those in Study 2E (.02 to .22). This difference may be due to the differences in time requirement of 90 seconds vs. 60 seconds (see above in this section).

Third, the proportions explained in Study 2E and Ukrainetz and Blomquist (2002) were similar, and the results of Ukrainetz and Blomquist's study (e.g., .06 to .10 in the number of tokens) were within the range of the proportions (.08 to .26) in Study 2E. Between the two studies, there were some differences (e.g., L2 learners [junior and senior high school students] in Study 2E vs. L1 learners [children] in Ukrainetz & Blomquist; monologues with tape-mediated recording in Study 2E vs. conversational, narrative, and expository discourse in Ukrainetz & Blomquist). Given these differences, the results of these studies (i.e., presence and absence of moderate or strong relationships between productive vocabulary knowledge and lexical elements [i.e., the number of tokens and types (or lemmas)]) may be more generalized, although further

research is needed.

The comparison between the current study and the previous studies overall leads to the conclusion that the results of moderate or strong relationships between productive vocabulary knowledge and vocabulary-related speaking performance are generally consistent with each other. However, the comparison between Studies 1C and 2E indicates that the measurement of speaking time, a difference in time requirement (e.g., 90 seconds vs. 60 seconds), test formats (e.g., face-to-face format vs. tape-mediated format), and task characteristics (e.g., self-introduction vs. picture description and comparison) affect the relationships as well. The next question is to what degree the findings in the current study can be generalized. While productive vocabulary knowledge that was dealt with in the current study seems to cover a relatively wide area of size and depth, generalizability regarding speaking performance needs to be carefully examined because this study investigated a rather narrow type of speaking performance: "description and comparison" in the "transactional function" and "monologues without pre-task planning time."

First, of the three types of speaking function (i.e., interactional, transactional, and ludic; Tarone, 2005; p. 486; see Table 12.21), the "transactional" function was the focus of this study. Speaking performance in the "interactional" and "ludic" functions is expected to be different from that in the "transactional" function in terms of the content and the way of expressing it. Since the purpose of interactional speaking is to "establish or maintain social relationships" (Tarone, p. 486), in order to achieve this purpose appropriate content and the way to express it need to be considered in more detail, depending on who is the listener. Since the cognitive load of such a process is predicted to be heavier and it may be too difficult for learners at the novice level to tackle this function from scratch, the speaking performance is expected to (a) fall on

more formulaic expressions and/or (b) become less varied among novice learners because of the increased difficulty. The use of more formulaic phrases and the process of producing speaking performance in the "interactional function" and "ludic function" may be similar to those of self-introduction (Task 1 in the current study) because in self-introduction, appropriate content is also selected and an appropriate way is used with the listener (i.e., a person who the speaker meets for the first time) in mind. If this is the case, the results of Task 1 (see section 12.3.4) suggest that in speaking performance in the "interactional function" and "ludic function," relationships may become weaker between productive vocabulary knowledge and vocabulary-related speaking performance. The less varied speaking performance among novice learners may also make relationships between productive vocabulary knowledge and speaking performance weaker. The weaker associations may appear because factors other than productive vocabulary knowledge tend to affect speaking performance more, and because the process of retrieving productive vocabulary knowledge and vocabulary-related speaking performance is expected to be less similar.

Table 12.21
Functions of Speaking Performance

Function	Further division	
Interactional		
Transactional	Factually oriented talk:	description, narration, instruction, comparison
	Evaluative talk:	explanation, justification, prediction, decision
Ludic		

Note. The same table as Table 2.19. From Tarone (2005), Bygate (1987), and Luoma (2004, pp. 31-32).

Second, the "transactional function" is divided into "factually oriented talk" and "evaluative talk." The former was selected mainly because it is basically based on "factually oriented talk" (Bygate, 1987, p. 24; see section 2.3.1). In contrast, to

produce “evaluative talk,” speakers need to have the integrated structures to covey information to a listener in order to produce the whole argument, and “evaluative talk” is expected to be more difficult than “factually oriented talk.” As a result, novice learners are not likely to be able to accomplish this task, which leads to low variations in speaking performance and weaker relationships between productive vocabulary knowledge and speaking performance. It seems unlikely that more formulaic sentences would be used in “evaluative talk” since novice learners may not know formulaic expressions that can be used for this purpose.

Third, “factually oriented talk” is separated into four types (i.e., description, narration, instruction, and comparison), and the current study used “description” and “comparison.” Since “narration” is generally more difficult than “(simple) description” (Brown, T. & Yule, 1983, p. 107), “narration” derived from novice level learners is expected to produce low variations in speaking performance, which may result in weaker relationships between productive vocabulary knowledge and speaking performance. “Instruction” is predicted to produce similar speaking performance to description (Brown, T. & Yule, p. 46), and relationships between productive vocabulary knowledge and speaking performance are expected to be similar.

Fourth, speaking performance is divided into monologues and dialogues depending on whether there is interaction between the speaker and the listener. In this study, “monologues” were selected. Since “dialogues” tend to be more affected by external factors (e.g., interviewer’s reactions, Brown, A., 2003, 2004), relationships in “dialogues” between internal factors (e.g., productive vocabulary knowledge) and speaking performance are expected to be weaker.

Fifth, speaking performance is separated into utterances with planning time (i.e., prepared speech) and without planning time (i.e., spontaneous speech). Strictly

speaking, there are two types of planning: planning conducted before starting speaking (i.e., pre-task planning) and planning conducted during speaking (i.e., on-line or within-task planning; see Ellis, 2005c; Kawauchi, 2005). In the present study, "speaking performance without pre-task planning" was selected. In "speaking performance with (both pre-task and on-line) planning," it is expected that the process of retrieving productive vocabulary knowledge will be similar to that of producing words in speaking performance unless more formulaic expressions are used. Then, stronger relationships between productive vocabulary knowledge and speaking performance may be expected.

Table 12.22 shows the summary of predicted results for types of speaking performance not examined in this study. Since these predictions are mostly based on

Table 12.22
Predicted Relationships Between Productive Vocabulary Knowledge and Speaking Performance

Types of speaking performance	Predicted result of relationships between productive vocabulary knowledge and speaking performance
Interactional	Weaker relationships than in the general "transactional" function (similar relationships to self-introduction (Task 1) of the "transactional" function)
Transactional	(Examined in the current study)
Ludic	Weaker relationships than in the general "transactional" function (similar relationships to self-introduction (Task 1) of the "transactional" function)
Factually oriented talk	(Examined in the current study)
Evaluative talk	Weaker relationships than "factually oriented talk"
Description	(Examined in the current study)
Narration	Weaker relationships than "description" and "comparison"
Instruction	Similar relationships to "description"
Comparison	(Examined in the current study)
Monologues	(Examined in the current study)
Dialogues	Weaker relationships than "monologues"
Speaking performance with planning time (i.e., prepared speech)	Stronger relationships than "speaking performance without planning time"
Speaking performance without planning time (i.e., spontaneous speech)	(Examined in the current study)

logical inferences, further studies are necessary to investigate actual relationships between productive vocabulary knowledge and speaking performance.

12.5 Summary of Chapter 12

This chapter examines relationships between size and depth of productive vocabulary knowledge and speaking performance. It is concluded that there is a moderate or strong tendency in some tasks for those (Japanese learners of English at the novice level) with larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types and better speaking performance related to vocabulary when they produce monologic description and comparison without pre-task planning time. It is also shown that all four aspects of speaking performance (i.e., fluency, accuracy, syntactic complexity, and lexical complexity) are related to productive vocabulary knowledge, but only one element of each aspect is found to be related. It is also demonstrated that size of productive vocabulary knowledge is related more to accuracy in a task, and that depth (especially knowledge of antonyms) of productive vocabulary knowledge is related more to one element of fluency (speaking speed), and that both size and depth of productive vocabulary knowledge are related to one element of syntactic complexity and one element of lexical complexity in monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level.

Part Ⅳ Conclusion

Chapter 13 Conclusion

13.1 Overview of Findings

This dissertation aims to examine relationships between size and depth of productive vocabulary knowledge and speaking performance, with a specific focus on Japanese learners of English at the novice level (i.e., third-year junior and first-year to third-year senior high school students in Japan). In this study, vocabulary size refers to the number of words whose written forms a person can write in response to a primary meaning (Adapted from Nakanishi & Shimamoto, 2003), whereas depth is defined as the degree to which a person can write written forms of derivatives, antonyms, and collocates of stimulus words (Adapted from Nakanishi & Shimamoto, 2003). The type of speaking performance mainly investigated in this study was description and comparison, which are monologues without pre-task planning time. In order to achieve this purpose, two studies were conducted: Studies 1C and 2E. Other studies were conducted in order to select a more appropriate scoring method and speaking performance measures and to examine the validity of inferences and uses of test scores used for Studies 1C and 2E. Figure 13.1 shows where each of the previous chapters was located in the overall research.

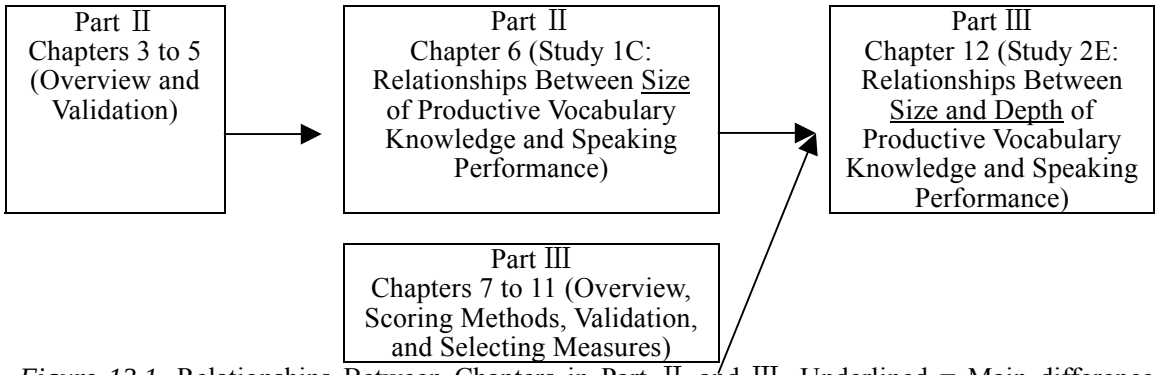


Figure 13.1. Relationships Between Chapters in Part Ⅱ and Ⅲ. Underlined = Main difference between Study 1 (1C) and Study 2 (2E).

In this section, findings related to relationships between productive vocabulary knowledge and speaking performance are reported in section 13.1.1, and other findings related to the tests and measures used are presented in section 13.1.2. It should be noted that the type of speaking performance investigated in this study was only description and comparison, which are monologue without pre-task planning time. While speculations on the cases of other types of speaking performance were made in section 12.4, the summary of findings is restricted to the type of speaking performance examined in this study.

13.1.1 In Relation to the Purpose of the Current Study

Study 1C (Chapter 6) investigated relationships between size of productive vocabulary knowledge and speaking performance. One Research Question was addressed for this purpose.

Research Question: What aspects of speaking performance are related to size of productive vocabulary knowledge to a moderate or strong degree?

As an answer to the Research Question, one element of syntactic complexity, as measured by SC3 (the number of tokens per AS-unit), and one element of lexical complexity, as measured by LC1 (the Guiraud index: the number of types divided by the square root of the number of tokens), are related to size of productive vocabulary knowledge. The interpretation of the results of Study 1C was that those who have greater productive vocabulary knowledge tend to produce more tokens and types and

excel in vocabulary-related aspects of speaking performance (see section 6.3). This was used for setting the hypotheses in Study 2 (2E).

Study 1C was conducted as a pilot study of Study 2E, which was modified based on problems encountered in Study 1 (1C). Study 2E (Chapter 12) investigated relationships between size and depth of productive vocabulary knowledge and speaking performance. Three Hypotheses and one Research Question were addressed for this purpose.

Hypothesis 1: There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types in speaking performance across tasks.

Hypothesis 2: There is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to have better speaking performance related to vocabulary across tasks.

Hypothesis 3: There is a moderate or strong tendency for those who produce a greater number of tokens and types in speaking performance to have better speaking performance related to vocabulary across tasks.

Hypothesis 4: Size of productive vocabulary knowledge is related more to speaking performance than depth of productive vocabulary knowledge is.

Research Question: Are there any individual factors other than productive vocabulary knowledge related to speaking performance to a moderate or strong degree?

The Hypothesis 1 of Study 2E was partially supported because it was upheld only in four tasks out of the five. It is concluded that there is a moderate or strong

tendency in some tasks for Japanese learners of English at the novice level with larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types when they produce monologic description and comparison without pre-task planning time (see section 12.3.2).

The Hypothesis 2 was supported because vocabulary-related aspects that were loaded on by measures derived from all the tasks were all associated with productive vocabulary knowledge. The results show that there is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to have better speaking performance mainly related to vocabulary across tasks. It is also shown that all four aspects of speaking performance (i.e., fluency, accuracy, syntactic complexity, and lexical complexity) are related to productive vocabulary knowledge, but only one element of each aspect is found to be related. To be specific, elements related to productive vocabulary knowledge are (a) one element of fluency (speaking speed), (b) accuracy in a task, (which is not extremely difficult), (c) one element of syntactic complexity, as mainly measured by the AS-unit length (the number of tokens per AS-unit), and (d) one element of lexical complexity, as mainly measured by the Guiraud index (LC1: the number of types divided by the square root of the number of tokens), the number of lexical word types divided by the square root of the number of tokens (LC5), and the number of sophisticated word types divided by the square root of the number of tokens (LC12J101). Productive vocabulary knowledge is not associated with other elements of speaking performance, such as dysfluency for each task (one element of fluency) and proportions of clauses and subordinate clauses (one element of syntactic complexity), some elements of lexical complexity (apart from PLC1 [Lexical complexity derived divided by the square root of the number of tokens]; see section 12.3.3).

With regard to the Hypothesis 3, it was supported. Based on the results of the three hypotheses, it is concluded that there is a moderate or strong tendency in some tasks (or when the content of the talk is already decided and when few formulaic expressions are used) for Japanese learners of English at the novice level who have larger and deeper productive vocabulary knowledge (than others of this level) to produce a greater number of tokens and types and better speaking performance mainly related to vocabulary when they produce monologic description and comparison without pre-task planning time. In addition, it is also found that there is generally a moderate association between productive vocabulary knowledge and accuracy unless tasks are extremely difficult. This relationship with accuracy tends to be weaker than the one with vocabulary-related aspects of speaking performance (see sections 12.3.3 and 12.3.4).

The Hypothesis 4 was partially supported because it was confirmed for accuracy in one task, and not supported for the number of tokens and types and speaking speed, and partially supported for one element of lexical complexity and that of syntactic complexity. It is concluded that size of productive vocabulary knowledge is related more to accuracy in a task, and that at least one aspect of depth (antonyms, or word connections) is related more to one element of fluency (speaking speed) in monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level. It is also shown that both size and antonyms of productive vocabulary knowledge are related to one element of syntactic complexity and one element of lexical complexity (see section 12.3.6).

Concerning the Research Question, it is demonstrated that there are no individual factors other than productive vocabulary knowledge that can be assessed by the questionnaire items and associated with speaking performance (monologic

description and comparison without pre-task planning time) to a moderate or strong degree (see section 12.3.7).

Three other findings are as follows. First, although some proportions of some elements of speaking performance are explained by productive vocabulary knowledge, the maximum proportion explained is approximately 40％, which indicates that contributions of factors other than productive vocabulary knowledge are larger. It should be noted that these results do not mean that productive vocabulary knowledge is not very important in speaking performance. In fact, productive vocabulary knowledge is an essential requirement of speaking performance. However, this study indicates that in actual speaking performance, factors other than productive vocabulary knowledge have a greater effect on speaking performance (see sections 12.3.6, 12.3.7, and 12.4).

Second, comparisons of results between the tasks and between the two studies demonstrate that some minor differences in task characteristics tend to affect relationships between productive vocabulary knowledge and speaking performance. The differences may include characteristics to elicit more formulaic expressions, the length of time during which test takers are required to speak, and differences in test formats (see sections 12.3.4 and 12.4).

Third, lexical density and lexical sophistication with the criterion of considering 101 to 8,000 words in the JACET8000 (JACET Basic Words Revision Committee, 2003) as sophisticated (J101) are negatively related to productive vocabulary knowledge in some tasks. In other words, those with more productive vocabulary knowledge tend to use more grammatical words (i.e., function words) and non-sophisticated words (i.e., basic words) that belong to the most frequent word group of up to 100 words (see section 12.3.6).

13.1.2 In Relation to the Tests and Measures Used

This section illustrates characteristics of tests and measures found in the current study, namely (a) the Productive Vocabulary Knowledge Test (Productive VKT), and (b) the Speaking Performance Measures.

13.1.2.1 In Relation to the Productive Vocabulary Knowledge Test

With regard to the Productive VKT, test formats of assessing size and depth of productive vocabulary knowledge provided positive evidence for validity (see sections 4.3.4 and 8.3). One problem found was that the definition of collocations in Study 2 may have been too broad and the test format may need to be changed in order to assess collocations, which are defined normally as combinations of adjacent words with some degree of exclusivity (see section 12.3.6). Regarding the scoring method of the Productive VKT, it is shown that a scoring method of giving a point for words with correct misspelling is closely correlated with a method of giving a point for words with minor misspelling (see sections 4.3.5 and 8.3), and that results of scoring methods for the written (paper-and-pencil) version are closely associated with results of the oral version (see section 8.3).

13.1.2.2 In Relation to the Speaking Performance Measures

The analyses of the Speaking Performance Measures demonstrate the following six points: Firstly, the Speaking Performance Measures provided positive evidence for validity (see sections 5.3.4 and 11.3.6).

Secondly, the AS-unit length (the number of tokens per AS-unit) is a syntactic complexity measure, rather than a fluency measure, at least to learners at the novice

level (see sections 5.3.2 and 11.3.3).

Thirdly, the degree of generalizability of results across tasks varies, depending on elements of speaking performance. Fluency measures are divided into measures that assess "speaking speed" and "dysfluency." Dysfluency varies across tasks, whereas speaking speed does not, and tends to have generalizability across tasks. Accuracy varies across tasks and tends to lack generalizability across tasks. Syntactic complexity measures are categorized into measures that assess (a) "syntactic complexity that varies across tasks," and (b) "syntactic complexity that does not vary across tasks." Lexical complexity measures have complicated structures and are classified into 12 types, and some have generalizability, whereas others do not (see section 11.3.4).

The fourth finding related to the Speaking Performance Measures is that there are strong correlations between "speaking speed" and "one element of syntactic complexity" (r = .51), and between "speaking speed" and "one element of lexical complexity" (r = .80). The results suggest that those who speak faster tend to produce utterances with more syntactic complexity and lexical complexity (see section 11.3.5).

Fifthly, in measuring speaking time, a method of giving test takers the same amount of time is better than a method of measuring only the time during which test takers speak without a long silence, when relationships between productive vocabulary knowledge and speaking performance are examined (see section 12.4).

Sixthly, for sophisticated words, 14 definitions of "sophisticated" were compared and the criterion of considering 101 to 8,000 words as sophisticated is strongly associated with other criteria with larger standard deviations (see section 10.3). When the criterion of considering 101 to 8,000 words as sophisticated is used, moderate relationships between productive vocabulary knowledge and speaking performance are found in some tasks, whereas when the criterion of considering 1,001

to 8,000 words as sophisticated is used, there are no relationships. Thus, in relation to relationships between productive vocabulary knowledge and speaking performance, using the criterion of considering 101 to 8,000 words as sophisticated may function better (see section 12.3.4).

13.2 Pedagogical and Methodological Implications for Language Teaching and Testing

The findings in the current study suggest five pedagogical implications and four methodological implications.

13.2.1 Pedagogical Implications for Language Teaching and Testing

As for five pedagogical implications, first, the results of this study (see section 13.1.1) provide evidence and rationales for teachers having long-term perspectives of teaching and testing vocabulary and speaking. The results suggest that there are relationships between productive vocabulary knowledge and some elements of speaking performance to a moderate or strong extent, but there are almost zero or small relationships with other elements of speaking performance in the case of monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level. The maximum proportion of speaking performance explained by productive vocabulary knowledge is about 40%. In productive vocabulary knowledge, both size and knowledge of antonyms contribute to better speaking performance. Based on these results, it can be inferred that in producing speaking performance (to be specific, in monologic description and comparison without pre-task planning time of Japanese learners of English at the novice level), declarative knowledge of productive vocabulary is important, but this is

not enough. As seen in section 12.3.7, there are other factors that may be related to speaking performance, and among them, the most important factor seems to be the procedural knowledge of productive vocabulary. Increasing both size and depth (especially word links) of productive vocabulary knowledge is also needed in order to build network of words (Aitchison, 2003, p. 196-199). It should be noted that depth should be systematically taught in a way that does not confuse students (Akase, 2005, p. 149). In addition, in assessing speaking, it is necessary to conduct a speaking test that elicits speaking performance because a paper-and-pencil test is limited in what it can assess. Furthermore, specific information on what aspect of speaking performance is related to productive vocabulary knowledge is also useful for thinking about language learning and teaching in the long view, which may lead to the principled and systematic integration of teaching productive vocabulary knowledge in a speaking syllabus.

Second, the information on the speaking aspects related to productive vocabulary knowledge can be utilized for giving diagnostic information in teaching and testing. For example, although these relationships may not be causal, when presenting activities for increasing productive vocabulary knowledge, teachers can say to students, “These techniques will help enhance your vocabulary-related performance in the long run.” It may be also possible to say to students who try hard to remember declarative or static productive vocabulary knowledge and have good scores on paper-and-pencil tests but have problems when speaking, “There is always a gap between the results of paper-and-pencil tests and speaking performance. Just remembering words does not give you the ability to use them or better speaking. Why don’t you spend more time using what you have remembered so far?” In addition, for those who want to increase speaking speed (one element of fluency), the following

piece of advice can be given: “It may be better to learn more about connections of word meaning rather than to try to learn to increase the number of separate words you know. This way of studying can enhance speaking speed to some degree.” It seems that diagnostic information should be easy to understand for learners, so some degree of simplification in the explanation may be necessary. At the same time, a balance also needs to be struck between what has been found so far in strict terms (e.g., restricting the findings to the case of monologic description and comparison without pre-task planning time) and some degree of generalization.

The third pedagogical implication is that there have not been many tests to assess the various aspects of vocabulary knowledge (Aizawa, 2005; see section 2.2.2). The Productive VKT developed in this study can be used for pedagogical and research purposes, although the Collocation Subsection may need to be revised. The test is significant not only because it went through a validation process, but also because the vocabulary knowledge intended to be measured was made explicit, based on Chapelle (1994), Nation (2001), and Read and Chapelle (2001). Assessing depth may also have beneficial effects of having students recognize the importance of various aspects of productive vocabulary knowledge (Aizawa, 2005). The results of vocabulary assessment of size and depth may give general and diagnostic feedback to students and teachers on information regarding the degree and characteristics of vocabulary acquisition of size and depth, which may help planning in regard to instructions, setting an appropriate goal, and selecting appropriate material.

Fourth, as for the scoring methods used in assessing productive vocabulary knowledge, the present study shows that in most cases for beginners, the method of giving one point only for words with correct spelling is sufficiently valid in comparison with other methods of giving credit for words with minor misspelling.

This information seems to lead to greater efficiency in testing of productive vocabulary knowledge.

Fifth, with regard to speaking assessment, two points can be made.

●First, this research provides information on relationships between speaking performance measures. This information can be used for selecting appropriate measures for research and pedagogic purposes. In particular, the result that the AS-unit length (the number of tokens per AS-unit) tends to mean not fluency but syntactic complexity seems crucial in selecting measures. For research purposes, the selection of appropriate measures can enhance the interpretations of the results and advance relevant research. For pedagogic purposes, the measures can also be used for giving feedback on fluency, accuracy, syntactic complexity, and lexical complexity to students and teachers.

●Second, since results vary according to measures and tasks used (except for some measures, such as speaking speed), it is recommended that multiple measures and tasks should be used, and that results should be interpreted in combination with these measures and tasks. In addition, generalizing of results should be done cautiously. Although speaking assessment is time consuming and labor intensive, it seems preferable to include more tasks to reveal more elements of speaking performance. Moreover, previous studies may need to be reviewed with consideration of the tasks in mind. Furthermore, the recognition is needed that results derived from speaking performance using a small number of measures and tasks may reflect only part of learners' speaking performance or ability.

13.2.2 Methodological Implications for Language Teaching and Testing

There are four methodological implications that can be suggested. First, concerning validation methods, this study covered all the six aspects of Messick's (1996) validity framework. Since there are not many studies that use this framework comprehensively, this study can be one example of how to make a validity argument based on theoretical and empirical evidence.

Second, in relation to validation, the problematic nature of the Collocation Subsection was revealed in examining relationships with speaking performance (see section 12.3.6), but not through the validation process of examining the Productive VKT (see section 9.3). This suggests that using 14 methods of validation was not enough, and including more validation methods may reveal new aspects of test scores examined.

The third methodological implication is that although there are strong correlations between size and depth (especially antonyms), these two aspects contribute to speaking performance. Therefore, even among novice learners, both size and depth of productive vocabulary knowledge need to be assessed and examined in order to investigate overall relationships between productive vocabulary knowledge and speaking performance, even when vocabulary size is small.

Fourth, although Study 2E used five tasks with characteristics clearer than Study 1C, some of the task differences were difficult to interpret. Therefore, in order to provide meaningful interpretations between tasks, it is necessary to use tasks with more controlled characteristics.

13.3 Limitations and Suggestions for Future Research

The results of the present study may be related to the Productive VKT, the

Speaking Test used, and the targeted aspects of productive vocabulary knowledge and of speaking performance. Replication studies are needed for generalization (Kline, 2004; Polio & Gass, 1997). This section includes suggestions from the following five perspectives: (a) overall points, (b) the Productive VKT, (c) the Speaking Performance Measures, (d) validation, and (e) areas that were not investigated in the current study.

13.3.1 Limitations and Suggestions Regarding Overall Points

Overall, a range of participants that more precisely represent the population as a whole should be selected in future research. In addition, the proficiency levels of the participants in Study 2 need to be located, for example on the Common Reference Levels in the Common European Framework (Council of Europe, 2001), as has been done in Study 1 (see Note 5.3). Furthermore, this study analyzed the data using all the participants from various schools as one group. However, it is also necessary to analyze the data using students at one school as one separate group to analyze characteristics of the school in order to provide more concrete feedback to teachers at each school. Moreover, in order to conduct rigid theory building, previous studies and the current study should be reinterpreted with test validity and confidence intervals of effect sizes in mind, and meta-analysis should be done in order to integrate the results of previous studies and this study (e.g., Kline, 2004; Shiba & Haebara, 1990; Tango, 2002).

13.3.2 Limitations and Suggestions Regarding the Productive Vocabulary Knowledge Test

In relation to the Productive VKT, there are two points to be made. First, in order to reduce the effects of the test method (Bachman, 1990), multiple sections using

multiple test formats are necessary not only in the Depth Section but also in the Size Section. Moreover, the Productive VKT needs to be further refined, especially the Collocation Subsection.

Second, the test should be improved by including an increased number of items, both more and less difficult ones, with more content representativeness. Furthermore, when a wide range of participants is the target, a few test versions (e.g., more difficult and easier versions of vocabulary tests) need to be made so that a test can be produced that fits students' ability and that is not too difficult or too easy for participants. Making comparable or parallel test versions and comparing results on the same scale is rather difficult when classical test theory is used, but test equation procedures using item response theory (e.g., Hambleton, Swaminathan, & Rogers, 1991; Ohtomo, 1996) can make this possible.

13.3.3 Limitations and Suggestions Regarding the Speaking Performance Measures

With regard to the Speaking Performance Measures, three points require further investigation. First, in order to further reduce effects of measures and tasks on results, more measures and tasks need to be used. Examples of measures to be included are those that use pauses for fluency, specific measures of accuracy, and Index of lexical diversity (D) for lexical complexity. For complexity measures, there seem to be other elements of complexity than the ones that this study examined, such as complexity of noun phrases (e.g., "The apple on the table was cut." is more complex than "The apple is on the table."). Furthermore, although using speaking performance measures was appropriate in the current research, in which target learners were novice learners, more aspects of speaking performance may be revealed using rating scales as

well.

Second, more aspects of speaking performance need to be targeted so that results can more precisely reflect speaking in real-life situations. In particular, dialogues and performance with pre-task planning time and other types of speaking performance, such as narration and explanation need to be investigated for future research. Moreover, perspectives of speaking performance should not be restricted to fluency, accuracy, syntactic complexity, and lexical complexity. For example, in order to examine the differences between what students try to say and what they can say, the cover rate (Matsubara, 2004, p. 361) can be used, which may reveal more aspects of speaking performance. Analysis of the proportion of formulaic phrases in speaking performance also needs to be carried out.

Third, although this study focused on speaking performance without pre-task planning time and the tasks asked test takers to start speaking right after the instruction ended, some test takers did not immediately follow this instruction and plan before starting speaking. A better method to control pre-task planning needs to be established in future studies.

13.3.4 Limitations and Suggestions Regarding Validation

As for validation, there are three points to be made. First, since "validation is a continuing process" (Messick, 1989, p. 13), more aspects of validity should be examined further using more validation methods. In particular, special attention should be paid to checking the process that test takers utilize in solving the Productive VKT and the Speaking Test. This is in order to examine, for example, whether instructions and/or examples regarding parts of speech in the Productive VKT are understood by students, and whether knowledge of parts of speech affect test performance and test

scores. This is the substantive aspect of validity in Messick (1996). Problems found, if any, should be revised in future studies.

Second, in this study, the validation was done based on hypotheses, some of which can be set up by general knowledge of language, not by theory. According to Messick (1989), there are two types of validation: weak and strong. The validation of the current study is of the weak validation type. When the theory is developed in the future, strong validation, which is based on theory, should be conducted.

Third, after Messick (1996), the concept of validity and the validation framework have been discussed, and some proposals for improvement have been made (e.g., Bachman, 2005; Borsboom et al., 2004; Chapelle et al., 2004; Kane, 2001; Zumbo, 2005b). Thus, better ways for validation should also be explored.

13.3.5 Suggestions Regarding Areas Not Investigated in the Present Study

There are seven suggestions for areas that were not investigated in the current study. First, qualitative studies should be conducted as well as the quantitative studies done in this study. In order to examine how productive vocabulary knowledge is used in speaking performance, one possible way is to conduct interviews to collect information on the speaking process. This analysis along with detailed analyses of responses in the Productive Vocabulary Knowledge Test and the Speaking Test may reveal reasons that some learners have difficulty speaking English and cannot use productive vocabulary knowledge when they have sufficient knowledge in memory.

Second, both cross-sectional and longitudinal studies are necessary in order to examine whether increasing productive vocabulary knowledge will improve speaking performance, and to confirm the causal effects of productive vocabulary knowledge on speaking performance. One example of a longitudinal study is that vocabulary teaching

is done first, and later, changes in speaking performance are assessed and interpreted.

Third, the current study only examined the relationship between productive vocabulary knowledge and speaking performance of Japanese beginner level learners of English. Further investigation should focus on intermediate and advanced learners as well as learners at the starting level (i.e., first-year and second-year junior high school students) to grasp the overall picture. An anticipated result based on Adams (1980) and Higgs and Clifford (1982) is that there is a decrease in the degree to which speaking performance can be explained by vocabulary knowledge.

The fourth suggestion as to points not examined in the present research is that the productive vocabulary knowledge examined in this study was declarative but procedural knowledge of productive vocabulary knowledge also needs to be examined for systematic exploration.

Fifth, an assessment of productive vocabulary knowledge from the proficiency perspective using the JACET8000 (JACET Basic Words Revision Committee, 2003) was conducted in this study. Further analysis is needed to examine relationships with vocabulary from achievement perspectives. Examining vocabulary that students have learned from all the input to which they are exposed may demonstrate a clearer picture of acquisition and the relationships between productive vocabulary knowledge and speaking performance.

Sixth, this study only dealt with knowledge of one-word individual words. In the future, multi-word level vocabulary (e.g., Moon, 1997, Read, 2000, 2005; Schmitt, 2004; Wray, 2002) should also be explored.

Seventh, in order to examine in greater detail the relative effects of test takers' internal factors and contextual factors on speaking test performance, other aspects apart from productive vocabulary knowledge and tasks used in this study should also

be examined (see section 12.3.7 for possible aspects).

13.4 Concluding Remarks

This research focused on relationships between size and depth of productive vocabulary knowledge and speaking performance by using multiple aspects of productive vocabulary knowledge and multiple speaking performance measures that assessed fluency, accuracy, syntactic complexity, and lexical complexity. The analyses from multiple perspectives seem to have the advantage of garnering a more complex but wider picture of the phenomenon that may be more realistic.

The current study mainly demonstrates the following two points. First, there is presence and absence of relationships between productive vocabulary knowledge and speaking performance (i.e., monologic description and comparison without pre-task planning time) of Japanese learners of English at the novice level, depending on the elements of speaking performance targeted. In particular, the essential results seem to be that there is a moderate or strong tendency for those who have larger and deeper productive vocabulary knowledge to produce a greater number of tokens and types in speaking performance and better speaking performance related to vocabulary in some tasks. These findings seem significant because previous studies have not examined the relationships between productive vocabulary knowledge and speaking performance from multiple perspectives, and the relative degrees of relationships when target elements of speaking performance are different have not been known previously. Second, the findings show the importance of both size and depth (especially antonyms) of productive vocabulary knowledge in speaking performance even among novice learners, which indicates the necessity of including depth aspects in teaching and testing. Overall, the findings in this study represent a step toward building an empirical

model of speaking performance, although the progress made is very small.

Some may think that some of the findings of this study seem so obvious that empirical examination was not needed. However, it is believed that the empirical evidence derived using tests and measures with more validity may lead to more solid theories and stronger confidence in learning, teaching, and testing. It is hoped that the present study may help the progress of research into relationships between productive vocabulary knowledge and speaking performance, as well as enhance theory development and improve vocabulary teaching.

Notes

Chapter 1 Introduction

[1.1]In this study, the terms "communicative language ability" and "language ability" were used interchangeably.

[1.2]Tasks refers to "what a test taker is required to do during a test or part of a test" and "more specifically, a type of test item involving complex performance in a test of productive skills" (Davies et al., 1999, p, 196).

[1.3]Ability and knowledge are similar and both involve using (i.e., recognizing and producing) a language, but the term "knowledge" was used in this study only when there are no contexts (i.e., discourse; Read, 2000, p. 11; see section 2.2.1), whereas the term "ability" was used when there are contexts.

[1.4]"A word" was used in the definition of productive vocabulary knowledge based on Nation's (2005) and Schmitt's (2000) descriptions. That is, Nation defined "productive knowledge" as "the kind of knowledge needed to use a word in speaking and writing" (p. 585, underlined mine), whereas Schmitt stated that "if we are able to produce a word of our own accord when speaking and writing, then that is considered *productive knowledge*" (p. 4, underlined mine).

Chapter 2 Literature Review

[2.1]This chapter is a modified version of Koizumi (2003, 2004a, 2004b, 2005a, 2005b, 2005c, 2005d, 2005e, 2005f, 2005g, in press).

[2.2]Declarative knowledge is sometimes called explicit knowledge, while procedural knowledge is called implicit knowledge. According to Ellis (2005a, pp. 148-151), explicit and implicit knowledge are different in seven aspects: (a) awareness (conscious vs. intuitive), (b) type of knowledge (declarative vs. procedural), (c) systematicity and certainty of L2 knowledge (anomalous and inconsistent vs. variable but systematic), (d) accessibility of knowledge (controlled vs. automatic), (e) use of L2 knowledge (access to knowledge during planning difficulty vs. during fluent performance), (f) self-report (verbalizable vs. nonverbalizable), and (g) learnability (any age vs. potentially only within critical period). Since the current study only focused on the second aspect (i.e., type of knowledge), this study did not use the term "explicit/implicit knowledge" but "declarative/procedural knowledge."

[2.3]This speaking activity based on Bachman and Palmer (1996) seems similar to Levelt's (1993) speaking process (see section 2.3.1). In the example, the goal setting in Bachman and Palmer appears to take place in the conceptualizer in Levelt. Assessment 1 seems to be related both to the conceptualization and the formulation, while the assessment 2 and plannings 1 and 2 appear to occur in the formulator with access to the lexicon and other storehouses of knowledge. Assessment 3 appears to be an activity

in which the internal speech is monitored.

[2.4]According to Meara (1996), there are two dimensions of vocabulary: size and organization (p. 47). Although he did not use the term "depth," the current study interpreted organization as one aspect of depth and thus, his statement can be interpreted as "size is a more important dimension than depth," which was used for the hypothesis.

[2.5]According to Schmitt and Zimmerman (2002), three types of information were utilized in order to construct a list of answer keys: (a) four learner dictionaries (e.g., *Cambridge International Dictionary of English*, 1995, as cited in Schmitt & Zimmerman), (b) the frequency in the BNC (British National Corpus), and (c) responses from 36 native-English-speaking university students. When there were consistent disagreements among native speakers, all of their responses were considered acceptable as answers (p. 156).

[2.6] According to Wolter (2001), syntagmatic associates "bear a sequential or collocational relationship to the prompt word and, as such, are usually (but not always) from a different word class than is the prompt word (e.g., *dog* vs. *bite*, or *bark*, p. 43).

[2.7]According to Wolter (2001), paradigmatic associates are "words from the same word class as the prompt word, and, as such, could presumably perform the same grammatical function within a given sentence" (p. 43), such as coordinates (e.g., *dog* and *cat*), superordinates (e.g., *dog* and *animal*), subordinates (e.g., *dog* and *terrier*), and synonyms (e.g., *dog* and *canine*, p. 43).

[2.8]Paradigmatic knowledge and syntagmatic knowledge are the one related to paradigmatic and syntagmatic associates (see Notes 2.2 and 2.3 above).

[2.9]The term "lemmas" is used for two meanings: (a) the method of counting a base form, inflected forms, and reduced forms (e.g., *n't*) as one word (see section 2.2.1) and (b) a part of the mental lexicon consisting of information on meanings and syntax (Levelt, 1993).

[2.10]The term "lexemes" is used for two meanings: (a) a part of the mental lexicon consisting of information on meanings and syntax (Levelt, 1993) and (b) multi-word units, which are also called "lexical units" or "lexical items" (Schmitt, 2000, p. 2).

[2.11]It should be noted that monologues can be elicited using direct and semi-direct methods (see section 2.3.2).

[2.12]While it was possible to cover more types of speaking performance, this study attempted to use the multiple number of tasks for each type rather than increasing more types of speaking performance (see section 2.6.2 the fifth characteristic of this study for the rationale).

[2.13]According to Malvern, Richards, Chipere, and Durán (2004, p. 26), the Guiraud index (i.e., the number of different word types divided by the square root of the number of tokens = type/√token) was proposed by Guiraud (1960, pp. 84-90) and is "essentially the same measure" as Carroll's (1964, p. 54) Corrected Type Token

Ratio (TTR; the number of types divided by the square root of two times the number of tokens = type/√(2*token), as seen in the following formulae (p. 26).

Guiraud index = Guiraud's Root TTR
= type/√token
= (type*√token)/(√token*√token)
= (type*√token)/token
= √token*(type/token)

Carroll's Corrected TTR
= type/√(2*token)
= (type*√token)/(√(2*token)*√token)
= (type*√token)/(√2*√token*√token)
= (type*√token)/(token*√2)
= (√token*type)/(token*√2)
= √token/√2*(type/token)
= √(token/2)*(type/token)

The two formulae show that as the number of tokens increases, the square root of the number of tokens increases, and values of the Guiraud index and Carroll's Corrected TTR increases. LC1 (the Guiraud index; lexical complexity measure) differs from F1 (the number of tokens per minute; fluency measure), whose value increases simply by producing more tokens.

[2.14]The Index of lexical diversity (D) is an improved version of Type token ratio and the Guiraud index (Kormos & Dénes, 2004; Malvern & Richards, 1997, 2002; Malvern et al., 2004). D can be computed using Meara and Miralpeix's (2004) D_Tools.

[2.15]Comrey and Lee (1992) presented criteria of sample sizes for using factor analysis or principal component analysis: 50 as very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1,000 as excellent.

[2.16]For example, Uenishi (2005) used 12 independent variables (IVs) for multiple regression analysis (n = 36 to 70). When the critical ratio is set at α = .05 and the ratio of Type 2 error is specified at β = .20, a sample size of more than "50 + 8*m" (m = the number of IVs) is necessary for testing multiple correlations, whereas a sample size of more than "104 + m" is required for testing individual predictors. Thus, for Uenishi's analysis 146 (i.e., 50 + 8m = 50 + 8*12 = 50 + 96) or 116 (i.e., 104 + m = 104 + 12) test takers are needed.

[2.17]Although Cohen's (1988) criteria have been used substantially across fields, the criteria of small, medium, and strong effects and of their meaningfulness vary according to each field of study (e.g., Kline, 2004, pp. 132-136). These multiple criteria have not been developed sufficiently, so further investigation is required.

Chapter 3 Overview of Study 1

[3.1]This chapter is a modified version of Koizumi (2003, 2004a, 2004b, 2005b, 2005c, 2005d, 2005e, 2005f, 2005g, in press).

[3.2]The increasing and decreasing effects of providing the first letter on item difficulty were supported by four test takers' free responses (from the Productive VKT Questionnaire (3-2); see Appendix ?) toward overall impressions of the Productive VKT.

[Both beneficial and detrimental]

Student A: The first alphabet (i.e., letter) written on the test item helped me write the answer on some occasions, but it perplexed me on other occasions when the word I remembered (in response to the Japanese meaning) had a different first alphabet.

Student C: The first letter specified helped me a lot. There were words I could not have remembered without it, but there were other cases where I could have written an answer if there were no first letters.

[Beneficial]

Student B: There were words I remembered thanks to the first alphabet (i.e., letter), but there were also words I came within an inch of remembering but couldn't.

Student C: Without the first letter, I would have scored only half.

[3.3]Although the Grammar Test was not used as a main instrument of the main analysis (Study 1C; Chapter 6), more items that provide better content coverage and representativeness are needed in future studies.

[3.4]Both filled and unfilled pauses are considered pauses, based on Lennon (1990, p. 405; filled pauses = pauses with fillers; unfilled pauses = pauses without fillers). A filler is a sound that "functions unambiguously as a hesitation marker" (Lennon, 1990, p. 406; e.g., *er, erm, mm, um*、*ah, uh, eh, matte, wakannai*. This study sees the following as fillers all together: Japanese and English fillers, and muttering to him/herself.

Chapter 4 Study 1A: Validation of the Productive Vocabulary Knowledge Test

[4.1]This chapter is a modified version of Koizumi (2003, 2005g).

Chapter 5 Study 1B: Validation of the Speaking Performance Measures

[5.1]This chapter is a modified version of Koizumi (2004a, 2004b, 2005f).

[5.2]An example of proficiency descriptors at the Smattering level is "Can use some basic greetings; can say *thank you, sorry*" and an example at the Waystage Plus level is "Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary" (North, 2000, pp. 274-275). North's proficiency scale was used as the basis of the Common Reference Levels of Common European Framework (Council of Europe, 2001). In the Common Reference Levels, most of the participants in this study belonged to the level of basic users or below.

[5.3]Table A shows the distribution of participants at each speaking level based on Koizumi's (2005a) results.

Table A
Number of Participants at Each Speaking Level

North's Level	School B	School C	School D	School E
Waystage Plus or above	0	0	0	4
Waystage	1	1	1	3
Breakthrough	6	5	3	4
Tourist	11	6	0	2
Smattering	11	14	0	0
Below Smattering	0	2	0	0
Total	29	28	4	13

Note. $n = 74$.

[5.4]All forms of *be, do, have* were considered to be grammatical based on O'Loughlin (personal communication, June 7, 2004).

[5.5]The correlation was very high between the number of sophisticated words with proper nouns and Japanese words and the number of ones without ($r = .93$, $p < .01$).

[5.6]For each task, there were ten combinations: (a) "AS-unit length [ASL] and F1" and "ASL and SC1," (b) "ASL and F1" and "ASL and SC2," (c) "ASL and F3" and "ASL and SC1," (d) "ASL and F3" and "ASL and SC2," (e) "ASL and F4" and "ASL and SC1," (f) "ASL and F4" and "ASL and SC2," (g) "ASL and F5" and "ASL and SC1," (h) "ASL and F5" and "ASL and SC2," (i) "ASL and F6" and "ASL and SC1," (j) "ASL and F6" and "ASL and SC2."

[5.7]10 (for each task)*3 = 30.

[5.8]10 (for each task)*3*3 = 90.

Chapter 6 Study 1C: Relationships Between Size of Productive Vocabulary Knowledge and Speaking Performance

[6.1]This chapter is a modified version of Koizumi (2005b).

Chapter 7 Overview of Study 2

[7.1]This chapter is a modified version of Koizumi (2005c, 2005e, in press).

[7.2]It should be noted that some of the productive vocabulary knowledge tests with which the authors explicitly said they intended to assess "depth" (i.e., Wesche & Paribakht, 1993; Shimamoto, 2005) in fact assess synonyms or word associations, which includes synonyms (see section 2.2.2). In addition, the word-associates test (Read, 1993, 1998), which is widely used as a depth test (e.g., Akase, 2005; Ishizuka,

2000; Noro, 2002; Qian, 1999, 2002), assesses synonyms. These facts seem to suggest that even knowledge of synonyms has been considered as a part of depth in vocabulary studies.

[7.3]As for "varying degrees of exclusivity," Schmitt (2000) provided an example of a word that has strong degree of exclusivity: *blond.* It comes "almost exclusively with the word *hair* and a few other animate nouns like *woman* or *lady*" (p. 77). Table B shows varying levels of exclusivity and their examples. Exclusivity decreases from 1) to 4).

Table B
Levels of Collocational Complexity (Degrees of Exclusivity)

Level	Examples
1) Idiom	*bite the dust* (i.e., die), *shoot the breeze* (i.e., engage in idle talk)
2) Invariable collocation	*break a journey* (*stop**/*suspend**; Mochizuki et al., 2003, pp. 62-63), *from head to foot*
3) Collocation with limited choice at one point	*take/have/be given precedence* [*over* noun phrase] *have/feel/experience a need* [*for* noun phrase]
4) Collocation with limited choice at two points	*as dark/black as night/ink* *get/have/receive a lesson/tuition/instruction* [in noun phrase]

Note. Based on Cowie & Howarth (1995, as cited in Schmitt, 2000, pp. 78-79). * = Not acceptable.

[7.4]Collocations usually includes idioms, compound words, phrasal verbs, fixed phrases, proverbs, lexical phrases, which are called multi-word units, that is, "a string of words with a single meaning" (Schmitt, 2000, pp. 97-101). Idioms refer to "multi-word items which are not the sum of their parts" (e.g., *spill the beans*, Moon, 1997, p. 46). "Proverbs differ from idioms in that they display shared cultural wisdom" (Schmitt, p. 100). However, since this study only dealt with two words next to each other, not all the categories above were involved (e.g., proverbs).

Moon (1997) set three criteria for defining multi-word units (MWU): (a) institutionalization (i.e., "To the extent that language speakers use a MWU in a similar and consistent way, it is institutionalized."), (b) fixedness (i.e., the degree to which combinations and their orders are determined; based on Schmitt, 2000, pp. 97-99), (c) noncompositionality (i.e., "the degree in which their meaning can be derived from a word-by-word analysis"). An example of idioms with strong fixedness is *kick the bucket*, which does not allow *kick the pail** (Schmitt, p. 78), *kick the big bucket** (Schmitt, p. 97) but allows *He kicked the bucket* (Nation, 2001, p. 331). Another example of fixed phrases with strong fixedness is *ladies and gentlemen*, which does not allow *gentlemen and ladies** (Schmitt, p. 99). As for noncompositionality, an example of phrases with relatively transparency is *running on all cylinders* (i.e., operating very well) and an opaque example of idioms is *kick the bucket*.

[7.5]As can be seen in section 9.2.3.2.1, all the acceptable combinations of words were scored as correct and treated equally. In other words, free combinations (e.g.,

strong dog), words next to each other with strong degrees of exclusivity (e.g., *hot dog*), and words next to each other with high frequency (e.g., *happy birthday* in contrast to *fun birthday*) were scored the same, and no weight was given to words next to each other with stronger degrees of exclusivity.

In addition, some may wonder if the Collocation Subsection assesses grammar to some degree rather than vocabulary, which was not specified in the description of the intended construct (see Appendix ?). For example, although the explanation of parts of speech was put in the instruction of the Derivation Subsection (above the Collocation Subsection), knowledge of parts of speech may have been needed to some degree. This possibility needs to be examined in future studies.

All the test takers' responses can be seen in Appendix ?. Most responses can be categorized as free combinations (e.g., (5-1) *perfect* ( ): *ability, action, address, house, human, idea*). Although the analysis of response patterns and factors that affect them is interesting, it was beyond the scope of this research.

Chapter 9 Study 2B: Validation of the Productive Vocabulary Knowledge Test

[9.1]This chapter is a modified version of Koizumi (in press).

[9.2]In scoring collocations, it was possible to use the existent corpus to determine the acceptability of collocations. However, it was not used because there did not seem to be clear criteria of how many times of existence of collocations can be considered acceptable, and because the fact that there are no entries in the corpus does not mean these collocations are unacceptable. It should be noted that the dictionaries used to make the criteria were made based on word frequency but when dictionaries are compiled, experts' judgment as well as frequency information are involved, so using dictionaries was judged acceptable.

[9.3]When there are multiple dependent variables and they are correlated with each other, one choice of method is multivariate analysis of variance (MANOVA). However, this study did not use MANOVA because the use of MANOVA reduces the power when there are positive high correlations (Tabachnick & Fidell, 2001, p. 357). One advantage to use MANOVA over separate (univariate) ANOVAs on each dependent variable is that MANOVA can control of familywise Type 1 error. However, this error rate can be controlled by utilizing separate ANOVAs on each dependent variable and using a Bonferroni correction to each ANOVA. Therefore, multiple one-way ANOVAs and Bonferroni corrections were used in order to control Type 1 error. The significance level (α) was changed for each case using the following formula: .05/(the total number of significance tests). When there was a significant difference in one-way ANOVA, the Bonferroni method was used for testing multiple comparisons.

Chapter 11 Study 2D: Validation of the Speaking Performance Measures

[11.1]This chapter is a modified version of Koizumi (2005f).

[11.2]Although it is interesting to examine what types of task can elicit particular aspects of speaking performance more, it is rather difficult to state the relationships clearly because many studies in SLA and language testing have reported conflicting results (e.g., Iwashita et al., 2001).

[11.3]Although it was possible to perform PCA with all the aspects together and examine the structures, this study did not do so because it assumes a framework which posits that speaking performance is classified into four aspects (i.e., fluency, accuracy, syntactic complexity, and lexical complexity). There are two ways of analysis when there are moderate or strong correlations between different constructs: (a) to analyze correlated constructs together, based on the empirical structures, and (b) to analyze them separately, based on the theoretical or logical framework. For example, sometimes grammatical knowledge and vocabulary knowledge are highly correlated. The first type of analysis is to analyze the two kinds of knowledge together and combine them as language proficiency, whereas the second type is to analyze them individually. The current study used the latter method.

One point to be added is that the first method (i.e., analyzing the four aspects at the same time) was also conducted but the four components did not appear very clearly. Thus, the validity of the four-component framework also needs to be examined in future studies. However, it seems preferable that the structures are examined using a wider range of learners from novice to advanced levels.

Chapter 12 Study 2E: Relationships Between Size and Depth of Productive Vocabulary Knowledge and Speaking Performance

[12.1]This chapter is a modified version of Koizumi (2005c, 2005e).

[12.2]The participants were not divided into levels because the reliability of the size and depth sections was not sufficiently high (less than $\alpha = .70$).

[12.3]The discourse markers (e.g., “in the left picture” and “in the right picture” in Task 2 and “before” and “after” in Task 5) indicate cohesion in the utterances based on the definition below. Cohesion is defined as “the grammatical and/or lexical relationships between the different elements of a text” (e.g., “A: Is Jenny coming to the party? B: Yes, she is.: There is a link between *Jenny* and *she* and also between *is* ... *coming* and *is*.” Richards & Schmidt, 2002, p. 86), whereas coherence is defined as “the relationships which link the meanings of utterances in a discourse or of the sentences in a text.” (e.g., “A: Could you give me a lift home? B: Sorry, I’m visiting my sister.: There is no grammatical or lexical link between A’s question and B’s reply but the exchange has coherence because both A and B know that B’s sister lives in the

opposite direction to A's home." Richards & Schmidt, pp. 85-86). It should be noted that Halliday and Hasan (1976) did not deal with coherence (Hirabayashi, 1994, p. 51) and that cohesion and coherence have some common areas to share (Hirabayashi, p. 51).

[12.4]Another possibility is that some test takers were familiar with the words "before" and "after" because there is a TV program "Daikaizou gekiteki before after [Great dramatic remodeling before after]" that can currently be watched every Sunday in Japan. In this program, the words "before" and "after" can be heard (as Japanized English) several times.

References

Adams, M. L. (1980). Five coocurring factors in speaking proficiency. In J. R. Firth (Ed.), *Measuring spoken language proficiency* (pp. 1-6). Washington, DC: Georgetown University Press.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24,* 425-438.

Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 38-49). Amsterdam, the Netherlands: John Benjamins.

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon* (3rd ed.). Malden, MA: Blackwell.

Aizawa, K. (1998). Developing a vocabulary size test for Japanese EFL learners. *ARELE (Annual Review of English Language Education in Japan), 9,* 75-85.

Aizawa, K. (2005). Assessing the depth of receptive vocabulary knowledge. *Facilitating vocabulary acquisition--JACET (Japan Association of College English Teachers) summer seminar proceedings No. 4:* (pp. 18-28). Tokyo: JACET.

Aizawa, K., Ochiai, N., & Osaki, S. (2003). The effects of teaching on vocabulary knowledge: Receptive vs. productive. *ARELE (Annual Review of English Language Education in Japan), 14,* 151-160.

Akase, M. (2005). The roles of breadth and depth of vocabulary knowledge in EFL reading comprehension: With a focus on English major students. *ARELE (Annual Review of English Language Education in Japan), 16,* 141-150.

Albert, Á., & Kormos, J. (2004). Creativity and narrative task performance: An exploratory study. *Language Learning, 54,* 277-310.

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

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.

Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing, 22,* 301-320.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA/APA/NCME). (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

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, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Ano, K. (2002). Koukousei eigo gakushuusha no hatsuwa ni okeru ryuuchousa to seikakusa no kankei [Relationship between fluency and accuracy in speaking of high school students learning English]. *STEP (Society for Testing English Proficiency) Bulletin, 14,* 39-49.

Arbuckle, J. L. (1999). Amos (Version 4.02) [Computer software]. Chicago: SmallWaters Corporation.

Arbuckle, J. L., & Wothke, W. (1995). *Amos 4.0 user's guide.* Chicago: SmallWaters Corporation.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal, 84,* 84-94.

Arnaud, P. J. L. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 133-145). London: Macmillan Academic and Professional Ltd.

Baba, K. (2004). Vocabulary size and lexical communication strategies in L2 language production. In M. Minami, H. Kobayashi, M. Nakayama, & H. Sirai (Eds.), *Studies in language sciences (3): Papaers from the Third Annual Conference of the Japanese Society for Language Sciences* (pp. 197-212). Tokyo: Kurosio.

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

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2,* 1-34.

Bachman, L. F., & Cohen, A. D. (Eds.). (1998). *Interfaces between second language acquisition and language testing research.* Cambridge University Press.

Bachman, L. F., Davidson, F., Ryan, K., & Choi, I.-C. (1995). *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study.* Cambridge University Press.

Bachman, L. F., & Eignor, D. R. (1997). Recent advances in quantitative test analysis. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of language and education Vol. 7: Language testing and assessment* (pp. 227-242). Dordrecht, the Netherlands: Kluwer Academic.

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

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

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6,* 254-279.

Bieber, S. L. (1988). Multiple regression and its alternatives. *Social Science Journal, 25*(1), 1-19.

Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (Eds.), The handbook of applied linguistics (pp. 367-396). Malden, MA: Blackwell.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*, 14-31.

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20,* 89-110.
Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061-1071.
Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20,* 1-25.
Brown, A. (2004). Interviewer variability in oral proficiency interviews. *Language Testing Update, 35,* 99-100.
Brown, A., & Iwashita, N. (1998). The role of language background in the validation of a computer-adaptive test. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 195-207). Mahwah, NJ: Lawrence Erlbaum Associates.
Brown, A., & Lumley, T. (1997). Interviewer variability in specific-purpose language performance tests. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC 96* (pp. 137-150). University of Jyväskylä and University of Tampere.
Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English.* Cambridge University Press.
Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy.* Cambridge University Press.
Buck, G. (1997). The testing of listening in a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment* (pp. 65-74). Dordrecht, the Netherlands: Kluwer Academic.
Buck, G. (2001). *Assessing listening.* Cambridge University Press.
Bygate, M. (1987). *Speaking*. Oxford University Press.
Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). Oxford, U.K.: Heinemann.
Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23-48). Essex, U.K.: Pearson Education Limited.
Cambridge University Press. (2004). Cambridge Dictionaries Online: The web's favourite learner dictionaries: Cambridge Dictionary of American English. Retrieved September 4, 2005, from http://dictionary.cambridge.org/default.asp?dict=A
Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 2-27). London: Longman.
Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.
Carroll, J. B. (1964). *Language and thought.* Englewood Cliffs, NJ: Prentice-Hall.
Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69).

Oxford University Press.
Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal, 52,* 43-56.
Centre for Applied Language Studies (CALS, University of Wales, Swansea). (n.d.). KR21 Calculator. Retrieved August 12, 2005, from http://www.swan.ac.uk/cals/calsres/encyclopedia/kr21.htm
Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tests and rater groups. *Language Testing, 12,* 16-33.
Chalhoub-Deville, M. (1996). Performance assessment and the components of the oral construct across different tests and rater 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 University Press.
Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20,* 369-383.
Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10,* 157-187.
Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32-70). Cambridge University Press.
Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics, 19,* 254-272.
Chapelle, C. A., Enright, M. K., & Jamieson, J. (2004, March). *Issues in developing a TOEFL validity argument.* Paper presented at the 26th Language Testing Research Colloquium, Temecula, CA.
Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 316-325.
Clapham, C. (1996). *Studies in language testing 4: The development of IELTS: A study of the effect of background knowledge on reading comprehension.* Cambridge University Press.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Lawrence Erlbaum.
Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge University Press.
Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America, 107,* 989-999.
Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21,* 107-145.

Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24,* 197-222.
Davies, A. (1984). Validating three tests of English language proficiency. *Language Testing, 1,* 50-69
Davies, A. (1989). Communicative competence as language use. *Applied Linguistics, 10,* 157-170.
Davies, A. (2004). The native speaker in applied linguistics. In A. Davies & C. Elder (Eds.), The handbook of applied linguistics (pp. 431-450). Malden, MA: Blackwell.
Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Studies in language testing 7: Dictionary of language testing*. Cambridge University Press.
Davies, A., & Elder, C. (2005). Validity and validation in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 795-813). Mahwah, NJ: Lawrence Erlbaum.
de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13,* 1-24.
de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46,* 529-555.
de Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition, 19,* 309-329.
de Jong, J. H. A. L. (2004, March). *Modeling development of spoken language ability.* Paper presented at the 26th Language Testing Research Colloquium, Temecula, CA.
Dewaele, J.-M. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 221-249). Clevedon, England: Multilingual Matters.
Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly, 29,* 55-85.
Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition, 20,* 349-385.
Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). Cambridge University Press.
Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing, 11,* 125-144.
Douglas, D. (2000). *Assessing languages for specific purposes.* Cambridge University Press.
Douglas, D. (2001). Performance consistency in second language acquisition and language testing research: A conceptual gap. *Second Language Research, 17,* 442-456.
Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25,* 220-242.

Egusa, C., & Yokoyama, Y. (2004a). Communication task ni okeru fukuzatsusa no sokutei houhou ni tsuite: Tougoteki fukuzatusa to goiteki hukuzatusa no kousatsu [Measures of complexity in communication tasks: Syntactic complexity and lexical complexity]. *HELES (The Hokkaido English Language Education Society) Journal, 4,* 53-63.

Egusa, C., & Yokoyama, Y. (2004b). The effects of task types on second language speech production among Japanese university students: Fluency, accuracy, complexity, and trade-off effects. *ARELE (Annual Review of English Language Education in Japan), 15,* 129-138.

Ellis, R. (2001). Some thoughts on testing grammar: An SLA perspective. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. et al. (Eds.), *Studies in language testing 11: Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 251-263). Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford University Press.

Ellis, R. (2005a). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27,* 141-172.

Ellis, R. (2005b). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Amsterdam, the Netherlands: John Benjamins.

Ellis, R (Ed.). (2005c). *Planning and task performance in a second language.* Amsterdam, the Netherlands: John Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language.* Oxford University Press.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4,* 139-155.

Eskey, D. (2005). Reading in a second language. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 563-579). Mahwah, NJ: Lawrence Erlbaum.

Foster, P. (1996). Doing the task better: How planning time influences students' performance. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 126-135). Oxford, U.K.: Heinemann.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 75-93). Essex, U.K.: Pearson Education Limited.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18,* 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354-375.

Fraser, C. (1988). *NOHARM: An IBM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory.* Armidale, New South Wales, Australia: University of New England, Center for Behavioral Studies.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H.

Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor, MI: The University of Michigan Press.
Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26,* 275-301.
Fujimori, C. (2002). *The effects of form-focused and meaning-focused pre-task activities on accuracy, fluency and complexity in communicative tasks among Japanese senior high school students.* Unpublished master's thesis, Joetsu University of Education.
Fujimori, C. (2004). Speech production no sokutei houhou: seikakusa, ryuuchousa, fukuzatsusa [Measures of speech production: Accuracy, fluency and complexity]. *KATE (Kanto-koshinetsu Association of Teachers of English) Bulletin, 18,* 41-52.
Fulcher, G. (1997). The testing of speaking in a second language. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of language and education Vol. 7: Language testing and assessment* (pp. 75-85). Dordrecht, the Netherlands: Kluwer Academic.
Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied Linguistics, 20,* 221-236.
Fulcher, G. (2003). *Testing second language speaking.* Essex, U.K.: Pearson Education Limited.
Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguisitic output. *Language Learning, 49,* 549-581.
Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston, MA: Allyn & Bacon.
Gotoh, S. (1991). *Eigo onseigaku kenkyuu: Hatsuon to tsuzuri ji* [English phonology: Pronunciation and spelling]. Osaka, Japan: Kyoiku Tosho.
Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics, 12,* 345-364.
Guerrero, M. D. (2000). The unified validity of the Four Skill Exam: Applying Messick's framework. *Language Testing, 17,* 397-421.
Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique.* Paris: Presses Universitaires de France.
Halliday, M. A. K., & Hasan, R., (1976). *Cohesion in English.* Essex, U.K.: Longman.
Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.
Hasselgren, A. (1998). *Smallwords and valid testing.* Unpublished Ph.D. dissertation, University of Bergen.
Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS Academic Module. In L. Cheng & Y. Watanabe, with A. Curtis (Eds), *Washback in language testing: Research contexts and methods* (pp. 97-111). Mahwah, NJ: Lawrence Erlbaum Associates.
Henning, G. (1987). *A guide to language testing: Development, evaluation, research.*

Boston, MA: Heinle & Heinle.
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21,* 303-317.
Hidai, S. (1999). Chuugakusei no gakusyuusha corpus no goi, bunpou kara mita eigo kentei shiken mondai (5 kyuu, 4 kyuu, 3kyuu) [STEP Test items (5th, 4th, and 3rd grade) from the perspectives of vocabulary and grammar in the learner corpus of junior high school students]. *STEP (Society for Testing English Proficiency) Bulletin, 11,* 95-105.
Higgs, T. V., & Clifford, R. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Lincolnwood, IL: National Textbook.
Hirabayashi, T. (1994). Kaku nouryoku wo nani de hakaruka: Koukousei no jiyuu eisakubun ni okeru cohesion to coherence [What should be used for assessing writing ability: Cohesion and coherence in free composition of senior high school students]. *STEP (Society for Testing English Proficiency) Bulletin, 6,* 49-57.
Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.
Ichikawa, S., Dutcher, D., Boyd, S., & Sawamura, K (Eds.). (1995). *The Kenkyusha dictionary of English collocations* (EPWING CD-ROM ed.). Tokyo: Kenkyusha.
Ikeura, S., Izumi, M., & Itakura, T. (1990). *Saishin eigo onseigaku* [Better phonetics for better communication]. Tokyo: Seibido.
In'nami, Y. (2004). *The effects of listening proficiency and task on the listening test performance of Japanese EFL learners.* Unpublished master's thesis, University of Tsukuba, Japan.
Ishizuka, H. (2000). Goi chishiki no fukasa to speaking nouryoku no soukan: native-like fluency wa goi chishiki niyori takamaruka [Correlations between depth of vocabulary knowledge and speaking ability: Can native-like fluency be enhanced by vocabulary knowledge?--]. *STEP (Society for Testing English Proficiency) Bulletin, 12,* 13-25.
Ishikawa, S. (2004). Daigakusei eigo gakusyuusha no juyou goiryoku to happyou goiryoku no kankei: goi size test oyobi essay corpus bunseki ni motozuku approach [A comparison between receptive vocabulary and productive vocabulary of college English learners]. *Bulletin of Chubu English Language Education Society, 34,* 337-344.
Iwashita, N. (2005). An investigation of lexical profiles in performance on EAP speaking tasks. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 3,* 101-111. English Language Institute, University of Michigan. Retrieved December 13, 2005, from http://www.lsa.umich.edu/eli/spaan/papers2005/spaan_working_papers_v3_iwashita.pdf
Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51,* 401-436.

Izumi, E., Uchimoto, K., & Isahara, H (Eds). (2004). *The NICT JLE Corpus.* Tokyo: ALC.
Japan Association of College English Teachers (JACET) Basic Word Revision Committee (Ed.). (2003). *JACET List of 8000 Basic Words.* Tokyo: Author.
Japan Institute for Educational Measurement. (2005). *Eigo communication nouryoku hantei test* [Computerized Assessment System for English Communication [CASEC]. Retrieved April 10, 2005, from http://casec.evidus.com/index.html
Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21,* 47-77.
Kadota, S. (2003). Mental lexicon towa? [What is mental lexicon?]」. In S. Kadota (Ed.), *Eigo no mental lexicon: Goi no kakutoku, syori, gakusyuu* [The English mental lexicon] (pp. 2-11). Tokyo: Shohakusha.
Kamimoto, T., & Kawauchi, C. (2000). The effects of repetition on EFL oral fluency: A quantitative approach. *Kumamoto Gakuen University Journal of Language and Literature, 13,* 1-24.
Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38,* 319-342.
Kaneko, E. (2004, October). *SST (Standard Speaking Test) and NICT JLE Corpus.* Presented at the Meeting at the University of Tsukuba, Japan.
Kano, Y., & Miura, A. (2002). *Amos, EQS, CALIS niyoru graphical tahenryou kaiseki: Medemiru kyoubunsan kouzou bunseki* [Graphical multivariate analyses using Amos, EQS, and CALIS: Visual approach to covariance structure analysis (Revised ed.)]. Kyoto, Japan: Gendai suugakusha.
Katagiri, K. (2001). *Developing the ten-minute vocabulary tests for quick and approximate estimates of general English ability of Japanese EFL learners.* Unpublished Ph.D. dissertation, Tokyo Gakugei University.
Katagiri, K., Komuro, T., Takayama, Y., Takeda, T. & Takei, A. (1997). Speaking nitsuite nani ga iware nani ga wakatteiruka: Kobetsu kenkyuu no shoukai [What is said and known regarding speaking: Introduction to individual studies]」. In A. Baba (Ed.), *Eigo speaking ron* [Studies of speaking English] (pp. 70-83). Tokyo: Kagensha.
Kawakami, A. (1994). The effect of proficiency in a second language on lexical-conceptual representation. *Japanese Journal of Psychology, 64,* 426-433.
Kawashima, T. (2004). Mensetsu houhou ga hatsuwa ni ataeru eikyou: Role play wo mochiita kobetsu mensetsu houshiki to pair mensetsu houshiki no hikaku [Effects of interview types on utterances: Comparison between one-on-one interview and pair interview using role plays]. *STEP (Society for Testing English Proficiency) Bulletin, 16,* 43-57.
Kawauchi, C. (1998). Cued/uncued oral narrative tasks in foreign language fluency. *JACET (Japan Association of College English Teachers) Bulletin, 29,* 83-96.
Kawauchi, C. (2005). *Pre-task planning in L2 oral performance: Quantitative and qualitative approaches.* Tokyo: Kinseido.
Kawauchi, C., & Kamimoto, T. (2000). Distinctive features of oral production by fluent and nonfluent EFL learners. *Language Laboratory, 37,* 21-36.

Kawauchi, C., & Nagasawa, S. (2000). Developing second language proficiency in a study abroad context. *Bulletin of the Institute of Foreign Language Education Kurume University, 7,* 49-79.

Kenkyusha. (2005). Excite jisho [Excite dictionary]: Kenkyusha online dictionary: Shin eiwa chu jiten [New English-Japanese dictionary] (6th ed.). Retrieved September 4, 2005, from http://www.excite.co.jp/dictionary/

Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography, 10,* 135-155.

Kimura, Y. (1995). An analysis of English communicative competence of high school returnees and non-returnees. *ARELE (Annual Review of English Language Education in Japan), 6,* 123-133.

Kimura, M. (2004). *Describing Japanese EFL learner's performance and the process of its development: Utilizing a large-scale learner corpus and data-driven techniques.* Unpublished Ph.D. dissertation, Tokyo Gakugei University, Japan.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *Modern Language Journal, 86,* 571-586.

Kobayashi, M., & Van Moere, A. (2005, July). *Frequency and quantity of output in group oral tests.* Paper presented at the 27th Language Testing Research Colloquium, Ottawa, Canada.

Koizumi, R. (2003). A productive vocabulary knowledge test for novice Japanese learners of English: Validity and its scoring methods. *JABAET (Japan-Britain Association for English Teaching) Journal, 7,* 23-52.

Koizumi, R. (2004a, August). *Speaking no ryuuchousa, seikakusa, fukuzatsusa no shihyou: Nihonjin tyuugakusei ni okeru datousei kensyou* [Measures of fluency, accuracy, syntactic complexity, and lexical complexity: Validation among Japanese junior high school students]. *Proceedings of the 30th annual meeting of the Japan Society of English Education in Nagano,* 30-33.

Koizumi, R. (2004b, March). *Validation of speaking performance measures: The case of Japanese learners of English with low speaking ability.* Poster session presented at the 26th Language Testing Research Colloquium, Temecula, California, U.S.

Koizumi, R. (2005a). Attainment levels of speaking ability of Japanese junior high school students. *JLTA (Japan Language Testing Association) Journal, 7,* 31-50.

Koizumi, R. (2005b, March). *Happyou goi chishiki ga speaking performance ni ataeru eikyou: Nihonjin shokyuusha wo taishouni* [Effects of productive vocabulary knowledge on speaking performance: The case of Japanese learners at the novice level]. Paper presented at the 6th JACET *(Japan Association of College English Teachers)* English lexicography workshop 2005: English dictionary and vocabulary, Tokyo Denki University, Japan.

Koizumi, R. (2005c, August). *Nihonjin chuu kou sei ni okeru happyou goi chishiki nohirosa, fukasa to speaking no kankei [Relationships between size and depth*

of productive vocabulary knowledge and speaking of Japanese junior and senior high school students. Proceedings of the 31th annual meeting of the Japan Society of English Education in Sapporo, 254-257.
Koizumi, R. (2005d). Predicting speaking ability from vocabulary knowledge. *JLTA (Japan Language Testing Association) Journal, 7,* 1-20.
Koizumi, R. (2005e, July). Relationships between productive vocabulary knowledge and speaking test performance using multiple measures. Poster session presented at the 27th Language Testing Research Colloquium, Ottawa, Canada.
Koizumi, R. (2005f). Speaking performance measures of fluency, accuracy, syntactic complexity, and lexical complexity. *JABAET (Japan-Britain Association for English Teaching) Journal, 9,* 5-33.
Koizumi, R. (2005g). Validating a productive vocabulary knowledge test for novice Japanese learners of English. *Facilitating vocabulary acquisition-- JACET (Japan Association of College English Teachers) summer seminar proceedings No. 4:* (pp. 64-69). Tokyo: JACET.
Koizumi, R. (in press). Nihonjin chuukousei ni okeru happyou goi chishiki no hirosato fukasa no kankei [Relationships between breadth and depth of productive vocabulary knowledge among Japanese junior and senior high school students. *STEP (Society for Testing English Proficiency) Bulletin, 17.*
Koizumi, R., & Kurizaki, I. (2002). Nihonjin chuugakusei no monologue niokeru speaking no tokuchou [Speaking characteristics of monologues given by Japanese junior high school students]. *Bulletin of the Kanto-Koshin-Etsu English Language Education Society, 16,* 17-28.
Koizumi, R., & Yamanouchi, I. (2003). Nihonjin chuugakusei no speaking ability no hattatsu [Development in speaking ability among Japanese junior high school students: Using self-introduction task]. *Bulletin of the Kanto-Koshin-Etsu English Language Education Society, 17,* 33-44.
Konishi, T (Ed.). (1996). *Taishukan's genius English-Japanese dictionary* (2nd ed.). Tokyo: Taishukan.
Konishi, T., & Minamide, K (Eds.). (2002). CD-ROM ban Genius eiwa dai jiten [Taishukan's unabridged Genius English-Japanese dictionary [CD-ROM version]. Tokyo: Shogakukan.
Konishi, T., Yasui, M., & Kunihiro, T (Eds.). (1987). *Shogakukan progressive English-Japanese dictionary* (2nd ed.). Tokyo: Shogakukan.
Kormos, J. (2000). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition, 22,* 145-167.
Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32,* 145-164.
Kosuge, A. (2003). *A longitudinal study of development in spoken performance by Japanese junior high school students.* Unpublished master's thesis, Tokyo Gakugei University, Japan.
Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing, 15,* 295-332.
Kunnan, A. J. (2005). Language assessment from a wider context. In E. Hinkel (Ed.),

Handbook of research in second language teaching and learning (pp. 779-794). Mahwah, NJ: Lawrence Erlbaum.
Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan Academic and Professional Ltd.
Laufer, B. (1995). Beyond 2000: A measure of productive lexicon in a second language. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 265-272). Amsterdam: Benjamins.
Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), Second language vocabulary acquisition (pp. 20-34). Cambridge University Press.
Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*, 255-271.
Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21,* 202-226.
Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54,* 399-436.
Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16,* 307-322.
Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16,* 33-51.
Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning, 48,* 365-391.
Lazaraton, A. (1992). The structural organization of a language interview: A conversation analysis 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 University Press.
Lennon, P. (1990a). The advanced learner at large in the L2 community: Developments in spoken performance. *IRAL (International Review of Applied Linguistics in Language Teaching), 28,* 309-324.
Lennon, P. (1990b). Investigating fluency in EFL: A qualitative approach. *Language Learning, 40,* 387-417.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MA: MIT Press.
Levelt, W. J. M. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: Walter de Gruyter.
Linacre, J. M. (1991). *A user's guide to FACETS: Rasch-Model computer programs.* Chicago: MESA Press.
Linacre, J. M. (2003). Facets: Rasch-measurement computer program (Version 3.45.2)

[Computer software]. Chicago: MESA Press.
Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English.* Malmö, Sweden: CWK Gleerup.
Lumley, T., & Brown, A. (2005). Research methods in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 833-855). Mahwah, NJ: Lawrence Erlbaum.
Luoma, S. (2001). *What does your test measure?: Construct definition in language test development and validation.* Unpublished Ph.D. dissertation, University of Jyväskylä.
Luoma, S. (2004). *Assessing speaking.* Cambridge University Press.
Magnan, S. S. (1988). Grammar and the ACTFL Oral Proficiency Interview: Discussion and data. *Modern Language Journal, 72,* 266-276.
Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language: Papers from the annual meeting of the British Association for Applied Linguistics held at the University of Wales, Swansea, September 1996* (pp. 58-71). Clevedon, England: Multilingual Matters.
Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19,* 85-104.
Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment.* Hampshire, England: Palgrave Macmillan.
Matsubara, M. (2004). Sanshutsu no gap wo hakaru shihyou to shiteno Cover Rate [*Cover Rate* as a measurement of the gap between what was to be produced and what actually was]. *Bulletin of Chubu English Language Education Society, 34,* 361-366.
Matsumi, N. (1993). Retrieval processes of words for speaking in a second language. *Japanese Journal of Educational Psychology, 41,* 424-434.
McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge University Press.
McCarthy, M., & Carter, R. (1997). Written and spoken vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 20-39). Cambridge University Press.
McCarthy, M., & Carter, R. (2001). Ten criteria for a spoken grammar. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 51-75). Mahwah, NJ: Lawrence Erlbaum.
McNamara, T. (1996). *Measuring second language performance*. Essex, U.K.: Addison Wesley Longman Limited.
Meara, P. (1990). A note on passive vocabulary. *Second Language Research, 6,* 150-154.
Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance & competence in second language acquisition* (pp. 35-53). Cambridge University Press.
Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research, 18,*

393-407.
Meara, P. (2005). Lexical Frequency Profiles: A Monte Carlo analysis. *Applied Linguistics, 26,* 32-47.
Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4,* 142-154.
Meara, P., & Fitzpatrick, T. (2000). Lex 30: An improved method of assessing productive vocabulary in an L2. *System, 28,* 19-30.
Meara, P., & Miralpeix, I. (2004). _lognostics D_Tools (Version 1.0) [Computer software]. University of Wales Swansea. Retrieved December 7, 2005, from http://www.swan.ac.uk/cals/calsres/lognostics.htm
Mehnert, U. (1998). The effects of different length of time for planning on second language performance. *Studies in Second Language Acquisition, 20,* 83-108.
Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). Cambridge University Press.
Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational Measurement* (3rd ed., pp. 13-103). New York: National Council on Measurement in Education/American Council on Education.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.
Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-256.
Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement, 24,* 367-378.
Ministry of Education, Science & Culture. (1989). *Chuugakkou shidousho gaikokugo hen* [Instruction guidelines for junior high school]. Tokyo: Kairyudo.
Ministry of Education, Science & Culture. (1999a). *Chuugakkou gakushuu shidou youryo (heisei 10 nen 12 gatsu) kaisetsu—gaikokugo hen*-- [Explanation of the Course of Study for junior high school concerning foreign languages]. Tokyo: Tokyo shoseki.
Ministry of Education, Science & Culture. (1999b). *Koutougakkou gakushuu shidou youryo kaisetsu—gaikokugo hen—eigo hen* [Explanation of the Course of Study for senior high school concerning foreign languages with a special focus on English]. Tokyo: Kairyudo.
Ministry of Education, Culture, Sports, Science & Technology (MEXT). (2003). Regarding the establishment of an action plan to cultivate "Japanese with English abilities." Retrieved March 29, 2005, from http://www.mext.go.jp/english/topics/03072801.htm
Mochizuki, A. (1996). *New action α: Koukou hyoujun eigo* [New action α: Standard English for senior high school students]. Tokyo Shoseki.
Mochizuki, M. (1998). A Vocabulary Size Test for Japanese Learners of English. *IRLT Bulletin, 12,* 27-53.
Mochizuki, M. (2002a). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *ARELE (Annual Review of English Language*

Education in Japan), 13, 121-129.
Mochizuki, M. (2002b). Lex30 no kanousei: Go rensou wa happyou goi wo sokutei shiteiru noka [Possibilities of the Lex30: Does word association measure productive vocabulary?]. *Proceedings of the 28th (2nd as a united organization) annual meeting of the Japan Society of English Education in Kobe,* 461-462.
Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System, 28,* 291-304.
Mochizuki, M., & Aizawa, K. (2001). A validity study of the vocabulary size test of controlled productive ability. *Reitaku University Journal, 73,* 85-102.
Mochizuki, M., Aizawa, K., & Tono, Y. (2003). *Eigo goi no shidou manual* [Manual for teaching vocabulary in English]. Tokyo: Taishukan.
Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40-63). Cambridge University Press.
Murphy, J. (2003). Task-based learning: The interaction between tasks and learners. *ELT Journal, 57,* 352-360.
Nakamura, T. (2004). Goi shuutoku [Vocabulary acquisition]. In I. Koike, M. Terauchi, K. Kinoshita, & M. Narita (Eds.), *Dai ni gengo syuutoku kenkyuu no genzai* [Current state of second language acquisition studies] (pp. 123-142). Tokyo: Taishukan.
Nakanishi, Y., & Shimamoto, T. (2003). Goi chishiki to sono sokutei [Vocabulary and its assessment]. In S. Kadota (Ed.), *Eigo no mental lexicon: Goi no kakutoku, syori, gakusyuu* [The English mental lexicon] (pp. 31-62). Tokyo: Shohakusha.
Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.
Nation, I. S. P. (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 581-585). Mahwah, NJ: Lawrence Erlbaum.
Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22,* 287-293.
Newby, D. (2000). Grammar. In M. Byram (Ed.), *Routledge encyclopedia of language teaching and learning* (pp. 248-250). London: Routledge.
Nikolov, M., & Krashen, S. (1997). Need we sacrifice accuracy for fluency? *System, 25,* 197-201.
Nomura, M. (2004). Speaking nouryoku no koujou to goi: koukogaku teki shuhou wotooshite [Enhancement of speaking ability and vocabulary: Using an archaeological research method]. *Naruto Eigo Kyouiku* [Naruto English Education]*, 18,* 60-71. Naruto University of Education, Department of English.
Nomura, M. (2005). The development of productive vocabulary by Japanese learners of English: An analysis of vocabulary in speech manuscripts. *ARELE (Annual Review of English Language Education in Japan), 16,* 151-160.
Noro, T. (2002). The roles of depth and breadth of vocabulary knowledge in reading

comprehension in EFL. *ARELE (Annual Review of English Language Education in Japan), 13,* 71-80.
Noro, T., & Shimamoto, T. (2003). Eigo no goichishiki to gengo unyou [Vocabulary knowledge and second language production in English]. In S. Kadota (Ed.), *Eigo no mental lexicon: Goi no kakutoku, syori, gakusyuu* [The English mental lexicon] (pp. 141-147). Tokyo: Shohakusha.
North, B. (2000). *The development of a common framework scale of language proficiency.* New York: Peter Lang.
Ohtomo, K. (1996). *Koumoku outou riron nyumon* [Introduction to item response theory]. Tokyo: Taishukan.
Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434-447.
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). *Studies in language testing 13: The equivalence of direct and semi-direct speaking tests.* Cambridge University Press.
O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing, 19,* 169-192.
Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21,* 109-148.
Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24,* 492-518.
O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing, 19,* 277-295.
O'Sullivan, B. (2004). Modelling factors affecting oral language test performance: A large-scale empirical study. In M. Milanovic & C. Weir (Eds.), *Studies in language testing 18: European language testing in a global context: Proceedings of the ALTE Barcelona Conference July 2001* (pp. 129-142). Cambridge University Press.
Ota, H. (2002). *Oral language development of Japanese EFL learners: A longitudinal study of spoken performance by 101 junior high school students.* Unpublished master's thesis, Tokyo Gakugei University, Japan.
Ota, H., Hidai, S., & Kajiro, T. (2004). *Corpus kara hajimeru tango tsukai konasi eikaiwa* [Learning English vocabulary and conversation using the corpus]. Tokyo: Obunsha.
Ota, H., Kanatani, K., Kosuge, A., & Hidai, S. (2003). *Eigo ryoku wa donoyouni nobite yukuka: Chuugakusei no eigo shuutoku katei wo ou* [How English ability is developed: Exploring English acquisition process of junior high school students]. Tokyo: Taishukan.
Oxford University Press. (2004). *Oxford Advanced Learner's Dictionary.* Retrieved November 1, 2004, from http://www.oup.com/elt/global/products/oald/
Pavlou, P. (1997). Do different speech interactions in an oral proficiency test yield different kinds of language? In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S.

Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC 96* (pp. 185-201). University of Jyväskylä and University of Tampere.
Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20,* 26-56.
Pearsons, E., Braham, C. G., Stein, J., & Flexner, (2000). *Random House Roget's college thesaurus.* New York, NY: Random House.
Pennington, M. C. (1996). *Phonology in English language teaching.* Essex, U.K.: Addison Wesley Longman Limited.
Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47,* 101-143.
Polio, C., & Gass, S. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition, 19,* 499-508.
Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56,* 282-307.
Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52,* 513-536.
Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355-371.
Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum Associates.
Read, J. (2000). *Assessing vocabulary.* Cambridge University Press.
Read, J. (2004). Plumbing the depths: How should the construct of vocabulary be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-227). Amsterdam, the Netherlands: John Benjamins.
Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes, 20,* 12-16. Retrieved December 13, 2005, from http://www.cambridgeesol.org/rs_notes/rs_nts20.pdf
Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18,* 1-32.
Reed, D. J., & Cohen, A. D. (2001). Revisiting raters and ratings in oral language assessment. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. et al. (Eds.), *Studies in language testing 11: Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 82-96). Cambridge University Press.
Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). Essex, U.K.: Pearson Education Limited.
Rifkin, B., & Roberts, F. D. (1995). Error gravity: A critical review of research design. *Language Learning, 45,* 511-537.
Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14,* 423-441.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45,* 99-140.
Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22,* 27-57.
Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL (International Review of Applied Linguistics in Language Teaching), 43,* 1-32.
Robinson, P., & Ross, S. (1996). The development of task-based assessment in English for academic purposes program. *Applied Linguistics, 17,* 455-476.
Robinson, P., Ting, S. C., & Urwin, J. J. (1995). Investigating second language task complexity. *RELC Journal, 26,* 62-79.
Rost, M. (2005). L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 503-527). Mahwah, NJ: Lawrence Erlbaum.
Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48,* 281-317.
Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing, 16,* 189-216.
Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge University Press.
Schmitt, N (Ed.). (2004). *Formulaic sequences.* Amsterdam, the Netherlands: John Benjamins.
Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 20,* 17-36.
Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36,* 145-171.
Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et al. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning, 53,* 165-202.
Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26,* 173-199.
Selinker, L., & Han, Z. H. (2001). Fossilisation: Moving the concept into empirical longitudinal study. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. et al. (Eds.), *Studies in language testing 11: Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 276-291). Cambridge University Press.
Shiba, Y., & Haebara, T. (1990). *Koudou kagaku ni okeru toukei kaiseki hou* [Statistical methods in behavioral science]. Tokyo University Press.
Shiba, Y., & Watanabe, H. (1984). *Toukei teki houhou □: Suisoku* [Statistical methods : Inferences (Revised ed.)]. Tokyo: Shinyosha.
Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 405-450). Washington, DC: American

Educational Research Association.
Shimada, K. (2004). Chinmoku wa kin narazu: Speech kunren no ryuuchousa ni taisuru kouka [Silence is not golden—The effect of speech training on fluency]. *Bulletin of Chubu English Language Education Society, 34,* 9-14.
Shimamoto, T. (1988). Dokkai ni okeru goi size to goi houryaku ni tsuite [Vocabulary size and vocabulary strategy in reading comprehension]. *JASEC (Japanese Association for Studies in English Communication) Bulletin, 7,* 71-79.
Shimamoto, T. (2000). An analysis of receptive vocabulary knowledge: Depth versus and breadth. *JABAET (Japan-Britain Association for English Teaching) Journal, 4,* 69-80.
Shimamoto, T. (2005). Exploring lexical network systems of Japanese EFL learners through depth and breadth of word knowledge. *ARELE (Annual Review of English Language Education in Japan), 16,* 121-130.
Shimamoto, T & Kadota, S. (2004). Nihonjin eigo gakusyuusha no goi network wo saguru: goi chishiki no fukasa, hirosa, eigo sougouryoku wo tegakarini [Exploring lexical network systems of Japanese EFL learners: Using vocabulary breadth, depth, and English proficiency. *Proceedings of the 30th annual meeting of the Japan Society of English Education in Nagano,* 684-687.
Shimaoka, T (Ed.). (2002). *Wordpower fully-bilingual dictionary.* Shizuoka: Zoushinkai.
Shimizu, S. (2005). *JACET 8000 Level Marker.* Retrieved March 13, 2005, from http://www01.tcp-ip.or.jp/~shin/J8LevelMarker/j8lm.cgi
Singleton, D. (1997). Learning and processing L2 vocabulary. *Language Teaching, 30,* 213-225.
Shiotsu, T. (2003). *Linguistic knowledge and processing efficiency as predictors of L2 reading ability: A component skills analysis.* Unpublished Ph.D. dissertation, University of Reading.
Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11,* 99-123.
Shohamy, E. (2001). *The power of tests.* Essex, U.K.: Pearson Education Limited.
Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17,* 38-62.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 167-185). Essex, U.K.: Pearson Education Limited.
Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1,* 185-211.
Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49,* 93-120.
Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition*

and second language instruction (pp. 183-205). Cambridge University Press.
Society for Testing English Proficiency, Inc (STEP). (2004). *Eiken* [STEP Test: Society for Testing English Proficiency Test]. Retrieved August 26, 2005, from http://www.eiken.or.jp/
Spalding, R. B. (1986). *Spalding shiki phonics no genri* [The writing road to reading--A modern method of phonics for teaching children to read (2nd ed.)] (Furuhashi, S., & Fukada, M., Trans.). Nagoya: Chubu Nihon Kyoiku Bunkakai. (Original work published 1969)
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.
SPSS. (1999). *SPSS® base 10.0 applications guide.* Chicago: SPSS.
Statistical Package for the Social Sciences (SPSS). (2003). SPSS (Version 12.0.1) [Computer software]. Chicago: SPSS Inc.
Summers, D. et al (Eds.). (1995). *Longman dictionary of contemporary English* (3rd ed.). Essex, U.K.: Longman Group Ltd.
Summers, D. et al (Eds.). (2003). *Longman dictionary of contemporary English* (4th ed.). Essex, U.K.: Pearson Education Ltd.
Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing, 18,* 275-302.
Swan, M. (1995). Practical English usage (2nd ed.). Oxford University Press.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th international student ed.). Needham Heights, MA: Allyn & Bacon.
Takahashi, M. (2000). *Eigo kyouiku gaku gairon (kaitei shinban)* [Overview of English language teaching (Revised ed.)]. Tokyo: Kinseidou.
Takebayashi, S. (1988). *Tsuzuriji to hatsuon no rule: Eigo no phonics* [Phonics for Japanese students (Revised ed.)]. Tokyo: Japan Times.
Takiguchi, H. (2003). *A study of the development of speaking skills within the framework of fluency, accuracy and complexity among Japanese EFL junior high school students.* Unpublished master's thesis, Joetsu University of Education, Japan.
Takiguchi, H. (2004). Nihonjin EFL chuugakusei no speaking nouryoku no hattatu kenkyu [A study of the developments of speaking ability among Japanese junior high school students from the viewpoints of fluency, complexity, and accuracy]. *KATE (Kanto-koshinetsu Association of Teachers of English) Bulletin, 18,* 1-13.
Tango, T. (2002). *Meta-analysis nyuumon* [Introduction to meta-analysis: Statistical methods to integrate evidence]. Tokyo: Asakura shoten.
Tarone, E. (2005). Speaking in a second language. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 485-502). Mahwah, NJ: Lawrence Erlbaum.
Taylor, D. S. (1988). The meaning and use of the term 'competence' in linguistics and applied linguistics. *Applied Linguistics, 9,* 148-168.
Temple, L. (1992). Disfluencies in learner speech. *Australian Review of Applied*

Linguistics, 15 (2), 29-44.
Thomas, D. R. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research, 27,* 335-362.
Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). Variable importance in regression and related analyses. *Social Indicators Research: International Journal for Quality-of-Life Measurement, 45,* 253-273.
Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84-119.
Toyoda, H (Ed.). (2003). *Kyoubunsan kouzou bunseki: Gimon hen* [Questions about Covariance Structure Analysis]. Tokyo: Asakura.
Tsukamoto, S. (2004). KWIC Concordance for Windows (Version 4.7). Retrieved October 25, 2004, from http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/kwic.html
Uenishi, K. (2005). *An empirical study on factors predicting the speaking ability of Japanese EFL learners.* Unpublished Ph.D. dissertation, Hiroshima University, Japan.
Ukrainetz, T. A., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy, 18,* 59-78.
University of Cambridge Local Examinations Syndicate (UCLES). (1999). *Cambridge examinations, certificates & diplomas: Cambridge young learners sample papers.* Cambridge: Author.
University of Cambridge Local Examinations Syndicate (UCLES). (2001). *Cambridge Preliminary English Test 3 with answers: Examination papers from the University of Cambridge Local Examinations Syndicate.* Cambridge University Press.
University of Cambridge ESOL Examinations. (n.d.-a). Common European Framework. Retrieved October 2, 2005, from http://www.cambridgeesol.org/exams/cef.htm
University of Cambridge ESOL Examinations. (n.d.-b). YLE (Cambridge Young Learners English Tests) examination report, 2003, Starters. Retrieved December 12, 2005, from http://www.cambridgeesol.org/support/dloads/yle/yle_examrep03_str.pdf
University of Cambridge ESOL Examinations. (n.d.-c). YLE (Cambridge Young Learners English Tests): Handbook and sample papers: Movers sample papers Speaking. Retrieved September 5, 2005, from http://www.cambridgeesol.org/support/dloads/yle/yle_movers_s.pdf
Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing, 16,* 82-111.
Usuda, Y. (2002). Oral production test no hyouka to hatsuwa ryou no kankei: Hatsuwa ryou wa hyouka syakudo no hitotsu ni nariuruka [The relationship between students' oral test scores and the quantity of speech they produce: Can the amount of speech be used as a criterion?]」 *Journal of the English Literary Society of Hakodate, 41,* 45-60.

Van den Branden, K. (1997). Effects of negotiation on language learners' output. *Language Learning, 47,* 589-636.
van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures–oral performances of students aged 11-12 years. *Language Testing, 11,* 291-319.
Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics, 23*, 361-374.
Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17,* 65-83.
Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata, 1,* 53-68. Okayama, Japan: Notre Dame Seishin University. Retrieved September 9, 2004, from http://www1.harenet.ne.jp/~waring/papers/vocsize.html
Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng & Y. Watanabe, with A. Curtis (Eds), *Washback in language testing: Research contexts and methods* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum Associates.
Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27,* 33-52.
Webster's new world dictionary & thesaurus: Accent Software International (Version 2.0). (1998). Hampshire, U.K.: Macmillan.
Weir, C. J. (2005). *Language testing and validation: An evidenced-based approach.* New York: Palgrave Macmillan.
Wesche, M., & Paribakht, T. S. (1993). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53,* 13-40.
Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14,* 85-106.
Wigglesworth, G. (1998). The effect of planning time on second language test discourse. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 91-110). Mahwah, NJ: Lawrence Erlbaum Associates.
Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.
Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly, 26,* 693-711.
Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity.* Honolulu, HI: University of Hawai'i Press.
Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition,* 23, 41-69.
Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge University Press.
Yahoo Japan. (2005). Yahoo! Japan ® dictionaries: Shogakukan's Progressive eiwa chu jiten [Progressive English-Japanese dictionary] (3rd ed.). Retrieved

September 4, 2005, from http://dic.yahoo.co.jp/
Yamakawa, K. (2004). Task no kurikaeshi ga gakushuusha gengo ni oyobosu eikyou [Effects of task repetition on learner language]. *Journal of English Language and Literature (Yasuda Women's University), 13,* 173-185.
Yamamoto, K. (1999). Kyoubunsan kouzou bunseki to sono ouyou [Covariance structure analysis and its applications]. In K. Yamamoto & T. Onodera (Eds.), *Amos niyoru kyoubunsan kouzou bunseki to kaiseki jirei* [Covariance structure analysis and analysis examples using Amos] (pp. 1-22). Kyoto, Japan: Nakanishiya.
Yamaoka, T. (2004). Ninchi kara mita gengo shuutoku [Language acquisition from the cognitive perspective]. In I. Koike, M. Terauchi, K. Kinoshita, & M. Narita (Eds.), *Dai ni gengo syuutoku kenkyuu no genzai* [Current state of second language acquisition studies] (pp. 23-42). Tokyo: Taishukan.
Yashima, T., & Viswat, L. (1997). Acquisition of "fluency in L2" through an overseas study program. *JACET (Japan Association of College English Teachers) Bulletin, 28,* 193-206.
Yasui, I. (1992). *Onseigaku* [Phonology]. Tokyo: Kaitakusha.
Yoda, M. (2004). A study on the relationship between writing fluency and vocabulary knowledge. *LEO, 33,* 1-16. Graduate School, Tokyo Gakugei University.
Yoshikawa, M. (2005). Structure of Japanese EFL learners' speaking ability: An empirical approach. *ARELE (Annual Review of English Language Education in Japan), 16,* 51-60.
Yoshimura, M., Imai, Y., Hiraiwa, Y., Fuse, K., Edasawa, Y., & Mine, H. (2005). Listening ni okeru goi size to gakushuu houryaku [Vocabulary size and learning strategies in listening comprehension]. *Dai 44 kai (2005 nendo) JACET (Japan Association of College English Teachers) zenkoku daikai youryou* [JACET 44th annual convention proceedings] (pp. 180-181). Tokyo: JACET.
Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24,* 1-27.
Zumbo, B. D. (2005a, July). *Quantitative approaches to test research and development.* Workshop presented at the 27th Language Testing Research Colloquium, Ottawa, Canada.
Zumbo, B. D. (2005b, July). *Reflections on validity at the intersection of psychometrics, scaling, philosophy of inquiry, and language testing.* Keynote address presented at the 27th Language Testing Research Colloquium, Ottawa, Canada.

Appendixes

Appendix 2.1

Summary of ratio measures of speaking performance measures used in the previous studies

Measure	Source
Fluency	Arevart & Nation (1991), Dörnyei (1995),[a] Freed (2000), Freed et al. (2004), Fujimori (2004), Kamimoto & Kawauchi (2000), Kawashima (2004), Kawauchi (1998), Kawauchi & Kamimoto (2000), Kawauchi & Nagasawa (2000), Koizumi & Kurizaki (2002), Koizumi & Yamanouchi (2003), Lennon (1990a, 1990b), Murphy (2003),[a] Niwa (2000, as cited in Robinson, 2005),[a] Riggenbach (1991), Segalowitz & Freed (2004), Takiguchi (2003, 2004), Yamakawa (2004), Yashima & Viswat (1997)
No. of words per minute [F1]	
No. of words per unit [F2]	Ano (2002),[A] Freed (2000),[B] Freed, Segalowitz, & Dewey (2004),[CDE] Kimura (1995),[F] Lennon (1990b),[C] Niwa (2000, as cited in Robinson, 2005),[H] Robinson (1995),[C] Robinson (2001),[F] Robinson, Ting, & Urwin (1995),[G] Segalowitz & Freed (2004),[CDE] Shimada (2004)[C]
No. of clauses per minute [F3]	Takiguchi (2003, 2004)
No. of syllables per minute	Ellis (1990, as cited in Ellis & Barkhuizen, 2005), Kormos (2000), Kormos & Dénes (2004), Mehnert (1998), Ortega (1999),[a] Temple (1992),[a] Towell, Hawkins, & Bazergui (1996), van Gelderen (1994),[a] Yuan & Ellis (2003)
No. of syllables per unit	Kormos & Dénes (2004),[C] Mehnert (1998)[C]
No. of pauses per minute	Iwashita et al. (2001),[b] Kormos & Dénes (2004), Takiguchi (2003), Yashima & Viswat (1997)
No. of pauses per unit	Bygate (2001),[H] Lennon (1990a, 1990b),[H] Takiguchi (2003)[I], Yamakawa (2004)[H]
Number of pauses per word	Freed (2000),[d] Yamakawa (2004),[d] Yashima & Viswat (1997)
Length of pauses divided by the total No. of pauses	Cucchiarini et al. (2000), Kormos & Dénes (2004), Towell et al. (1996)
Length of time spent speaking divided by length of total delivery time (including pause time)	Cucchiarini et al. (2000), Kormos & Dénes (2004), Lennon (1990b),[c] Towell et al. (1996)
No. of dysfluency markers per minute [F4]	Iwashita et al. (2001),[b] Kormos & Dénes (2004), Takiguchi (2003), Yashima & Viswat (1997)
No. of dysfluency markers per word [F5]	Arevart & Nation (1991),[d] Bygate (1996), Douglas (1994), Freed (2000),[d] Lennon (1990b), Yashima & Viswat (1997)
No. of dysfluency markers per unit [F6]	Koizumi & Yamanouchi (2003),[H] Lennon (1990a, 1990b),[H] van Gelderen (1994)[J]
No. of clause containing dysfluency markers per clause	Wigglesworth (1997, 1998)
Accuracy	Egusa & Yokoyama (2004b), Foster (1996, 2001), Foster & Skehan (1996), Fujimori (2004), Iwashita et al. (2001), Matsubara (2004), Mehnert (1998), Murphy (2003), Kormos & Dénes (2004), O'Sullivan (2002), Skehan (2001), Skehan & Foster (1997, 1999), Takiguchi (2003), Yuan & Ellis (2003)
No. of error-free clauses per clause [A1]	
No. of error-free units per unit [A2]	Kawashima (2004),[eH] Kimura (1995),[F] Koizumi & Kurizaki (2002),[eH] Koizumi & Yamanouchi (2003),[eH] Niwa (2000, as cited in Robinson, 2005),[H] Robinson (2001),[F] Takiguchi (2003, 2004)[eI], Yamakawa (2004)[H]
No. of errors per word [A3]	Mehnert (1998),[d] Takiguchi (2003, 2004)
No. of errors per unit [A4]	Bygate (2001),[H] Koizumi & Kurizaki (2002),[H] Takiguchi (2003),[I] Williams (1992)[H]

(Appendix 2.1continues)

(Appendix 2.1 continued)

Measure	Source
Proportion of correct target features	Fujimori (2004, e.g., present progressives), Nikolov & Krashen (1997, e.g., plurals), Wigglesworth (1997, verbal morphology; 1998, plurals), Yuan & Ellis (2003, verbs)
No. of definite articles divided by total No. of definite and indefinite articles	Wigglesworth (1997, 1998)
Syntactic Complexity No. of clauses per unit [SC1]	Albert & Kormos (2004),[I] Egusa & Yokoyama (2004a, 2004b),[F] Foster (1996, 2001),[F] Foster & Skehan (1996),[F] Iwashita et al. (2001),[F] Kawauchi (2005),[H] Kawauchi & Nagasawa (2000),[F] Matsubara (2004),[A] Murphy (2003),[I] O'Sullivan (2000, 2002),[F] Robinson (2001),[F] Skehan (2001),[F] Skehan & Foster (1997, 1999),[F] Takiguchi (2003, 2004),[I] Williams (1992),[H] Yamakawa (2004),[H] Yuan & Ellis (2003)[H]
No. of subordinate clauses per unit [SC2]	Lennon (1990a),[H] Mehnert (1998),[H] Takiguchi (2003),[I] Van den Branden (1997),[H] Wigglesworth (1997, 1998)[K]
No. of S-nodes per unit	Mehnert (1998),[H] Niwa (2000, as cited in Robinson, 2005),[H] Robinson (1995),[H] Robinson et al. (1995)[H]
No. of words per unit [SC3]	Bygate (2001),[H] Egusa & Yokoyama (2004a),[F] Fujimori (2004),[I] Kawauchi (1998, 2005),[H] Kamimoto & Kawauchi (2000),[H] Kawashima (2004),[H] Kawauchi & Kamimoto (2000),[H] Koizumi & Kurizaki (2002),[H] Mehnert (1998),[F] Ortega (1999),[G] Takiguchi (2003, 2004),[I] Van den Branden (1997)[H]
Lexical Complexity Type token ratio (TTR): No. of different word types divided by No. of words	Bygate (1996), Daller et al. (2003), Egusa & Yokoyama (2004a), Fujimori (2004), Kimura (1995), Kosuge (2003), Malvern & Richards (2002), Malvern et al. (2004), Niwa (2000, as cited in Robinson, 2005), Ortega (1999), Ota (2002), Read (2004), Vermeer (2000), Yamakawa (2004)
Mean segmental type-token ratio (MSTTR)	Egusa & Yokoyama (2004a), Malvern & Richards (2002), Malvern et al. (2004), Yuan & Ellis (2003)
Guiraud index: No. of different word types divided by the square root of No. of words [LC1]	Daller et al. (2003), Vermeer (2000)
Index of lexical diversity (D)	Kormos & Dénes (2004), Malvern & Richards (1997, 2002), Malvern et al. (2004)
Lexical density: No. of lexical words per word [LC2]	Gass, Mackey, Alvarez-Torres, & Fernández-García (1999), Kawauchi (2005), O'Loughlin (1995, 2001), Robinson (1995), Vermeer (2000)[f]
Weighted lexical density: (No. of sophisticated lexical words and No. of basic lexical words given half the weight) divided by No. of words [LC3]	Mehnert (1998), O'Loughlin (1995, 2001)
No. of sophisticated word types per word [LC4]	Daller et al. (2003), Gass et al. (1999)

Note. [ ] = Measures used in Studies 1B and 1C; [a]Per second; [b]Divided by total amount of speech; [c]Total pause time divided by total delivery time; [d]Per 100 words; [e]Incorrect unit, not error-free unit; [f]Function words per word. The units used in the relevant study: [A]Sentence; [B]Unit defined by dysfluent pauses or hesitations; [C]Pausally defined unit; [D]Filler-free unit; [E]Dysfluency-marker-free unit; [F]c-unit; [G]Utterance; [H]T-unit; [I]AS-unit; [J]Tone group; [K]Clause. F2 and SC3 are the same measures.

Appendix 3.1

Purpose and design of the Productive and Receptive Vocabulary Knowledge Tests

TEST PURPOSE

Inferences: Trait definition of productive [receptive] vocabulary (vocabulary knowledge independent of contexts of use). *Item level*: Knowledge of a relationship between form and meaning of each from a sample of high-frequency words. *Test level*: The degree to which test takers have size of productive [receptive] vocabulary knowledge.

Uses: *Research*: A measure of the degree of productive [receptive] vocabulary knowledge for various kinds of L2 vocabulary research. *Instructional*: A classroom test intended to assist teachers in designing suitable vocabulary teaching plans for their students. A relatively low-stakes test for the students.

Intended Impacts: The test is hand-scored, which limits its application to a rather small scale [The test can be hand-scored or machine-scored]. It provides L2 researchers with a tool designed to measure size of productive [receptive] vocabulary knowledge.

TEST DESIGN

A discrete (i.e., an independent construct is measured; Read & Chapelle, p. 4), selective (i.e., specific target words are selected "as the focus of the assessment"; p. 5), context-independent (i.e., a response can be made "without referring to any context"; p. 4) vocabulary test with the words presented in isolation. A 40-item test. *Input*: Items are presented with a definition in Japanese and the first word [with four options]. *Expected response*: Japanese test takers write the missing part of the form [select one of four single-word options for each item] that matches each of the target words. *Scoring*: Items are scored dichotomously.

Note. Based on the test specification and its examples in Read & Chapelle (2001, pp. 26-32). Specification for the Receptive VKT is shown in [ ].

Appendix 3.2

Paper-and-pencil test in Study 1

語彙・文法テスト

＜試験時間＞　前半25分・後半20分 (合計45分)

前半と後半の間に約1分間の休憩があります。

＜注意事項＞

1　試験開始の合図があるまでは問題用紙の冊子を開かないでください。最初の問題が終わった後に、

ここから先は、指示があるまで進まないでください。

試験開始から25分たったら、休憩が1分間あります。

と書いてありますので、そこまで行ったら、そこまでのところで、見直しをするなどして待っていてください。

2　問題用紙と解答用紙2枚に名前などを書いてください。解答および必要事項は、すべて解答用紙の所定欄・マーク欄に、必ず **HB・B・2B の黒鉛筆かシャープペンシル**で**濃く**記入してください。

3　解答は直接解答用紙に記入してください(問題用紙にはメモしてもかまいませんが、後で解答用紙に解答を書き写す時間はありません)。

4　この問題用紙は回収します。解答用紙といっしょに先生に提出してください。

＜解答用紙の ID番号の記入法＞

以下のように、四角の中にそれぞれ番号を記入し、マーク欄をぬりつぶしてください。

性別：男は1、女は2　　**学年・組・出席番号**：あなたの番号

例：男3年2組15番の人の場合

	0	1	2	3	4	5	6	7	8	9
性別		O	O	O	O	O	O	O	O	O	O
学年		O	O	O	O	O	O	O	O	O	O
組		O	O	O	O	O	O	O	O	O	O
出席番号		O	O	O	O	O	O	O	O	O	O
		O	O	O	O	O	O	O	O	O	O

↓記入してください

＿＿＿＿＿学校

＿＿年＿＿組＿＿＿番

名前＿＿＿＿＿＿＿＿

(男・女)

[1] 単語を書く問題

日本語にあたる英単語を解答用紙に書いてください。始まりの文字は書いてあります。

わからない所があっても、一部分でもいいので、できるだけ書いてください。

例. 町 (t＿＿＿＿＿)

答えは town　　答えを解答用紙に書いてください。

1. 昼食 (l＿＿＿＿＿) lunch　Answers were not printed on the test
2. 重さ、体重 (w＿＿＿＿＿) weight
3. ～の後ろに (b＿＿＿＿＿) behind (common [in the receptive vocabulary section])
4. ～年 (y＿＿＿＿＿) year
5. 袋、手さげかばん (b＿＿＿＿＿) bag
6. ～のせいにする、責任を負わせる (b＿＿＿＿＿) blame
7. 病気の、気分が悪い (s＿＿＿＿＿) sick (common)
8. 音 (s＿＿＿＿＿) sound
9. 箱 (b＿＿＿＿＿) box
10. 戦い、戦闘 (f＿＿＿＿＿) fight
11. だいだい色 (o＿＿＿＿＿) orange
12. 変化 (c＿＿＿＿＿) change
13. 桃 (p＿＿＿＿＿) peach
14. 長さ (l＿＿＿＿＿) length
15. 野菜 (v＿＿＿＿＿) vegetable
16. しなければならない (m＿＿＿＿＿) must
17. 感じ、気持ち (f＿＿＿＿＿) feeling
18. 失敗する、しくじる、うまくゆかない (f＿＿＿＿＿) fail

19. ひびが入る、割れ目ができる (c_____) crack

20. すてきな、すばらしい (n_____) nice

21. ふれる、さわる (t_____) touch

22. 黄色 (y_____) yellow

23. トウモロコシ (c_____) corn

24. 星 (s_____) star

25. 動く、動かす (m_____) move

26. 安い (c_____) cheap

27. かき回す、かき混ぜる (s_____) stir

28. うさぎ (r_____) rabbit

29. 読む (r_____) read

30. おかしい、おもしろい、こっけいな (f_____) funny

31. 貸す (l_____) lend

32. 白 (w_____) white

33. 人気のある、評判の良い (p_____) popular

34. ニワトリ (c_____) chicken

35. 石 (s_____) stone

36. 要点、点 (p_____) point

37. 植物 (p_____) plant

38. 気がつく、見つける (n_____) notice

39. 耳 (e_____) ear

40. ひと月、～月 (m_____) month

[2] 文法問題

次の 1. から 15. までの ( ) に入れるのに最も適切なものを(a)～(d)の中から一つ選び、そのアルファベットのマーク欄をぬりつぶしてください。

1. ( ) play the guitar late at night, Bob.
(a) Isn't (b) Doesn't (c) Don't* (d) Not

2. *A:* ( ) your brother like music?

B: Yes. He has lots of CDs.
(a) Are (b) Am (c) Do (d) Does*

3. My brother got a letter ( ) in French yesterday.
(a) to write (b) written* (c) wrote (d) writing

4. *A:* ( ) bag is this?
B: It's Cindy's.
(a) Why (b) Who (c) Whose* (d) How

5. We spent the summer in Switzerland ( ) the beautiful scenery.
(a) enjoyed (b) enjoying* (c) will enjoy (d) been enjoying

6. Tomorrow is Tsuyoshi's birthday. He'll ( ) thirteen years old.
(a) is (b) am (c) are (d) be*

7. Steve has a brother ( ) is five years younger than he is.
(a) who* (b) whom (c) whose (d) which

8. I know Mr. White. ( ) is a teacher.
(a) He* (b) She (c) It (d) They

9. Michael, this is the restaurant ( ) your father and I had our first date.
(a) when (b) where* (c) which (d) what

10. Scott went to the park to ( ) for his dog.
(a) look* (b) looking (c) looks (d) looked

11. First, I'd like to introduce ( ). My name is Akiko Sato.
(a) mine (b) my (c) me (d) myself*

12. When I was playing video games in the living room, Mother asked me ( ) I would go shopping with her.
(a) if* (b) that (c) as (d) for

13. That temple was ( ) about two hundred years ago.
(a) built* (b) building (c) to build (d) build

14. Jason ran much ( ) than everyone else and easily won the race.
(a) fast (b) fastest (c) faster* (d) as fast as

15. Thank you for ( ) me to the party.
(a) invite (b) invited (c) inviting* (d) to invite

ここから先は、指示があるまで進まないでください。
試験開始から25分たったら、休憩が1分間あります。

このページより前には戻らないでください
---

[3] 単語を選ぶ問題
日本語にあたる英単語を (a)〜(d) の中から一つ選び、そのアルファベットのマーク欄をぬりつぶしてください。

例. 町　(a) place　(b) scene　(c) square　(d) town*
答えは (d) town
　　a　b　c　d
例　O　O　O　O　解答用紙の (d) にマークしてください。

Note. * = answer.

1. 野球
(a) card　(b) data　(c) snow　(d) baseball*

2. 銃、鉄砲
(a) neck　(b) gun*　(c) seed　(d) trousers

3. 意味
(a) state　(b) fun　(c) generation　(d) meaning*

4. たまねぎ
(a) onion*　(b) pear　(c) rose　(d) tree

5. 赤ちゃん
(a) baby*　(b) engine　(c) room　(d) week

6. 影響
(a) comparison　(b) decision　(c) influence*　(d) feather

7. 世帯
(a) impression　(b) opposition　(c) remark　(d) household*

8. 電話
(a) bath　(b) lamp　(c) phone*　(d) pot

9. 年、年齢
(a) age*　(b) problem　(c) space　(d) trick

10. 押すこと、圧力
(a) attempt　(b) opportunity　(c) pressure*　(d) tap

11. さらに進んで
(a) abroad　(b) apparently　(c) aside　(d) further*

12. 食事
(a) air　(b) meal*　(c) piece　(d) sign

13. 覚え書き、メモ
(a) bus　(b) force　(c) note*　(d) sky

14. 害、傷害、危害
(a) conversation　(b) function　(c) harm*　(d) object

15. 予算
(a) budget* (b) loss (c) provision (d) confidence

16. 顔
(a) face* (b) hair (c) leg (d) shoe

17. 働く
(a) help (b) receive (c) show (d) work*

18. かむ、かじる
(a) approve (b) bite* (c) offer (d) refuse

19. 雇うこと、雇用
(a) employment* (b) prayer (c) relief (d) nail

20. 試験
(a) act (b) exam* (c) tennis (d) trouble

21. 病気の、気分が悪い
(a) every (b) late (c) sick* (d) wonderful

22. (損害などを)こうむる、受ける
(a) argue (b) describe (c) join (d) suffer*

23. 批判する
(a) annoy (b) criticize* (c) suck (d) insist

24. 持っている
(a) do (b) give (c) have* (d) raise

25. 教師、先生
(a) college (b) head (c) lesson (d) teacher*

26. 必要な
(a) asleep (b) necessary* (c) personal (d) violent

27. 区別する
(a) commit (b) distinguish* (c) comment (d) enable

28. 聞く
(a) listen* (b) continue (c) mean (d) understand

29. 落ちる、降る
(a) fall* (b) hope (c) ride (d) write

30. 証拠
(a) association (b) difficulty (c) evidence* (d) respect

31. 歓迎する
(a) fancy (b) discuss (c) hurry (d) welcome*

32. 大きい
(a) hot (b) large* (c) least (d) quite

33. 引く、引っ張る
(a) look (b) pull* (c) tell (d) worry

34. 特別な
(a) medical (b) particular* (c) slight (d) spare

35. 実用的な
(a) practical* (b) tough (c) sensitive (d) silly

36. 偉大な、りっぱな
(a) complete (b) early (c) great* (d) usually

37. 軍隊、集団
(a) entrance (b) solution (c) troop* (d) manner

38. まじめな、真剣な
(a) electric (b) mad (c) perfect (d) serious*

39. ～の後ろに
(a) behind* (b) ever (c) we (d) with

40. 私の
(a) below (b) my* (c) past (d) which

これで終わりです。ご協力ありがとうございました。

Appendix 3.3
Answer sheet for the paper-and-pencil test

語彙・文法テスト解答用紙

		0	1	2	3	4	5	6	7	8	9
性別		O	O	O	O	O	O	O	O	O	O
学年		O	O	O	O	O	O	O	O	O	O
組		O	O	O	O	O	O	O	O	O	O
出席番号		O	O	O	O	O	O	O	O	O	O
		O	O	O	O	O	O	O	O	O	O

_____学校

名前_____

[1] 例 town	1. (l)	2. (w)	3. (b)	4. (y)	5. (b)
6. (b)	7. (s)	8. (s)	9. (b)	10. (f)	11. (o)
12. (c)	13. (p)	14. (l)	15. (v)	16. (m)	17. (f)
18. (f)	19. (c)	20. (n)	21. (t)	22. (y)	23. (c)
24. (s)	25. (m)	26. (c)	27. (s)	28. (r)	29. (r)
30. (f)	31. (l)	32. (w)	33. (p)	34. (c)	35. (s)
36. (p)	37. (p)	38. (n)	39. (e)	40. (m)	

[2]	a	b	c	d
1.	O	O	O	O
2.	O	O	O	O
3.	O	O	O	O
4.	O	O	O	O
5.	O	O	O	O
6.	O	O	O	O
7.	O	O	O	O
8.	O	O	O	O
9.	O	O	O	O
10.	O	O	O	O
11.	O	O	O	O
12.	O	O	O	O
13.	O	O	O	O
14.	O	O	O	O
15.	O	O	O	O

[3]	a	b	c	d
例.	O	O	O	O
1.	O	O	O	O
2.	O	O	O	O
3.	O	O	O	O
4.	O	O	O	O
5.	O	O	O	O
6.	O	O	O	O
7.	O	O	O	O
8.	O	O	O	O
9.	O	O	O	O
10.	O	O	O	O
11.	O	O	O	O
12.	O	O	O	O
13.	O	O	O	O
14.	O	O	O	O

	a	b	c	d
15.	O	O	O	O
16.	O	O	O	O
17.	O	O	O	O
18.	O	O	O	O
19.	O	O	O	O
20.	O	O	O	O
21.	O	O	O	O
22.	O	O	O	O
23.	O	O	O	O
24.	O	O	O	O
25.	O	O	O	O
26.	O	O	O	O
27.	O	O	O	O
28.	O	O	O	O

	a	b	c	d
29.	O	O	O	O
30.	O	O	O	O
31.	O	O	O	O
32.	O	O	O	O
33.	O	O	O	O
34.	O	O	O	O
35.	O	O	O	O
36.	O	O	O	O
37.	O	O	O	O
38.	O	O	O	O
39.	O	O	O	O
40.	O	O	O	O

↑ここで、待っていてください

ご協力ありがとうございました。

Appendix 3.4

Analysis of the Productive and Receptive Vocabulary Knowledge Tests

[1]　単語を書く問題

日本語にあたる英単語を解答用紙に書いてください。始まりの文字は書いてあります。**わからない所があっても、一部分でもいいので、できるだけ書いてください。**

例.　町 (t　　)　　答えは town　　答えを解答用紙に書いてください。

1.　昼食　(l　　) lunch (1, 1)	21.　ふれる、さわる　(t　　) touch (1, 2)
2.　重さ、体重　(w　　) weight (2, 1)	22.　黄色　(y　　) yellow (1, 2)
3.　～の後ろに　(b　　) behind (1, 1; C)	23.　トウモロコシ　(c　　) corn (1, 3)
4.#　～年　(y　　) year (1, 1)	24.#　星　(s　　) star (1, 2)
5. 袋、手さげかばん　(b　　) bag (1, 1)	25.　動く、動かす　(m　　) move (1, 1)
6.　～のせいにする、責任を負わせる　(b　　) blame (2, 2)	26.　安い　(c　　) cheap (2, 1)
7.　病気の、気分が悪い　(s　　) sick (1, 1; C)	27.　かき回す、かき混ぜる　(s　　) stir (2, 3)
8.#　音　(s　　) sound (1, 1)	28.#　うさぎ　(r　　) rabbit (1, 3)
9.　箱　(b　　) box (1, 1)	29.　読む　(r　　) read (1, 1)
10.　戦い、戦闘　(f　　) fight (1, 1)	30.　おかしい、おもしろい、こっけいな　(f　　) funny (2, 1)
11.　だいだい色　(o　　) orange (1, 3)	31.　貸す　(l　　) lend (1, 3)
12.#　変化　(c　　) change (1, 1)	32.#　白　(w　　) white (1, 3)
13.×　桃　(p　　) peach (1, ×)	33.　人気のある、評判の良い　(p　　) popular (1, 2)
14.　長さ　(l　　) length (1, 2)	34.　ニワトリ　(c　　) chicken (1, 2)
15.　野菜　(v　　) vegetable (1, 3)	35.　石　(s　　) stone (2, 2)
16.#　しなければならない　(m　　) must (1, 1)	36.#　要点、点　(p　　) point (1, 1)
17.　感じ、気持ち　(f　　) feeling (1, 1)	37.　植物　(p　　) plant (1, 2)
18.　失敗する、しくじる、うまくゆかない　(f　　) fail (2, 2)	38.　気がつく、見つける　(n　　) notice (1, 1)
19.　ひびが入る、割れ目ができる　(c　　) crack (2, 3)	39.　耳　(e　　) ear (1, 2)
20.#　すてきな、すばらしい　(n　　) nice (1, 1)	40.#　ひと月、～月　(m　　) month (1, 1)

Note. e.g., (1, 3) = 1,000 word frequency of written words; 3,000 frequency words of spoken words in the *Longman Dictionary of Contemporary English* (3rd ed.; Summers et al., 1995). × = excluded analysis; # = From the first version of the Mochizuki Test; C = common items with the receptive vocabulary section. Answers were not printed on the test. The original version had more space. All words were content words.

[3]　単語を選ぶ問題
日本語にあたる英単語を (a)〜(d) の中から一つ選び、そのアルファベットのマーク欄をぬりつぶしてください。
例. 町　(a) place　(b) scene　(c) square　(d) town*　答えは (d) town
　　a　b　c　d
例　O　O　O　●　　解答用紙の (d) にマークしてください。

1. 野球　(a) card (1, 1)　(b) data (1, 1)　(c) snow (1, 3)　(d) baseball* (1, 3)
2. 銃、鉄砲　(a) neck (2, 2)　(b) gun* (1, 2)　(c) seed (2, 3)　(d) trousers (2, 2)
3. 意味 (a) state (1, 3)　(b) fun (3, 2)　(c) generation (3, 2)　(d) meaning* (3, 2)
4.# たまねぎ　(a) onion* (1, 3)　(b) pear (1, 3)　(c) rose (1, 3)　(d) tree (1, 1)
5. 赤ちゃん　(a) baby* (1, 1)　(b) engine (1, 1)　(c) room (1, 1)　(d) week (1, 1)
6. 影響　(a) comparison (2, 3)　(b) decision (2, 1)　(c) influence* (2, 2)
　(d) feather (2, 3)
7. 世帯　(a) impression (3, 3)　(b) opposition (3, 2)　(c) remark (3, 3)
　(d) household* (3, 2)
8.# 電話　(a) bath (1, 2)　(b) lamp (1, 3)　(c) phone* (1, 1)
　(d) pot (1, 2)
9. 年、年齢　(a) age* (1, 1)　(b) problem (1, 1)　(c) space (1, 1)　(d) trick (1, 3)
10. 押すこと、圧力　(a) attempt (2, 2)　(b) opportunity (2, 1)　(c) pressure* (2, 1)
　(d) tap (2, 3)
11. さらに進んで　(a) abroad (3, 2)　(b) apparently (3, 1)　(c) aside (3, 3)
　(d) further* (3, 1)
12.# 食事　(a) air (1, 1)　(b) meal* (1, 2)　(c) piece (1, 1)　(d) sign (1, 3)
13. 覚え書き、メモ　(a) bus (1, 1)　(b) force (1, 3)　(c) note* (1, 1)　(d) sky (1, 2)
14. 害、傷害、危害　(a) conversation (2, 1)　(b) function (2, 3)　(c) harm* (2, 3)
　(d) object (2, 3)
15. 予算　(a) budget* (3, 1)　(b) loss (3, 2)　(c) provision (3, 3)
　(d) confidence (3, 2)
16.# 顔　(a) face* (1, 1)　(b) hair (1, 1)　(c) leg (1, 1)　(d) shoe (1, 2)
17. 働く　(a) help (1, 1)　(b) receive (1, 1)　(c) show (1, 1)　(d) work* (1, 1)
18. かむ、かじる　(a) approve (2, 2)　(b) bite* (2, 2)　(c) offer (2, 1)　(d) refuse (2, 2)
19. 雇うこと、雇用　(a) employment* (3, 1)　(b) prayer (3, 3)　(c) relief (3, 2)　(d) nail (3, 3)
20.# 試験　(a) act (1, 1)　(b) exam* (1, 1)　(c) tennis (1, 3)　(d) trouble (1, 1)
21. 病気の、気分が悪い　(a) every (1, 1)　(b) late (1, 1)　(c) sick* (1, 1)
　(d) wonderful (1, 1)
22. (損害などを) こうむる、受ける　(a) argue (2, 1) (b) describe (2, 1) (c) join (2, 1)　(d) suffer* (2, 1)
23. 批判する　(a) annoy (3, 3)　(b) criticize* (3, 3) (c) suck (3, 3)　(d) insist (3, 3)
24.# 持っている　(a) do (1, 1)　(b) give (1, 1)　(c) have* (1, 1)　(d) raise (1, 2)
25. 教師、先生　(a) college (1, 1)　(b) head (1, 1)　(c) lesson (1, 2)　(d) teacher* (1, 1)
26. 必要な　(a) asleep (2, 2)　(b) necessary* (2, 1)　(c) personal (2, 1)
　(d) violent (2, 3)
27. 区別する　(a) commit (3, 2)　(b) distinguish* (3, 3)　(c) comment (3, 1)
　(d) enable (3, 3)
28.# 聞く　(a) listen* (1, 1)　(b) continue (1, 1)　(c) mean (1, 1)　(d) understand (1, 1)
29. 落ちる、降る　(a) fall* (1, 1)　(b) hope (1, 1)　(c) ride (1, 3)　(d) write (1, 1)
30. 証拠　(a) association (2, 3)　(b) difficulty (2, 1)　(c) evidence* (2, 3)
　(d) respect (2, 1)
31. 歓迎する　(a) fancy (3, 3)　(b) discuss (2, 3)　(c) hurry (2, 3)　(d) welcome* (1, 3)
32.# 大きい　(a) hot (1, 1)　(b) large* (1, 1)　(c) least (1, 1)　(d) quite[1] (1, 1)
33. 引く、引っ張る　(a) look (1, 1)　(b) pull* (1, 1)　(c) tell (1, 1)　(d) worry (1, 1)

34. 特別な (a) medical (2, 2) (b) particular* (2, 1) (c) slight (2, 2) (d) spare (2, 3)
35. 実用的な (a) practical* (3, 3) (b) tough (3, 2) (c) sensitive (3, 3) (d) silly (3, 3)
36.# 偉大な、りっぱな (a) complete (1, 2) (b) early (1, 1) (c) great* (1, 1) (d) usually (1, 1)
37. 軍隊、集団 (a) entrance (3, 3) (b) solution (3, 2) (c) troop* (3, 3) (d) manner (2, 3)
38. まじめな、真剣な (a) electric (2, 2) (b) mad (2, 2) (c) perfect (2, 2) (d) serious* (2, 1)
39. ～の後ろに (a) behind* (1, 1) (b) ever (1, 1) (c) we (1, 1) (d) with (1, 1)
40. 私の (a) below (1, 2) (b) my* (1, 1) (c) past (1, 1) (d) which (1, 1)

....

Note. e.g., (1, 3) = 1,000 word frequency of written words; 3,000 frequency words of spoken words; * = answer; # = From the first version of the Mochizuki Test. [1]*quite* (32d) was not originally used in the Mochizuki Test but it was in the original 1,000-word list, so Item 32 was included in the analysis. All words but two (Items 39 *with* & 40 *which*) were content words in the Receptive VKT.

Appendix 3.5
Item description of the Grammar Test

No.	Description in the Course of Study[a]	School	Structure (Answer underlined)	Grade/Year
1	Negative imperative sentences.	JHS	Don't play the guitar …	5/1998
2	Interrogative sentences which begin with auxiliary verbs like *does*.	JHS	Does your brother like music?	4/1998
3	The adjectival use of past participles.	JHS	… a letter written in French …	3/1988
4	Interrogative sentences which begin with interrogatives *whose*.	JHS	Whose bag is this?	5/1997
5	'Subject + Verb + Object + Complement' pattern, in which the complement is a present participle.	SHS	… spent the summer … enjoying …	pre-2/1998
6	Future forms using auxiliary verbs.	JHS	He'll be thirteen years old.	4/1996
7	The basic restrictive use of relative pronouns *who* used as the subject.	JHS	… a brother who is five …	3/1996
8	Personal pronouns.	JHS	I know Mr. White. He is a teacher.	5/1996
9	Relative adverbs.	SHS	… restaurant where your father …	pre-2/1997
10	Basic use of to-Infinitives.	JHS	… went to the park to look for …	4/1996
11	Personal pronouns.	JHS	… I'd like to introduce myself.	3/1998
12	'Subject + Verb + Indirect Object + Direct Object' pattern, in which the direct object is a clause beginning with *if*.	SHS	… Mother asked me if I would go …	pre-2/1996
13	The past forms of passive voice.	JHS	That temple was built …	3/1991
14	The comparative forms of adverbs.	JHS	Jason ran much faster than …	4/1997
15	Basic use of gerunds.	JHS	Thank you for inviting me to the party.	3/1990

Note. [a]Translations were from Takahashi (2000). School (JHS/SHS) = the Course of Study for junior/senior high school. Grade/Year: e.g., 5/1990 means that the item was derived from the 5th grade test administered in 1990.

Appendix 3.6
Instruction Card of Task 1

----

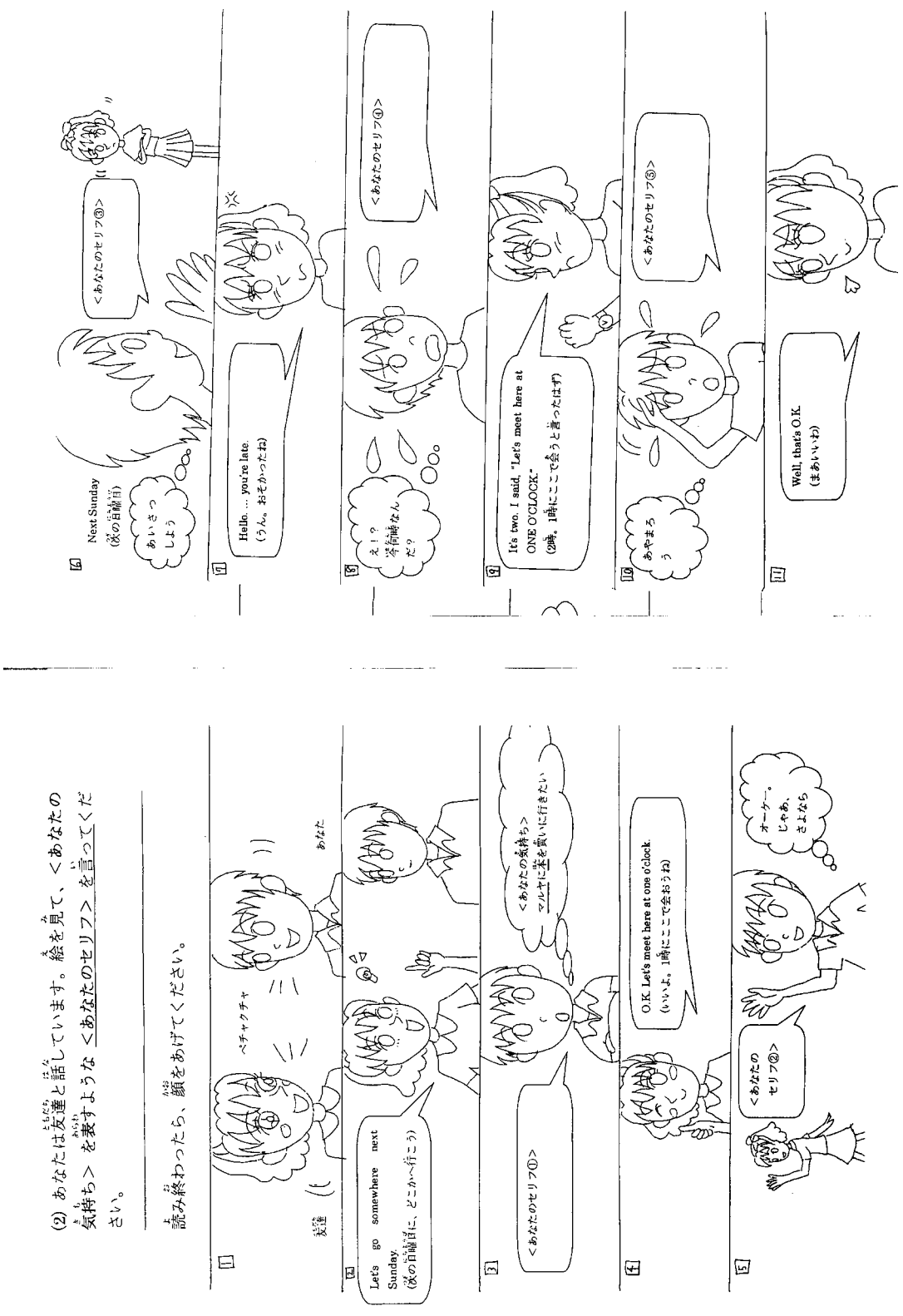
(1) 1分半ぐらいで、あなたの自己紹介をしてください。

☆自己紹介で話すことの例：名前・学年・学校・好きなこと・家族・友達など
(いろいろなことを話してください)

——————————————————————————

読み終わったら、顔をあげてください。

Appendix 3.7
Picture of Task 2 in the Speaking Test



Appendix 3.8
Instruction Card of Task 3

---

(3) あなたは学校新聞を書く係です。先週転校してきた男の子をインタビューし、記事を書こうとしています。メモを見ながら、先生がその男の子だと思って、質問してください。
＜メモ：質問したいこと＞
①この学校は好きか
②今どこに住んでいるか
③前はどこに住んでいたのか
④部活は何に入るつもりか
◎その他 (他にも何か思いついたら聞いてみよう)

読み終わったら、顔をあげてください。

Appendix 3.9
Instruction Card of Task 4

---

(4) あなたの好きな歌手について、1 分半ぐらいで、話してください。考える時間は 1 分間あります。
☆話すことの例 :①好きな理由、②人気はあるか

---

1 分たったら、合図があります。

---

(4) あなたの好きなテレビ番組について、1 分半ぐらいで、話してください。考える時間は 1 分間あります。
☆話すことの例 :①好きな理由、②人気はあるか

---

1 分たったら、合図があります。

---

(4) あなたの好きな動物について、1 分半ぐらいで、話してください。考える時間は 1 分間あります。
☆話すことの例 :①好きな理由、②人気はあるか

---

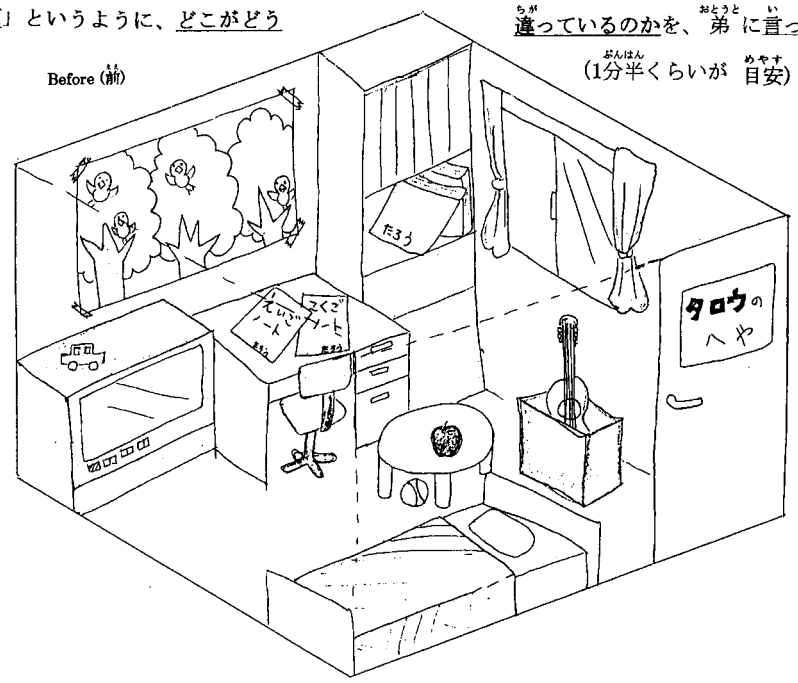
1 分たったら、合図があります。

Appendix 3.10
Picture of Task 5

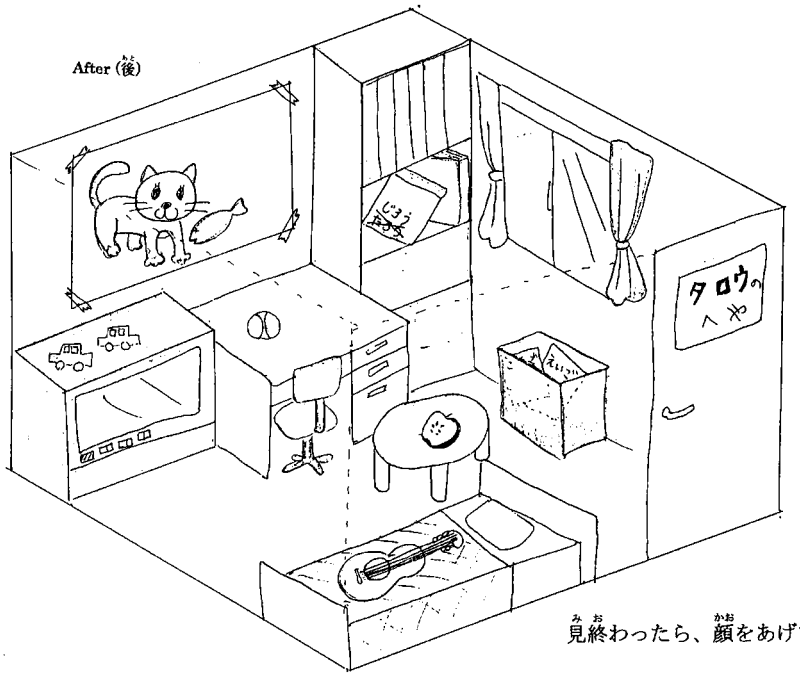
(5) あなたの弟はとてもいたずら好きです。あなたが学校に行っている間に、あなたの部屋をいじってしまいました。弟をしかると、「どこも変わっていない」と言います。「前はどうだったが、今はこうだ」というように、どこがどう違っているのかを、弟に言ってください。

(1分半くらいが目安)

Before (前)



After (後)



見終わったら、顔をあげてください。

Appendix 3.11
Speaking Test Interviewers' Manual

○Principles of being a good interviewer

・Be a sympathetic listener.
・Don't interrupt a student's speech.
・If you forget to time it, allow for more time rather than finishing early.

○Rule for use of English and Japanese

First, speak only English.
→When you are unsure about whether a student understood you or not, and there is a five-second period of silence, say the same thing again with the aid of Japanese translation.

Activity	Conversations and behavior (15 minutes per student)
Warm-up (1 minute)	○When you are ready, open the door and say "Please come in." ＊Students are told to arrive five minutes before their allotted test time. ○A student enters the room. Interviewer (I): Hello. Expected responses from the student (S): (Hello.) ＊Parenthesized utterances are not always required. I: Please sit down. S: (Thank you.) I: Please. ○Ask the student to put a microphone on his/her clothes as closely as possible to their mouth. During instruction before the test, students are told that their voices will be recorded for accurate assessment of the results. ○Use two tape recorders and check the movement of the tapes. I: My name is Hanako Hanayama. May I have your name, please? S: My name is Taro Tsukuba. ○The student may start talking about him/herself prematurely, but since there is a self-introduction task later, try to stop them naturally and go on to the next task. I: Nice to meet you. S: (Nice to meet you, too.) I: Do you like English? ○Start off with easy questions to make the student feel comfortable. S: Yes./No. I: Oh. Who is your English teacher? S: Mr. Ibaraki./Ms. Ibaraki. I: I see. Do you like him/her/the teacher? S: So-so.
Task 1 Self- introduction (2 minutes)	I: Now let's begin our first test. Please read this. ○Give the student a card written in Japanese. ○The student reads it.

	＊Card to be shown to the student ―――――――――――――――――――――――――――――― <Card 1> Please introduce yourself for a minute and a half. ☆Examples：Name, grade, school, favorites, family, friends (Please talk about many things.) ―――――――――――――――――――――――――――――― When you finished reading, please look up at me. ―――――――――――――――――――――――――――――― ○When the student (a) looks up or (b) keeps on reading for more than 30 seconds, say I: O.K. Please start. ○Use a stopwatch. ＊Utterances to be assessed ―――――――――――――――――――――――――――――― S: (Example) My name is Kayoko Ogawa. I go to Aozora Junior High School. I am a third year student. My hobby is listening to music. There are four people in my family. I have many friends. (Waystage+) ―――――――――――――――――――――――――――――― ○Nod without making any comments. Don't stare at the student. Look at him/her approximately once every five seconds. ○If the student remains silent for five seconds before he/she begins, point to examples on the card. ○When the student (a) pauses for 15 seconds or more, (b) speaks for one and a half minutes and stops just for a moment, or (c) says "That's all," say I: O.K. Thank you. ○Collect the card and give some brief comments. I: So you like music. ○If the student does not speak at all, ask an easy question such as, "Do you like dancing?" and try not to make him/her feel bad.
Task 2: Conversation with a friend (3 minutes)	I: Next. Please read this. ○Give the student a card written in Japanese. ○The student reads it. ―――――――――――――――――――――――――――――― <Card 2> You are talking with your friend. Express what you're talking about as shown in the picture. (Picture comes below) ―――――――――――――――――――――――――――――― When you finished reading, please look straight ahead. ―――――――――――――――――――――――――――――― ○When the student (a) looks straight ahead or (b) continues reading for more than 30 seconds, say

I: O.K. Let's begin. Let's go somewhere next Sunday.

Task 2①

I want to go to Maruya to buy a book. (Waystage)

○When the student says nothing for five seconds, point to the picture and the sentence ① while saying "Number 1."

○While pointing to the sentence ① and there is continuous silence for 15 seconds, say,

I: Do you want to buy a book at Maruya?
S: Yes.

○If the student tries to speak , wait for a while.

○When the student says something incomplete, like "I want to buy a book," wait for five seconds. Continue by saying the next line if there is continued silence.

I: O.K. Let's meet here at one o'clock.
S: O.K.

Task 2②

S: Good bye. (Smattering)

○When the student says nothing for 15 seconds, say the next line without pointing to the picture. (The same is applied to the following in Task 2.)

I: Next Sunday.

○Gesture to encourage the next message.

Task 2③

S: Hello. (Smattering)

○When the student says nothing for 15 seconds, go on to the next line.

I: Hello. … you're late.

Task 2④

S: What time is it now? (T)

○When the student says nothing for 15 seconds, go on to the next line.

I: :It's two. I said, "Let's meet here at ONE O'CLOCK."

	Task 2⑤ ―――――――――――――――――――――――――――――――― S: I am sorry. (Smattering) ―――――――――――――――――――――――――――――――― ○When the student says nothing for 15 seconds, go on to the next line. I: Well, that's O.K. ○When the student mumbles, wait for 30 seconds. Then you can go on to the next line. ○When the student says something different from the message, try to respond and modify the conversation so that it will agree with the original task at hand. ○When the student skips a line, instead of going on to the next line, point to the messages he/she has skipped. When there is a silence, wait for 15 seconds and go on to the next line. I: O.K. Thank you. ○Collect the card
Task 3: Interviewing a new student (3 minutes)	I: Next. Please read this. ○Give the student a card written in Japanese. ○The student reads it. ―――――――――――――――――――――――――――――――― <Card 3> You are a reporter for your school newspaper. You will interview a boy who transferred from another school last week, and then write a report. Look at your notes and ask him questions about himself. The teacher in front of you will play the role of the new student. ＜Notes: Things to be asked＞ (Written in Japanese) ①Do you like this school? ②Where do you live now? ③Where did you live before? ④What club are you going to join? ◎Others (Let's ask other things) ―――――――――――――――――――――――――――――――― When you finished reading, please look straight ahead. ―――――――――――――――――――――――――――――――― ○When the student (a) looks straight ahead or (b) keeps on reading for more than 30 seconds, say I: O.K. Please start. Task 3① ―――――――――――――――――――――――――――――――― Do you like this school? (Breakthrough) ―――――――――――――――――――――――――――――――― ○When the student says nothing for five seconds, point to the picture and the sentence ① while saying "Number 1."

	○When the student says something different from the task, try to respond and modify the conversation so that it will fit with the original task at hand. ○When pointing to the sentence ① and there is continuous silence for 15 seconds, say, ”Number 2.” and point to ② on the card. I: Yes, I like it very much. ○When pointing to the sentence ② and there is continuous silence for 15 seconds, say, ”Number 3.” and point to ③ on the card. Task 3② ―――――――――――――――――――――――――――――――― Where do you live? (Breakthrough) ―――――――――――――――――――――――――――――――― I: I live in Tochigi ○When pointing to the sentence ③ and there is continuous silence for 15 seconds, say, ”Number 4.” and point to ④ on the card. Task 3③ ―――――――――――――――――――――――――――――――― Where did you live before? (Waystage) ―――――――――――――――――――――――――――――――― I: I was in Ibaraki. Task 3④ ―――――――――――――――――――――――――――――――― What club are you going to join? (Waystage) ―――――――――――――――――――――――――――――――― I: I want to join the volleyball club. What club are you in? ＊DO NOT forget this question. ○When pointing to the sentence ④ and there is continuous silence for 15 seconds, say, “What club are you in?” Task 3⑤ ―――――――――――――――――――――――――――――――― I am in the soccer club. (Breakthrough) ―――――――――――――――――――――――――――――――― I: Well, do you have any other questions? S: ”Yes.” →Keep talking until the student asks questions. S: “No,” or if there is continuous silence for 15 seconds, I: O.K. Thank you. ○Collect the card ○When the student says something incomplete, he/she may continue to speak, so

	wait for five seconds. ○When the student skips a question, answer it and point to the part that he/she has skipped later. ○When a question is fragmentary and hard to understand, try to respond. If it is caused by bad pronunciation, say “Pardon?” If you still don’t understand, say “Yes.” with a smile. If he/she keeps asking the same question, say “Do you mean …?” ○Respond as if you are a new student. e.g., How old are you? →I am 15.
Task 4: Describing a favorite thing (3 minutes)	I: By the way, do you like music? S: Yes. I: Who is your favorite singer? S: I like Arashi. ○If the student says “No.” say I: Do you like TV? S: Yes. I: What TV program do you like? S: I like Sakura. ○If the student says “No.” say I: Do you like animals? S: Yes. I: What animal do you like? S: I like rabbits. ○If the student says “No.” to all the questions, go on to the next task. ○Ask questions in the order of “music, TV, and animals.” When the student says something related to his/her favorites, say I: Please tell me about … (e.g., Arashi/rabbits/Sakura.) ○If you know his/her favorites already beforehand from the self-introduction, you can just say, I: Tell me about … I: Please tell me about your favorite singer. ○Give the most appropriate card written in Japanese. ○The student reads it. ―――――――――――――――――――――――――――――――― <Card 4> Tell me about your favorite singer, TV programs, or animal for a minute and a half. You have one minute to prepare. ☆Example: reasons, how popular ―――――――――――――――――――――――――――――――― The teacher will give you the cue to start after one minute.

	――――――――――――――――――――――――――――――――― ○Use a stopwatch to give him/her a minute to think. I: (After one minute) O.K. Please start. ○Use a stopwatch again to time the student for a minute and a half. ○Stop the student who begins talking immediately, and point to the card and say, I: You have one minute to think. Task 4 ――――――――――――――――――――――――――――――――― S: (Example) I like Arashi because their songs are nice. They are very popular among junior high school students. (Waystage) ――――――――――――――――――――――――――――――――― ○If the student remains silent for five seconds before he/she starts, point to examples on the card. ○When the student (a) pauses for 15 seconds or more, (b) speaks for one and a half minutes and stops just for a while, or (c) says "That's all," say I: O.K. Thank you. ○Collect the card and give some brief comments or ask an easy question. I: O.K. I like V6. Do you like V6, too? S: Yes./No. I: Oh. Thank you.
Task 5: Comparison between two pictures (2 minutes)	I: Next. Please read this. ○Give the student a card written in Japanese. ○The student reads it. ――――――――――――――――――――――――――――――――― <Card 5> Your brother likes scattering things about. While you are at school, he scattered your things about your room. When you scolded him about it, he said "nothing has changed at all." Tell him how the room has changed by comparing how it was before with how it is now. You have a minute and a half to speak. (Picture comes below) ――――――――――――――――――――――――――――――――― When you finished reading, please look straight ahead. ――――――――――――――――――――――――――――――――― ○When the student (a) looks straight ahead or (b) keeps on reading for more than 30 seconds, say I: O.K. Please start. ○Use a stopwatch. Task 5

	―――――――――――――――――――――――――――――――――――― The ball was under the table. But it is now on the desk. There were English and Japanese notebooks on the desk. But now they are in the box. There was a guitar in the box. But now it is on the bed. There was a picture of three trees and four birds on the wall. But now there is a picture of a fish and a cat on it. There was one (toy; not in the Course of Study word list) car on the TV (set; not in the list). But there are two (toy) cars on it. There was an apple on the table. But there is half of the apple on it. The (note)book was Taro's. But now it is Jiro's. (Waystage) ―――――――――――――――――――――――――――――――――――― ○If the student remains silent for five seconds before he/she starts, point to the ball in the two pictures. ○When the student (a) pauses for 15 seconds or more, (b) speaks for one and a half minutes and stops momentarily, or (c) says “That’s all,” say I: O.K. Thank you. ○Intended responses to be elicited: ・singular and plural forms ・*half* ・possession ・*on, under,* and *in* ○If the student does not speak at all, ask easy questions such as, ”What’s this?” and try to make him/her feel that he/she was able to speak. I: O.K. Thank you. ○Collect the card and give some brief comments.
Wind-down (one minute)	I: By the way, do you have any plans for summer vacation? S: Yes./No. I: Are you going to go swimming? ○Task 5 is very difficult, so ask easy Yes/No questions. Don’t ask difficult questions. I: O.K. That’s all. Thank you. (In Japanese) Please fill out and put this questionnaire into an envelope provided on the table in the hall. ○Stop the two tape recorders. Collect the microphone. Give the student the questionnaire. I: Good bye.

＜Q & A＞

Q: Should I speak very slowly?
A: Try to speak slowly.

Q: What should I do if the student does not appear at his/her allotted test time?
A: Look for him/her in the hall or on the veranda. If the next student has already arrived, you can go first with him/her. I would be glad if you left a message for the student who failed to come.

Q: What should I do if I made a mistake and said something wrong?
A: Just say “Oh, sorry.” and correct it. You don’t need to go back if it’s a minor problem.

Q: What if the student starts reading the card aloud?
A: Tell him/her to stop by using gestures.

Q: What if I don’t understand the student?
A: Just nod and say “Yes.” “Oh.” as if you understand.

Q: Should I be strict about counting the 15 seconds?
A: This is an approximate time, so you have some leeway. But it is my hope that you give the student longer than 15 seconds, rather than less, if you forget to count time for the analysis.

Q: When there is continuous silence for 15 seconds, we say “O.K.” but asking the student “O.K.?” sounds nice to me.
A: I agree that your way is probably better for the student. However, time for each student is restricted. Moreover, some might begin talking again if they are asked “O.K.?” while others might feel that they should stop, so each student will have a different reaction. I would like to take the stance of saying “O.K.” to everyone in order to be fair. But I hope you say “O.K.” in a nice way.

Q: What should I do if I forget to use a stopwatch?
A: Remember to use it. If you forget, then simply calculate the time that has elapsed and begin timing from there. In total, the student should have approximately a minute and a half. When you are unsure about how much time has passed, take more rather than less.

Q: Should I make the sound of a stopwatch when time is up?
A: Let’s try not to make noise.

Q: Is it O.K. to say “Good?” in responding to the student’s utterance?
A: Avoid using words to evaluate. It is of course all right if you say it unintentionally.

Q: Is it O.K. if the time of the student’s speaking test goes beyond 15 minutes?
A: It can happen. But try not to extend the time too much because other students are waiting.

Q: What should I say if the student says “How do you say … in English?”
A: Say “Sorry. I can’t tell you.”

Q: Should I be honest when asked questions?
A: Either is O.K. During Task 3, please respond as if you are a boy.

Q: How should I respond to errors?
A: Grammatical errors and errors in meaning need not be taken heed of. Just continue talking normally. It is the rater’s job to evaluate the student.

Q: What if the student goes into a panic?
A: Try to calm him/her down, saying “It’s O.K.” in English. Point to the Japanese translation or say something in Japanese. I have never come across such the student, so I don’t think we should worry too much about it.

Q: When should I point to the Japanese translation? While speaking in English, should I point to it at the same time or later?
A: Do it at the same time in order to save time.

Q: What if I forget to point to the Japanese translations when I say it a second time?
A: Point to the translations when you notice. If the student understands, there's no need to use the translations again.

Q: Is saying "Mm…" considered silence or is it speaking?
A: It is considered silence.
Examples of words that don't convey meaning: *um, mm, eetto, doushiyou* (in Japanese), *wait a minute*
Examples of words that are not considered silence: *well, let me see, let's see*
When it is unclear whether the student said *um* or *and*, consider it *and*.

Q: What should I do if the student misunderstands what to do or if he/she says "I don't understand."
A: Tell them the task if it is written on the paper by pointing to the description. When it is not written, write it on the paper and show it to him/her.

Q: How should I react if I feel that the student does not speak because he/she does not understand the task? For example, in the self-introduction, the student sticks to one topic and struggles to keep on talking about that particular topic.
A: If it lasts for more than five seconds, point to the description of the task.

Q: What should I do if the student asks questions when he/she should be speaking about him/herself? For instance, during the self-introduction task, the student says "I like English. Do you like English?"
A: Answer it concisely with a word or two, and ask him/her to keep on talking by using gestures.

Q: Between tasks we should talk briefly in order to relax the student. Is there anything to be careful about?
A: Try to ask Yes/No questions and to make them easy. WH-Questions are rather difficult and some students may be at a loss for an answer.

Q: When I say, "Let's begin." the student asks me to wait.
A: Wait for another five seconds.

Q: What should I do if a school announcement is heard during the speaking test?
A: Wait until it is finished. Have the student wait, too. If you continue talking, the test will probably not be assessed due to two much background noise on the tape.

Appendix 4.1

Item measurement report of the Receptive Vocabulary Knowledge Test

No.	Answer	1st M		InfitMS	InfitZ	2nd M		InfitMS	InfitZ
1	baseball	-4.45	0.71	0.87	-0.20	-4.45	0.71	0.86	-0.21
2	gun	-1.22	0.19	0.87	-1.15	-1.19	0.19	0.87	-1.09
3	meaning	0.54	0.14	0.74	-4.51	0.63	0.14	0.74	-4.30
4	onion	-2.38	0.28	0.91	-0.40	-2.36	0.29	0.90	-0.44
5	baby	-3.31	0.42	0.86	-0.39	-3.30	0.42	0.86	-0.40
6	influence	1.20	0.13	1.44	6.18	excluded			
7	household	0.22	0.14	0.87	-1.95	0.30	0.15	0.90	-1.49
8	phone	-1.22	0.19	0.85	-1.33	-1.19	0.19	0.87	-1.15
9	age	-0.36	0.16	0.78	-2.98	-0.30	0.16	0.78	-2.95
10	further	0.88	0.14	1.09	1.36	0.99	0.14	1.15	2.25
11	abroad	3.09	0.15	1.38	4.51	3.37	0.15	1.52	5.58
12	meal	0.63	0.14	1.01	0.19	0.73	0.14	1.05	0.76
13	note	-0.08	0.15	0.93	-0.93	-0.01	0.15	0.96	-0.62
14	harm	2.20	0.14	0.79	-3.37	2.41	0.14	0.81	-2.75
15	budget	2.77	0.14	1.23	2.92	3.03	0.15	1.32	3.70
16	face	-1.30	0.20	0.85	-1.21	-1.26	0.20	0.87	-1.05
17	work	-0.76	0.17	0.78	-2.49	-0.72	0.17	0.79	-2.31
18	bite	0.50	0.14	0.95	-0.75	0.59	0.14	1.00	-0.01
19	employment	1.33	0.13	1.03	0.55	1.47	0.14	1.11	1.60
20	exam	0.65	0.14	0.80	-3.49	0.75	0.14	0.80	-3.28
21	sick	-1.34	0.20	0.87	-1.08	-1.30	0.20	0.85	-1.26
22	suffer	2.39	0.14	1.11	1.48	2.62	0.14	1.21	2.59
23	criticize	2.39	0.14	1.87	9.00	excluded			
24	have	-1.85	0.23	0.92	-0.50	-1.82	0.24	0.93	-0.42
25	teacher	-2.46	0.29	0.86	-0.61	-2.45	0.30	0.86	-0.61
26	necessary	0.69	0.14	0.91	-1.44	0.79	0.14	0.94	-0.96
27	distinguish	1.67	0.13	1.32	4.50	1.84	0.14	1.44	5.47
28	listen	-1.55	0.21	0.79	-1.53	-1.52	0.21	0.80	-1.50
29	fall	0.32	0.14	0.80	-3.33	0.41	0.15	0.82	-2.80
30	evidence	1.44	0.13	0.93	-1.08	1.59	0.14	0.99	-0.10
31	welcome	-1.05	0.18	0.82	-1.73	-1.01	0.19	0.83	-1.59
32	large	-1.30	0.20	0.94	-0.48	-1.26	0.20	0.96	-0.36
33	pull	-1.50	0.21	0.88	-0.88	-1.47	0.21	0.90	-0.77
34	particular	2.03	0.14	0.84	-2.45	2.23	0.14	0.90	-1.42
35	practical	1.36	0.13	1.28	4.05	1.51	0.14	1.39	5.11
36	complete	0.22	0.14	0.85	-2.31	0.30	0.15	0.87	-2.00
37	troop	3.04	0.15	1.17	2.14	3.33	0.15	1.28	3.23
38	serious	0.71	0.14	0.81	-3.22	0.81	0.14	0.83	-2.74
39	behind	-1.15	0.19	0.92	-0.73	-1.11	0.19	0.94	-0.54
40	my	-3.00	0.37	0.87	-0.41	-2.99	0.37	0.86	-0.46
M		0.00	0.19	1.00	-0.30	0.00	0.20	1.00	-0.20
SD		1.81	0.10	0.20	2.80	1.88	0.11	0.20	2.40

Note. 1st *M* = Item difficulty in the first analysis; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic

standardized toward a unit-normal distribution";

Appendix 4.2
Item measurement report of the Grammar Test

No.		1st M	*SE*	InfitMS	InfitZ	2nd M	*SE*	InfitMS	InfitZ
1	Don't	-0.58	0.14	1.05	0.72	-0.52	0.15	1.09	1.36
2	Does	-0.81	0.15	0.85	-2.36	-0.76	0.15	0.86	-2.23
3	written	0.20	0.14	1.00	-0.05	0.30	0.14	1.05	0.75
4	Whose	-0.96	0.15	0.80	-3.16	-0.91	0.15	0.81	-3.02
6	be	-0.92	0.15	0.80	-3.24	-0.87	0.15	0.79	-3.35
7	who	0.24	0.14	0.82	-3.00	0.34	0.14	0.86	-2.11
8	He	-1.90	0.18	1.04	0.43	-1.87	0.18	1.06	0.58
9	where	1.19	0.15	1.46	5.30	excluded			
10	look	0.54	0.14	1.00	-0.02	0.65	0.15	1.04	0.61
11	myself	-0.27	0.14	0.91	-1.55	-0.20	0.14	0.92	-1.25
12	if	2.13	0.17	1.22	2.32	2.38	0.18	1.36	3.35
13	built	0.77	0.14	0.97	-0.40	0.89	0.15	1.00	-0.01
14	faster	-0.39	0.14	1.08	1.28	-0.33	0.15	1.12	1.87
15	inviting	0.77	0.14	0.95	-0.71	0.89	0.15	0.99	-0.19
M		0.00	0.15	1.00	-0.30	0.00	0.15	1.00	-0.30
SD		1.00	0.01	0.20	2.30	1.04	0.01	0.10	1.90

Note. 1st *M* = Item difficulty in the first analysis; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic standardized toward a unit-normal distribution";

Appendix 4.3

Item measurement report of the Productive Vocabulary Knowledge Test by Method 1

No.	Answer	1st M		InfitMS	InfitZ	2nd M		InfitMS	InfitZ
1	lunch	-3.10	0.13	1.07	0.90	-3.26	0.13	1.11	1.43
2	weight	0.47	0.12	0.66	-5.43	0.41	0.12	0.67	-5.15
3	behind	0.18	0.12	1.13	1.73	0.12	0.12	1.16	2.14
4	year	-3.35	0.14	0.70	-4.34	-3.52	0.14	0.72	-3.91
5	bag	-3.06	0.13	1.61	7.23	excluded			
6	blame	5.29	0.20	0.99	-0.06	5.28	0.20	1.00	-0.02
7	sick	-2.05	0.12	1.12	1.98	-2.17	0.12	1.15	2.36
8	sound	-1.46	0.12	1.00	0.00	-1.57	0.12	1.03	0.49
9	box	-5.70	0.22	1.22	1.45	-6.03	0.23	1.36	2.18
10	fight	-0.52	0.12	0.85	-2.40	-0.60	0.12	0.86	-2.06
11	orange	-1.06	0.12	0.98	-0.35	-1.16	0.12	1.01	0.15
12	change	0.32	0.12	1.01	0.14	0.26	0.12	1.03	0.44
13	peach	excluded				excluded			
14	length	2.84	0.13	0.74	-4.51	2.83	0.13	0.73	-4.54
15	vegetable	2.34	0.12	0.99	-0.18	2.32	0.12	1.00	-0.01
16	must	-2.30	0.12	0.97	-0.47	-2.43	0.12	0.98	-0.25
17	feeling	2.68	0.13	0.81	-3.11	2.66	0.13	0.82	-2.89
18	fail	2.48	0.12	0.93	-1.09	2.46	0.13	0.94	-0.96
19	crack	5.81	0.24	0.85	-0.91	5.81	0.24	0.85	-0.91
20	nice	-2.15	0.12	1.30	4.56	-2.27	0.12	1.34	5.03
21	touch	0.70	0.12	0.89	-1.66	0.65	0.12	0.90	-1.41
22	yellow	-0.88	0.12	0.94	-1.01	-0.97	0.12	0.96	-0.56
23	corn	-0.46	0.12	1.31	4.18	-0.54	0.12	1.34	4.50
24	star	-2.73	0.13	1.14	1.96	-2.87	0.13	1.17	2.43
25	move	-1.77	0.12	0.97	-0.46	-1.89	0.12	0.99	-0.12
26	cheap	2.35	0.12	0.95	-0.80	2.33	0.12	0.95	-0.74
27	stir	6.63	0.32	0.88	-0.49	6.63	0.32	0.88	-0.50
28	rabbit	0.27	0.12	1.06	0.84	0.20	0.12	1.09	1.16
29	read	-3.52	0.14	0.80	-2.78	-3.70	0.14	0.81	-2.49
30	funny	2.07	0.12	0.86	-2.25	2.04	0.12	0.87	-2.07
31	lend	1.13	0.12	0.71	-4.71	1.09	0.12	0.73	-4.41
32	white	-1.33	0.12	0.93	-1.18	-1.43	0.12	0.95	-0.78
33	popular	-0.22	0.12	1.03	0.43	-0.29	0.12	1.06	0.79
34	chicken	2.55	0.13	1.10	1.53	2.54	0.13	1.11	1.66
35	stone	0.15	0.12	0.65	-5.65	0.09	0.12	0.66	-5.37
36	point	-1.34	0.12	0.90	-1.68	-1.44	0.12	0.92	-1.38
37	plant	-0.35	0.12	0.67	-5.39	-0.42	0.12	0.68	-5.16
38	notice	1.39	0.12	0.79	-3.37	1.35	0.12	0.80	-3.21
39	ear	-1.14	0.12	1.13	1.96	-1.24	0.12	1.17	2.53
40	month	-1.16	0.12	1.06	0.92	-1.26	0.12	1.08	1.24
M		0.00	0.13	1.00	-0.60	0.00	0.14	1.00	-0.50
SD		2.60	0.04	0.20	2.80	2.64	0.04	0.20	2.60

Note. 1st *M* = Item difficulty in the first analysis; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic standardized toward a unit-normal distribution";

Appendix 4.4

Item measurement report of the Productive Vocabulary Knowledge Test by Method 2

No.	Answer	1st M		InfitMS	InfitZ	2nd M		InfitMS	InfitZ
1	lunch	-3.36	0.15	0.71	-3.75	-3.52	0.15	-3.52	0.15
2	weight	0.78	0.12	0.62	-6.06	0.73	0.12	0.73	0.12
3	behind	0.52	0.12	1.06	0.83	0.46	0.12	0.46	0.12
4	year	-3.02	0.14	0.77	-3.20	-3.17	0.14	-3.17	0.14
5	bag	-2.78	0.13	1.56	6.60	excluded			
6	blame	5.81	0.21	0.98	-0.12	5.84	0.21	5.84	0.21
7	sick	-1.71	0.12	1.06	1.08	-1.82	0.12	-1.82	0.12
8	sound	-1.45	0.12	1.16	2.62	-1.56	0.12	-1.56	0.12
9	box	-5.30	0.23	1.25	1.53	-5.53	0.23	-5.53	0.23
10	fight	-0.32	0.12	0.81	-3.24	-0.40	0.12	-0.40	0.12
11	orange	-1.78	0.12	0.91	-1.66	-1.89	0.12	-1.89	0.12
12	change	0.31	0.12	1.17	2.27	0.24	0.12	0.24	0.12
13	peach	excluded				excluded			
14	length	3.25	0.13	0.71	-4.98	3.26	0.13	3.26	0.13
15	vegetable	1.17	0.12	0.86	-2.06	1.13	0.12	1.13	0.12
16	must	-1.85	0.12	0.94	-1.09	-1.97	0.12	-1.97	0.12
17	feeling	3.15	0.13	0.84	-2.54	3.16	0.13	3.16	0.13
18	fail	2.92	0.13	0.96	-0.70	2.92	0.13	2.92	0.13
19	crack	6.29	0.23	0.85	-0.98	6.32	0.23	6.32	0.23
20	nice	-1.76	0.12	1.29	4.57	-1.88	0.12	-1.88	0.12
21	touch	0.10	0.12	1.24	3.32	0.03	0.12	0.03	0.12
22	yellow	-1.12	0.12	0.86	-2.47	-1.22	0.12	-1.22	0.12
23	corn	-1.10	0.12	1.35	5.39	-1.19	0.12	-1.19	0.12
24	star	-2.95	0.14	1.19	2.41	-3.09	0.14	-3.09	0.14
25	move	-1.45	0.12	0.92	-1.48	-1.56	0.12	-1.56	0.12
26	cheap	0.98	0.12	0.71	-4.53	0.94	0.12	0.94	0.12
27	stir	7.21	0.31	0.76	-1.15	7.24	0.31	7.24	0.31
28	rabbit	-0.97	0.12	0.88	-2.09	-1.07	0.12	-1.07	0.12
29	read	-3.08	0.14	0.82	-2.40	-3.23	0.14	-3.23	0.14
30	funny	2.03	0.12	0.89	-1.79	2.01	0.12	2.01	0.12
31	lend	1.55	0.12	0.78	-3.45	1.53	0.12	1.53	0.12
32	white	-0.96	0.12	0.90	-1.76	-1.06	0.12	-1.06	0.12
33	popular	-0.70	0.12	1.02	0.35	-0.79	0.12	-0.79	0.12
34	chicken	1.42	0.12	1.20	2.72	1.39	0.12	1.39	0.12
35	stone	0.00	0.12	0.95	-0.75	-0.07	0.12	-0.07	0.12
36	point	-0.92	0.12	0.84	-2.85	-1.01	0.12	-1.01	0.12
37	plant	-0.72	0.12	0.90	-1.73	-0.80	0.12	-0.80	0.12
38	notice	1.82	0.12	0.78	-3.48	1.80	0.12	1.80	0.12
39	ear	-1.14	0.12	0.99	-0.12	-1.24	0.12	-1.24	0.12
40	month	-0.85	0.12	1.10	1.68	-0.94	0.12	-0.94	0.12
M		0.00	0.13	1.00	-0.60	0.00	0.14	1.00	-0.50
SD		2.63	0.04	0.20	2.80	2.68	0.04	0.20	2.70

Note. 1st *M* = Item difficulty in the first analysis; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic standardized toward a unit-normal distribution";

Appendix 4.5

Item measurement report of the Productive Vocabulary Knowledge Test by Method 3

No.	Answer	1st M		InfitMS	InfitZ	2nd M		InfitMS	InfitZ
1	lunch	-2.93	0.15	0.81	-2.27	-2.97	0.15	0.81	-2.31
2	weight	0.28	0.11	0.77	-4.00	0.28	0.12	0.79	-3.65
3	behind	1.16	0.12	1.04	0.63	1.19	0.12	1.08	1.06
4	year	-2.73	0.15	0.80	-2.47	-2.77	0.15	0.80	-2.58
5	bag	-2.35	0.14	1.37	4.28	-2.39	0.14	1.37	4.35
6	blame	6.14	0.19	0.96	-0.35	6.28	0.19	0.97	-0.24
7	sick	-1.01	0.12	1.05	0.86	-1.03	0.12	1.07	1.22
8	sound	-0.86	0.12	1.10	1.81	-0.88	0.12	1.12	2.11
9	box	-4.58	0.23	1.24	1.36	-4.63	0.23	1.23	1.32
10	fight	-0.48	0.11	0.96	-0.73	-0.49	0.12	0.96	-0.69
11	orange	-2.46	0.14	0.91	-1.11	-2.51	0.14	0.93	-0.85
12	change	0.61	0.12	1.28	3.92	0.62	0.12	1.33	4.52
13	peach	excluded				excluded			
14	length	3.69	0.13	0.65	-6.13	3.81	0.13	0.65	-5.94
15	vegetable	-1.49	0.12	0.89	-1.80	-1.52	0.12	0.90	-1.59
16	must	-1.28	0.12	0.98	-0.41	-1.30	0.12	0.99	-0.16
17	feeling	3.69	0.13	0.80	-3.23	3.81	0.13	0.81	-3.01
18	fail	3.31	0.12	0.91	-1.47	3.42	0.13	0.92	-1.31
19	crack	6.33	0.20	1.04	0.30	6.48	0.20	1.04	0.32
20	nice	-1.06	0.12	1.26	4.20	-1.09	0.12	1.29	4.61
21	touch	-1.43	0.12	0.84	-2.72	-1.46	0.12	0.86	-2.47
22	yellow	-2.62	0.14	0.91	-1.09	-2.67	0.14	0.93	-0.89
23	corn	-1.30	0.12	1.17	2.75	-1.33	0.12	1.20	3.13
24	star	-3.14	0.16	0.87	-1.46	-3.19	0.16	0.87	-1.37
25	move	-0.73	0.12	0.92	-1.49	-0.75	0.12	0.92	-1.55
26	cheap	1.40	0.12	0.82	-2.76	1.44	0.12	0.84	-2.33
27	stir	7.52	0.29	0.88	-0.62	7.66	0.29	0.88	-0.62
28	rabbit	-1.91	0.13	0.88	-1.75	-1.95	0.13	0.89	-1.66
29	read	-2.71	0.15	0.79	-2.66	-2.75	0.15	0.78	-2.88
30	funny	2.53	0.12	0.84	-2.53	2.62	0.12	0.86	-2.16
31	lend	2.13	0.12	0.77	-3.65	2.20	0.12	0.77	-3.57
32	white	-1.69	0.12	1.31	4.33	-1.72	0.13	1.33	4.63
33	popular	-0.82	0.12	0.95	-0.89	-0.84	0.12	0.96	-0.70
34	chicken	0.37	0.11	1.50	6.96	excluded			
35	stone	-1.98	0.13	1.19	2.60	-2.02	0.13	1.20	2.69
36	point	-0.23	0.11	0.77	-4.28	-0.24	0.12	0.77	-4.23
37	plant	-0.07	0.11	0.83	-3.09	-0.08	0.12	0.84	-2.81
38	notice	2.39	0.12	0.74	-4.37	2.47	0.12	0.76	-3.74
39	ear	-1.04	0.12	0.88	-2.17	-1.06	0.12	0.90	-1.71
40	month	-0.61	0.11	0.97	-0.57	-0.63	0.12	0.97	-0.45
M		0.00	0.14	1.00	-0.70	0.00	0.14	1.00	-0.70
SD		2.72	0.04	0.20	2.80	2.82	0.04	0.20	2.60

Note. 1st *M* = Item difficulty in the first analysis; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic

standardized toward a unit-normal distribution";

Appendix 5.1

Examples of the Speaking Performance Measures From Task 4 (About a Favorite TV Program)

|My favorite is TV is SMAP SMAP :: because it is very interesting.| |It is a very popular TV.| |{Maybe mo} maybe {most of} most of people in my school {look it} look it.|

[Silence for 15 seconds and this task was finished; Speech time: 43 seconds; | | = AS-unit (3); :: = clause (4); { } = dysfluency marker (6); = error (4)]

F1: 27 / 43*60 = 37.67
F3: 4 / 43*60 = 5.58
F5: 6 / 43*60 = 8.37
F6: 6 / 27 = 0.22
F7: 6 / 3 = 2.00
A1: 1 / 4= 0.25
A2: 0 / 3 = 0.00
A3: 4 / 27 = 0.15
A4: 4 / 3 = 1.33

SC1: 4 / 3 = 1.33
SC2: 1 / 3 = 0.33
SC3/F2: 27 / 3 = 9.00
LC1: 18 / √27 = 3.46
LC2: 14 / 27 = 0.52 (favorite, TV, SMAP, SMAP, very, interesting, very, popular, TV, maybe, most, people, school, look)
LC3: (1 + 13*0.5) / 27 = 0.28
LC4: 1 / 27 = 0.04 (favorite)

Appendix 5.2.

Correlations between 17 Speaking Performance Measures within the same task

	F1	F3	F4	F5	F6	A1	A2	A3	A4	SC1	SC2	SC3
F1	--	.93**	-.21	-.35**	-.28*	.42**	.46**	-.47**	-.41**	.05	.00	.14
F3	.92**	--	-.27*	-.38**	-.35**	.48**	.51**	-.49**	-.49**	.22	-.03	-.07
F4	-.19	-.18	--	.88**	.94**	-.05	-.03	.00	.12	.16	.10	.42**
F5	-.32**	-.24*	.88**	--	.95**	-.04	-.02	.11	.14	.01	.02	.17
F6	-.26*	-.26*	.90**	.86**	--	-.08	-.07	.06	.19	.09	.13	.41**
A1	.03	.10	-.01	-.01	.01	--	.91**	-.74**	-.77**	.07	-.03	-.20
A2	.01	.10	.07	.10	.04	.82**	--	-.78**	-.83**	.09	-.08	-.21
A3	-.21	-.17	.00	.05	-.04	-.77**	-.66**	--	.87**	-.29*	-.08	-.09
A4	-.10	-.16	-.02	-.05	.04	-.69**	-.81**	.76**	--	-.16	.02	.31**
SC1	.15	.10	-.01	-.11	.11	.14	-.17	-.19	.28*	--	.32**	.34**
SC2	.13	.04	-.02	-.12	.07	.13	-.18	-.11	.30**	.84**	--	.44**
SC3	.18	-.06	.05	-.16	.19	.04	-.27*	-.23	.32**	.77**	.74**	--
LC1	-.01	-.21	.27*	.01	.27*	.24*	.09	-.29*	.00	.21	.22	.56**
LC2	.04	.06	-.29*	-.22	-.24*	-.21	-.01	.10	-.04	-.23*	-.19	-.22
LC3	.05	.05	-.29*	-.26*	-.26*	-.26*	-.09	.16	.05	-.17	-.14	-.16
LC4	.06	.04	-.21	-.23*	-.21	-.20	-.17	.17	.17	.00	.01	.01

Note. $n = 74$. Intercorrelations in Task 1 are presented above the diagonal, and intercorrelations in Task 4 are presented below the diagonal.
*$p < .05$. **$p < .01$.

	LC1	LC2	LC3	LC4			T5F1	T5F3	T5F4	T5F5	T5F6	T5A1	T5A2	T5A3	T5A4
F1	-.20	-.04	-.11	-.13		T5F1	--								
F3	-.36**	.02	.02	.01		T5F3	.68**	--							
F4	.34**	-.25*	-.26*	-.12		T5F4	-.10	-.11	--						
F5	.13	-.11	-.15	-.09		T5F5	-.30**	-.16	.45**	--					
F6	.27*	-.22	-.26*	-.14		T5F6	-.29*	-.19	.80**	.88**	--				
A1	-.38**	-.03	-.07	-.08		T5A1	.36**	.17	.21	-.05	.04	--			
A2	-.41**	-.10	-.10	-.06		T5A2	.34**	.16	.21	-.06	.04	.97**	--		
A3	.18	.34**	.22	.02		T5A3	-.23	.02	-.33**	.01	-.15	-.54**	-.56**	--	
A4	.39**	.12	.04	-.03		T5A4	-.24*	-.22	-.18	.03	-.07	-.31**	-.33**	.57**	--
SC1	-.01	-.38**	-.23*	.00		T5SC1	.22	.50**	.27*	.11	.19	.39**	.37**	-.32**	-.32**
SC2	.20	-.29*	-.27*	-.10		T5SC2	-.06	-.08	-.08	-.05	-.06	-.05	-.06	-.11	-.10
SC3	.54**	-.50**	-.48**	-.24*		T5SC3	.32**	-.10	.22	-.25*	-.04	.47**	.48**	-.66**	-.29*
LC1	--	-.19	-.14	-.04		T5LC1	.01	-.31**	.09	-.22	-.11	.19	.21	-.27*	-.01
LC2	-.36**	--	.78**	.25*		T5LC2	-.08	.13	-.28*	-.31**	-.34**	-.38**	-.38**	.73**	.36**
LC3	-.29*	.87**	--	.77**		T5LC3	-.04	.23*	-.23*	-.27*	-.28*	-.35**	-.35**	.73**	.28*
LC4	-.05	.22	.65**	--		T5LC4	-.02	.37**	-.06	-.07	-.07	-.23*	-.24*	.53**	.04

	T5SC1	T5SC2	T5SC3	T5LC1	T5LC2	T5LC3	T5LC4
T5SC1	--						
T5SC2	-.07	--					
T5SC3	.28	.12	--				
T5LC1	-.24*	.09	.31**	--			
T5LC2	-.29*	-.05	-.47**	-.08	--		
T5LC3	-.21	.00	-.49**	-.17	.95**	--	
T5LC4	.03	.15	-.46**	-.30**	.62**	.82**	--

Appendix 5.3

Correlations between 17 Speaking Performance Measures among different tasks

	T4F1	T4F3	T4F4	T4F5	T4F6	T4A1	T4A2	T4A3	T4A4	T4SC1	T4SC2	T4SC3
T1F1	.49**	.38**	-.04	-.13	-.08	-.12	.03	-.13	-.05	.10	.07	.14
T1F3	.44**	.36**	-.11	-.16	-.16	-.20	-.07	.00	.07	.16	.17	.11
T1F4	-.13	-.22	.45**	.35**	.54**	.02	-.13	-.07	.10	.10	.10	.32**
T1F5	-.25*	-.28*	.32**	.34**	.45**	-.03	-.18	.02	.13	.00	-.03	.16
T1F6	-.20	-.26*	.45**	.40**	.58**	-.01	-.14	-.04	.10	.05	.01	.25*
T1A1	.31**	.26*	-.25*	-.28*	-.25*	-.09	-.06	-.01	.03	-.02	.06	.03
T1A2	.30**	.26*	-.23*	-.26*	-.25*	-.06	-.01	.00	-.02	-.10	-.03	-.05
T1A3	-.39**	-.27*	.17	.40**	.19	.01	.05	.11	.03	-.06	-.10	-.22
T1A4	-.35**	-.28*	.24*	.34**	.31**	-.01	-.02	.07	.08	.04	-.02	-.03
T1SC1	.13	.05	.01	-.13	-.03	.01	-.10	-.08	.04	.29*	.31**	.34**
T1SC2	.11	.03	.08	-.01	.07	.09	-.02	-.05	.08	.24*	.15	.33**
T1SC3	.15	.00	.34**	.10	.40**	.18	.08	-.32**	-.10	.24*	.16	.47**
T1LC1	-.02	-.09	.40**	.23*	.37**	.27*	.11	-.23*	-.10	.09	.11	.27*
T1LC2	-.07	.04	-.04	.19	-.01	-.30**	-.22	.35**	.24*	-.21	-.12	-.30**
T1LC3	-.06	.05	-.06	.09	-.06	-.21	-.17	.42**	.31**	-.13	.01	-.25*
T1LC4	-.02	.04	-.04	-.05	-.06	-.05	-.10	.34**	.30**	.05	.19	-.04

Note. $n = 74$. $^{*}p < .05$. $^{**}p < .01$.

	T4LC1	T4LC2	T4LC3	T4LC4
T1F1	.07	.21	.20	.08
T1F3	-.06	.21	.21	.09
T1F4	.52**	-.31**	-.23*	-.03
T1F5	.33**	-.26*	-.18	.01
T1F6	.45**	-.31**	-.25*	-.07
T1A1	-.03	.06	.05	.03
T1A2	-.01	.07	.09	.10
T1A3	-.16	-.03	-.04	-.04
T1A4	.06	-.07	-.11	-.10
T1SC1	.17	-.16	-.14	-.04
T1SC2	.20	-.16	-.16	-.06
T1SC3	.59**	-.25*	-.24*	-.11
T1LC1	.59**	-.30**	-.35**	-.23*
T1LC2	-.46**	.35**	.25*	-.02
T1LC3	-.39**	.19	.16	.06
T1LC4	-.14	-.03	.04	.13

	T5F1	T5F3	T5F4	T5F5	T5F6	T5A1	T5A2	T5A3	T5A4
T1F1	.07	-.07	-.06	-.07	-.07	.08	.08	-.03	-.08
T1F3	.06	-.04	-.10	-.06	-.08	.03	.03	.02	-.03
T1F4	.09	-.03	.37**	-.02	.18	.16	.18	-.39**	-.26*
T1F5	.02	-.06	.23*	-.04	.11	.03	.04	-.23*	-.14
T1F6	.06	-.04	.31**	-.02	.16	.10	.10	-.31**	-.23*
T1A1	.03	.08	-.19	-.10	-.14	.02	.02	.13	.05
T1A2	-.03	.04	-.17	-.05	-.10	.05	.06	.06	-.02
T1A3	-.17	-.11	.06	.03	.03	-.19	-.20	.08	.14
T1A4	-.10	-.13	.15	.00	.07	-.10	-.10	-.04	.01
T1SC1	.20	.19	.16	.02	.11	.28*	.27*	-.27*	-.34**
T1SC2	.25*	.11	.02	-.01	-.01	.49**	.48**	-.16	-.21
T1SC3	.25*	.04	.27*	-.07	.08	.38**	.38**	-.44**	-.43**
T1LC1	.13	-.05	.36**	-.03	.14	.18	.22	-.35**	-.30**
T1LC2	-.20	-.01	.04	.11	.10	-.45**	-.44**	.30**	.24*
T1LC3	-.07	.02	.12	.12	.16	-.34**	-.31**	.11	.14
T1LC4	.11	.10	.13	.10	.14	-.05	-.03	-.11	-.03

	T5SC1	T5SC2	T5SC3	T5LC1	T5LC2	T5LC3	T5LC4
T1F1	.03	.14	.18	-.08	-.01	-.02	-.04
T1F3	.05	.11	.12	-.25*	.04	.03	.00
T1F4	.17	-.06	.38**	.20	-.22	-.20	-.13
T1F5	.00	-.01	.20	.14	-.09	-.08	-.08
T1F6	.07	-.02	.30**	.17	-.17	-.15	-.09
T1A1	.04	.20	-.01	-.28*	.08	.11	.12
T1A2	.08	.22	.04	-.29*	.05	.10	.17
T1A3	-.14	-.18	-.26*	.11	.08	.01	-.11
T1A4	-.16	-.19	-.08	.28*	.03	-.02	-.11
T1SC1	.21	-.05	.23	-.11	-.13	-.05	.14
T1SC2	.13	-.05	.18	-.05	-.15	-.11	-.01
T1SC3	.12	.00	.48**	.40**	-.27*	-.23*	-.10
T1LC1	.05	-.07	.30**	.38**	-.30**	-.29*	-.19
T1LC2	.01	.10	-.36**	-.32**	.18	.16	.10
T1LC3	.08	.09	-.19	-.28*	.08	.07	.07
T1LC4	.18	.05	.05	-.18	-.02	-.01	.05

	T5F1	T5F3	T5F4	T5F5
T4F1	.18	-.02	-.02	.05
T4F3	.02	-.10	-.08	.12
T4F4	-.03	.00	.53**	.07
T4F5	-.14	-.03	.38**	.05
T4F6	-.04	-.04	.48**	.06
T4A1	.11	-.10	-.21	-.27
T4A2	-.03	-.11	-.16	-.19
T4A3	-.15	.04	.12	.22
T4A4	.02	.05	.14	.07
T4SC1	.16	-.11	-.05	-.08
T4SC2	.24*	.04	-.04	-.10
T4SC3	.34**	.07	.14	-.12
T4LC1	.32**	.03	.31**	-.14
T4LC2	-.22	.05	-.24*	.10
T4LC3	-.16	.09	-.16	.30*
T4LC4	.01	.08	-.01	.43**

	T5F6	T5A1	T5A2	T5A3	T5A4	T5SC1	T5SC2	T5SC3	T5LC1	T5LC2	T5LC3	T5LC4
T4F1	.01	.19	.19	-.20	-.16	.06	.57**	.25*	-.03	-.16	-.11	.01
T4F3	.03	.03	.02	-.13	-.11	-.04	.76**	.09	-.05	-.13	-.09	.04
T4F4	.29*	.09	.09	-.27*	-.21	.28*	-.09	.14	.17	-.16	-.18	-.14
T4F5	.21	-.09	-.09	-.16	-.11	.16	-.07	-.05	.07	-.10	-.14	-.13
T4F6	.27*	.03	.04	-.27*	-.23	.18	-.05	.19	.21	-.17	-.19	-.15
T4A1	-.29*	.19	.19	-.34**	-.21	-.11	.22	.23*	.33**	-.35**	-.33**	-.22
T4A2	-.22	.02	.01	-.10	.00	-.20	.20	.02	.21	-.17	-.16	-.11
T4A3	.19	-.22	-.22	.38**	.27*	.10	-.19	-.30**	-.36**	.40**	.40**	.30**
T4A4	.11	.00	.02	.10	.05	.18	-.18	.05	-.22	.27*	.22	.08
T4SC1	-.07	.28*	.29*	-.35**	-.30*	.09	-.01	.42**	.07	-.23*	-.31**	-.30**
T4SC2	-.08	.24*	.26*	-.31**	-.28*	.24*	-.08	.32**	-.07	-.19	-.23*	-.19
T4SC3	.01	.51**	.51**	-.41**	-.30**	.27*	-.12	.60**	.16	-.28*	-.29*	-.23*
T4LC1	.04	.46**	.48**	-.46**	-.30*	.23	-.17	.56**	.35**	-.35**	-.32**	-.26*
T4LC2	-.02	-.34**	-.31**	.59**	.32**	-.17	.07	-.41**	-.22	.44**	.46**	.39**
T4LC3	.15	-.23*	-.22	.44**	.33**	-.06	.01	-.32**	-.23*	.32**	.32**	.24*
T4LC4	.30*	.02	.00	.00	.21	.09	-.08	-.03	-.16	-.04	-.09	-.13

Appendix 6.1

Correlations between productive vocabulary knowledge and speaking performance

	Task 1	Task 4	Task 5
Productive & F1	-.01	.11	.25*
Productive & F3	-.09	-.02	-.01
Productive & F4	.42**	.27*	.35**
Productive & F5	.20	.05	-.09
Productive & F6	.33**	.32**	.10
Productive & A1	-.17	.35**	.46**
Productive & A2	-.11	.14	.49**
Productive & A3	-.13	-.33**	-.55**
Productive & A4	.08	-.07	-.41**
Productive & SC1	.24*	.32**	.28*
Productive & SC2	.16	.33**	.08
Productive & SC3	.58**	.54**	.65**
Productive & LC1	-.42**	-.20	-.37**
Productive & LC2	.52**	.65**	.47**
Productive & LC3	-.38**	-.31**	-.37**
Productive & LC4	-.20	-.24*	-.34**
Productive & LC5	.05	-.02	-.17

Note. $n = 73$.

*$p < .05$. **$p < .01$.

Appendix 6.2

Regression Analysis Summary for productive vocabulary knowledge predicting each Speaking Performance Measure

Variable	B	*SEB*	*β*	*t*	*p*	R^2	SEE	*F*	*p*
(Constant)	46.83	3.78		12.39	0.00				
T1F1	-0.12	1.73	-0.01	-0.07	0.95	0.01	30.94	0.00	0.95
(Constant)	9.65	0.83		11.64	0.00				
T1F3	-0.27	0.38	-0.09	-0.72	0.47	0.01	6.79	0.52	0.47
(Constant)	4.76	0.50		9.55	0.00				
T1F5	0.89	0.23	0.42	3.90	0.00	0.18	4.09	15.23	0.00
(Constant)	0.13	0.02		8.28	0.00			2.95	
T1F6	0.01	0.01	0.20	1.72	0.09	0.04	0.13		0.09
(Constant)	0.72	0.08		8.66	0.00			8.77	
T1F7	0.11	0.04	0.33	2.96	0.00	0.11	0.68		0.00
(Constant)	0.67	0.03		25.21	0.00				
T1A1	-0.02	0.01	-0.17	-1.44	0.15	0.03	0.22	2.07	0.15
(Constant)	0.60	0.03		19.86	0.00				
T1A2	-0.01	0.01	-0.11	-0.89	0.37	0.01	0.25	0.80	0.37
(Constant)	0.11	0.01		9.65	0.00				
T1A3	-0.01	0.01	-0.13	-1.06	0.29	0.02	0.09	1.13	0.29
(Constant)	0.59	0.06		10.59	0.00				
T1A4	0.02	0.03	0.08	0.66	0.51	0.01	0.46	0.44	0.51
(Constant)	1.10	0.03		39.66	0.00				
T1SC1	0.03	0.01	0.24	2.08	0.04	0.06	0.23	4.32	0.04
(Constant)	0.02	0.01		2.62	0.01				
T1SC2	0.00	0.00	0.16	1.38	0.17	0.03	0.06	1.91	0.17
(Constant)	5.44	0.12		47.23	0.00				
T1SC3	0.32	0.05	0.58	6.00	0.00	0.34	0.94	35.94	0.00
(Constant)	0.76	0.02		42.48	0.00				
T1LC1	-0.03	0.01	-0.42	-3.87	0.00	0.17	0.15	14.96	0.00
(Constant)	3.82	0.11		34.70	0.00				
T1LC2	0.26	0.05	0.52	5.18	0.00	0.27	0.90	26.87	0.00
(Constant)	0.57	0.01		64.32	0.00				
T1LC3	-0.01	0.00	-0.38	-3.51	0.00	0.15	0.07	13.33	0.00
(Constant)	0.33	0.01		46.62	0.00				
T1LC4	-0.01	0.00	-0.20	-1.71	0.09	0.04	0.06	2.93	0.09
(Constant)	0.08	0.01		9.22	0.00				
T1LC5	0.00	0.00	0.05	0.38	0.70	0.00	0.07	0.15	0.70
(Constant)	41.30	3.26		12.65	0.00				
T4F1	1.45	1.49	0.11	0.97	0.33	0.01	26.73	0.94	0.33
(Constant)	8.73	0.92		9.46	0.00				
T4F3	-0.06	0.42	-0.02	-0.14	0.89	0.00	7.56	0.02	0.89
(Constant)	5.37	0.68		7.90	0.00				
T4F5	0.73	0.31	0.27	2.36	0.02	0.07	5.57	5.59	0.02
(Constant)	0.17	0.03		6.36	0.00				
T4F6	0.01	0.01	0.05	0.45	0.65	0.00	0.22	0.20	0.65
(Constant)	0.99	0.12		8.12	0.00				
T4F7	0.16	0.06	0.32	2.88	0.01	0.10	1.00	8.28	0.01
(Constant)	0.51	0.03		14.76	0.00				
T4A1	0.05	0.02	0.35	3.13	0.00	0.12	0.28	9.77	0.00
(Constant)	0.42	0.04		10.81	0.00				
T4A2	0.02	0.02	0.14	1.18	0.24	0.02	0.32	1.40	0.24
(Constant)	0.14	0.01		11.63	0.00				
T4A3	-0.02	0.01	-0.33	-2.96	0.00	0.11	0.10	8.79	0.00
(Constant)	0.80	0.07		11.74	0.00				
T4A4	-0.02	0.03	-0.07	-0.60	0.55	0.01	0.56	0.37	0.55
(Constant)	1.21	0.04		29.91	0.00				
T4SC1	0.05	0.02	0.32	2.80	0.01	0.10	0.33	7.85	0.01
(Constant)	0.24	0.03		6.99	0.00				
T4SC2	0.05	0.02	0.33	2.98	0.00	0.11	0.29	8.89	0.00

(Appendix 6.2 continues)

(Appendix 6.2 continued)

Variable	B	*SEB*	*β*	*t*	*p*	R^2	SEE	*F*	*p*
(Constant)	6.03	0.20		29.75	0.00				
T4SC3	0.50	0.09	0.54	5.43	0.00	0.29	1.66	29.43	0.00
(Constant)	0.77	0.02		42.35	0.00				
T4LC1	-0.01	0.01	-0.20	-1.72	0.09	0.04	0.15	2.97	0.09
(Constant)	3.26	0.08		41.70	0.00				
T4LC2	0.26	0.04	0.65	7.19	0.00	0.42	0.64	51.75	0.00
(Constant)	0.57	0.01		42.27	0.00				
T4LC3	-0.02	0.01	-0.31	-2.76	0.01	0.10	0.11	7.64	0.01
(Constant)	0.31	0.01		35.81	0.00				
T4LC4	-0.01	0.00	-0.24	-2.08	0.04	0.06	0.07	4.31	0.04
(Constant)	0.05	0.01		6.03	0.00				
T4LC5	0.00	0.00	-0.02	-0.18	0.86	0.00	0.07	0.03	0.86
(Constant)	36.09	2.39		15.10	0.00				
T5F1	2.40	1.09	0.25	2.20	0.03	0.06	19.57	4.83	0.03
(Constant)	5.04	0.61		8.22	0.00				
T5F3	-0.01	0.28	-0.01	-0.05	0.96	0.00	5.02	0.00	0.96
(Constant)	6.26	0.78		8.01	0.00				
T5F5	1.13	0.36	0.35	3.15	0.00	0.12	6.41	9.90	0.00
(Constant)	0.26	0.09		2.94	0.00				
T5F6	-0.03	0.04	-0.09	-0.80	0.43	0.01	0.74	0.64	0.43
(Constant)	1.33	0.24		5.64	0.00				
T5F7	0.09	0.11	0.10	0.85	0.40	0.01	1.94	0.73	0.40
(Constant)	0.14	0.02		5.94	0.00				
T5A1	0.05	0.01	0.46	4.41	0.00	0.22	0.19	19.46	0.00
(Constant)	0.14	0.02		6.36	0.00				
T5A2	0.05	0.01	0.49	4.72	0.00	0.24	0.19	22.30	0.00
(Constant)	0.36	0.03		13.04	0.00				
T5A3	-0.07	0.01	-0.55	-5.52	0.00	0.30	0.23	30.48	0.00
(Constant)	0.70	0.07		9.51	0.00				
T5A4	-0.13	0.03	-0.41	-3.74	0.00	0.16	0.61	13.99	0.00
(Constant)	0.76	0.04		20.50	0.00				
T5SC1	0.04	0.02	0.28	2.42	0.02	0.08	0.31	5.84	0.02
(Constant)	0.05	0.03		1.59	0.12				
T5SC2	0.01	0.01	0.08	0.68	0.50	0.01	0.24	0.46	0.50
(Constant)	5.82	0.16		37.11	0.00				
T5SC3	0.51	0.07	0.65	7.14	0.00	0.42	1.28	51.03	0.00
(Constant)	0.64	0.02		26.94	0.00				
T5LC1	-0.04	0.01	-0.37	-3.34	0.00	0.14	0.20	11.17	0.00
(Constant)	3.28	0.10		33.23	0.00				
T5LC2	0.20	0.05	0.47	4.45	0.00	0.22	0.81	19.77	0.00
(Constant)	0.42	0.02		23.00	0.00				
T5LC3	-0.03	0.01	-0.37	-3.34	0.00	0.14	0.15	11.17	0.00
(Constant)	0.27	0.01		20.93	0.00				
T5LC4	-0.02	0.01	-0.34	-3.04	0.00	0.12	0.11	9.23	0.00
(Constant)	0.10	0.01		9.63	0.00				
T5LC5	-0.01	0.00	-0.17	-1.45	0.15	0.03	0.09	2.12	0.15

Note. $n = 73$. T = Task. SEE = Standard Error of the Estimate. F (1, 71).

Appendix 7.1

Description of Each School

School	Place	Test administration period	PVKT	ST	ST facility	Order	Path after graduation (e.g., education continuance rate)
		Pilot study					
JHS X	Ibaraki	November, 2004	Class	After school	ST Kit	PVKT→ST	
SHS X	Tochigi	November, 2004	Class	--	--	--	University: 60.3%, Junior college: 4.9%, Vocational or others: 26.1%, Employed: 8.7%
University	Ibaraki	November, 2004	Varied	Varied	ST Kit	Both	
		Main experiment					
JHS A	Ibaraki	December, 2004	Class	Class	ST Kit	ST→PVKT	
JHS B	Ibaraki	November, 2004 to December, 2004	Class	Class	ST Kit	PVKT→ST	
JHS C	Ibaraki	December, 2004	Class	Class	ST Kit	PVKT→ST	
JHS D	Ibaraki	December, 2004	Class	After school	ST Kit	PVKT→ST	
JHS E	Ibaraki	December, 2004	Class	--	--	--	
JHS F	Tokyo	December, 2004	Class	--	--	--	
JHS G	Tokyo	November, 2004	Class	Class	ST Kit	Both	
SHS A	Ibaraki	December, 2004	Class	Class	LL	PVKT→ST	University: about 26.32%, Vocational: 57.89%, Employed: 15.79%
SHS B	Gifu	December, 2004	Class	Class	LL	Both	University: about 20%（Normally 12 to 13%）, Junior college: about 8%, Vocational or others: about 35%, Employed: about 20%（Normally 30%）, Others: about 17%
SHS C	Ibaraki	December, 2004	Class	Class	LL	PVKT→ST	90% of the students go to university right after or one year after graduating from senior high school.
SHS D	Ibaraki	December, 2004	Class	--	--	--	The data of students who graduated in March, 2004: University: 70%; Junior college: 1%; Vocational or others: 8%; Employed: 1%; Others (*Ronin*): 20%
SHS E	Ibaraki	December, 2004	Class OR After school	After school	ST Kit	PVKT→ST	University: 50%; Junior college or Vocational: 35%; Employed: 11%; *Ronin*: 4%

(Appendix 7.1 continues)

(Appendix 7.1 continued)

School	Place	Test administration period	PVKT	ST	ST facility	Order	Path after graduation (e.g., education continuance rate)
SHS F	Hiroshima	December, 2004	Class	--	--	--	University: 35 to 40%; Junior college: 20 to 25%; Vocational: 30%; Others: 10% (Employed: 2 to 3%)
SHS G	Tokyo	November, 2004	Class	Lunch OR After school	ST Kit	PVKT→ST	Higher education (mostly university): 50 to 60%; *Ronin*: 40 to 60 % (90 % decides on the school after one-year *Ronin*.); Junior college, vocational, or others: 0 to 2.5%
SHS H	Tokyo	February, 2005	Class	--	--	--	University: 60%; *Ronin*: 40%

Note. -- = The test was not administered. PVKT = Time when the Productive VKT was administered. ST = Time when the Speaking Test was administered. Class = The test was administered during class. After school = The test was administered after school. Lunch = The test was administered during lunch time. Varied = The test was administered on the one-on-one basis, so the time varied. ST facility = The type of Speaking Test facility used: the LL (Kit) or the ST (Speaking Test Kit) was used (see section 7.2.3.1 for details). Order = The order of the Productive VKT (PVKT) and the Speaking Test (ST): Both = The two tests were conducted in the order of PVKT→ST for some test takers, and in the order of ST→PVKT for others. The information on the path after graduation was derived from school homepages or from teachers, and the data was basically from students who graduated in March, 2005 unless the year was given. University = Four-year university. Vocational = Vocational technical school. Employed = A student who was employed. *Ronin* = A student who failed to be accepted at a university of his choice and was studying to take a test in another year.

Appendix 7.2

Characteristics of the participants: STEP Test grade (self-reported)

Grade	1.5	2	2.5	3	4	5	Grade total	Missing	
Frequency	1	48	166	274	82	20	591	322	913
Percent	0.11	5.26	18.18	30.01	8.98	2.19	64.73	35.27	100

Appendix 7.3

Characteristics of the participants: Experiences of going abroad (self-reported)

Abroad		Yes	No	Response total	Missing	Total
	Frequency	220	646	866	47	913
	Percent	24.1	70.76	94.85	5.15	100

Appendix 7.4

Characteristics of the participants: When students started learning English (self-reported)

	Before elementary school	During elementary school	After elementary school	Response total	Missing	Total
Frequency	35	228	612	875	38	913
Percent	3.83	24.97	67.03	95.84	4.16	100

Appendix 7.5

Purpose and design of the Size and Depth Section of the Productive Vocabulary Knowledge Test

TEST PURPOSE

Inferences: Trait definition of productive vocabulary (vocabulary knowledge independent of contexts of use). *Item level*: Knowledge of "a written form" and "form and meaning" ["a written form," "word parts," "associations," and "collocations"] of each from a sample of high-frequency words. *Test level*: The degree to which test takers have size [depth] of productive vocabulary knowledge.

Uses: *Research*: A measure of the degree of size [depth] of productive vocabulary knowledge for various kinds of L2 vocabulary research. *Instructional*: A classroom test intended to assist teachers in designing suitable vocabulary teaching plans for their students. A relatively low-stakes test for the students.

Intended Impacts: The test is hand-scored, which limits its application to a rather small scale. It provides L2 researchers with a tool designed to measure size [depth] of productive vocabulary knowledge.

TEST DESIGN

A discrete (i.e., an independent construct is measured; Read & Chapelle, p. 4), selective (i.e., specific target words are selected "as the focus of the assessment"; p. 5), context-independent (i.e., a response can be made "without referring to any context"; p. 4) vocabulary test with the words presented in isolation. A 96- [20-, 18-, 18-] item test. *Input*: Items are presented with a meaning in Japanese and the first word [with a stimulus word]. *Expected response*: Japanese test takers write the missing part of the form [write a derivational word, an antonym, and a collocate related to the stimulus word] that matches each of the target words. *Scoring*: Items are scored dichotomously.

Note. Based on the test specification and its examples in Read & Chapelle (2001, pp. 26-32). Specification for the Depth Section is shown in [ ].

Appendix 7.6

Productive Vocabulary Knowledge Test

Note. Answers were added after each item in [ ]. {} = Answers scored as incorrect. "Excluded: " = This item was excluded because there was a similar answer.

単語テスト

＜試験時間＞ 前半 25 分･後半 20 分 (合計 45 分)

＜注意事項＞

1 試験開始の合図があるまでは、この問題用紙の冊子を開かないでください。最初の問題が終わった後に、

ここから先は、指示があるまで進まないでください。

と書いてありますので、そこまで行ったら、見直しをするなどして待っていてください。

2 問題用紙に名前などを書いてください。

3 解答は直接問題用紙に書いてください。終了後、この問題用紙は回収します。

4 問題数は多いですが、あせらずによく考え、問題をといてください。

5 **わからない場合や正確なつづりが思い出せない場合も、一部分でもいいので、できるだけ書いてください。**

■ここに記入してください

＿＿＿＿＿＿＿＿学校 ＿＿年＿＿組＿＿＿番

名前＿＿＿＿＿＿＿＿＿＿＿＿ (男･女)

[1] 単語を書く問題 A
( ) 内に、日本語にあたる英単語を 1 語書いてください。はじめの文字は書いてあります。例.
町 (t) 答えは (town) または (towns)
(名詞のときは、単数·複数形のどちらでも O.K.)
ふれる、さわる (t) 答えは (touch)

1. 犬 (d) [dog(s)]
2. ネズミ (m) [mouse (mice)]
3. くつした (s) [sock(s) (sox)]
4. たぶん (m) [maybe]
5. 完全な、完ぺきな (p) [perfect (parfect, perfitt, poifect)]
6. 解釈 (i) [interpretation]
7. ～を信じている, 信じる (b) [believe]
8. 橋 (b) [bridge]
9. ～を洗う (w) [wash: Keyword]
10. 彼は、彼が (h) [he]
11. おば (a) [aunt]
12. 動物 (a) [animal: Keyword]
13. 運転する人、運転手 (d) [driver]
14. 独立、自立 (i) [independence]
15. 爆発する、～を爆発させる (e) [explode]
16. ～の後ろに (b) [behind]
17. 生きて、生きた状態で (a) [alive]
18. くだもの (f) [fruit(s): Keyword]
19. 集団、集まり (g) [group] {gathering}
20. ～を警告する、注意する (w) [warn] {watch}
21. 友達 (f) [friend(s) (frend(s)): Keyword]
22. 始まる、始める (s) start
23. 外国に、外国へ、海外に、海外へ (a) [abroad]
24. しきりに～したがっている (e) [eager] “Excluded: enthusiastic”
25. 通常、ふつうは (u) [usually]
26. ～を曲げる (b) [bend]
27. 暖かいこと、暖かさ (w) [warmth (warmness)]
28. 女性、婦人 (l) [lady]
29. 三 3 (t) [three: Keyword]
30. 招待 (i) [invitation]
31. 開いた、開いている (o) [open (opened)] {opening}
32. 劇場 (t) [theater (theatre)]
33. ～の価値がある、～に値する (d) [deserve]
34. 上に、頭上に (a) [above] {atop, ahead}
35. ～する余裕がある (a) [afford]
36. もう 1 つの, もう 1 人の (a) [another: Keyword]
37. ～を持ってくる (b) [bring]
38. 誕生日 (b) [birthday]
39. 管弦楽団 (o) [orchestra]
40. 王、国王 (k) [king]
41. 楽しみ、喜び (p) [pleasure]
42. 幸福に、楽しく (h) [happily]
43. 男, やつ (g) [guy]
44. 非常に広い、広大な (v) [vast]
45. 打ちとけた、形式ばらない (i) [informal]

46. じっと見つめる、〜をじろじろ見る (s) [stare]
47. 内容、中身 (c) [content] {component}
48. 悲劇 (t) [tragedy]
49. 機会、好機 (o) [opportunity] “Excluded: occasion]
50. 起源 (o) [origin]
51. 支持者、味方 (s) [supporter]
52. 〜を必要とする (r) [require]
53. 〜に食べ物･えさを与える (f) [feed]
54. 所有すること、入手すること (p) [possession] {property}
55. どんな〜でも (w) [whatever] “Excluded: whichever”
56. 客 (g) [guest]
57. 威厳 (d) [dignity]
58. 〜に乗る (r) [ride (ride on): Keyword]
59. 卵 (e) [egg(s)]
60. はう、腹ばいで進む (c) [crawl] “Excluded: creep”
61. 尊敬、敬意 (r) [respect(s)] “Excluded: reverence”
62. リンゴ (a) [apple(s)]
63. 財宝、宝物 (t) [treasure]
64. 〜を含む (i) [include] “Excluded: involve”
65. 生徒, 児童 (p) [pupil]
66. タマネギ (o) [onion]
67. 〜を楽しむ (e) [enjoy] {entertain}
68. 湖 (l) [lake]
69. ゆうれい (g) [ghost]
70. 日付 (d) [date]
71. 労働、労力 (l) [labor (labour)]
72. それにもかかわらず (r) [regardless]
73. 半分の (h) [half (hafe, haff, harf, haulf, hahf, halfe, haa(l)f, ha'f, hofe); Keyword]
74. 輸出 (e) [export] “Excluded: exportation]
75. 糸 (t) [thread]
76. 〜を買う (b) [buy: Keyword]
77. 〜を知らせる、通知する (i) [inform]
78. 優勝者 (c) [champion (champ, champeen, champean)]
79. 〜を支払う (p) [pay]
80. 軍人、兵隊 (s) [soldier]
81. 〜を破滅させる (r) [ruin]
82. 〜を料理する (c) [cook: Keyword]
83. 講演、講義 (l) [lecture] {lesson}
84. 刑務所 (j) jail
85. 〜を予期する、期待する (e) [expect] {estimate}
86. 塔 (t) [tower]
87. 家族 (f) [family (femily): Keyword]
88. 働く、仕事をする (w) [work (wark)]
89. 〜を破壊する (d) [destroy] “Excluded: damage, demolish”
90. 運河、人工水路 (c) [canal] {channel}
91. 島 (i) [island (iland, i(s)lant, ile-dand, i(s)le)]
92. 宗教 (r) [religion]
93. 類似して、同様に (s) [similarly]
94. それの、その (i) [its]
95. ぬれた、湿った (w) [wet]
96. 頼りになる、信頼できる (r) [reliable] {respectable} “Excluded: responsible”

97. 市長 (m) [mayor]
98. ちり、ほこり (d) [dust] “Excluded: dirt”
99. より小さい、少数派の (m) [minor] {micro}
100. 芸術, 美術 (a) [art]
101. どうぞ (p) [please]
102. ～を書く (w) [write: Keyword]
103. だれの (w) [whose]
104. 恐怖、大あわて (p) [panic] “Excluded: phobia, panic”
105. 作る人、製作者 (m) [maker] “Excluded: manufacturer”
106. 意識、正気 (c) [consciousness]
107. がん (c) [cancer]
108. 波 (w) [wave]

ここから先は、指示があるまで進まないでください。

ここから前には戻らないでください。

---

[2] 単語を書く問題 B
英単語を、[ ] の中に書かれた品詞にしてください。( ) 内には 1 語書いてください。ing と ed をつけた語は書かないでください。
例. kind [名詞：～なこと] 答えは (kindness)
＊「名詞」とは、主に、人や物の名前 (例：ネコ) を表す言葉です。
例：This is my cat. →cat が名詞
introduction [動詞：～する] 答えは (introduce)
＊「動詞」とは、主に、動き (例：走る) や状態 (例：住んでいる) を表す言葉です。
例：I live in Japan. →live が動詞
wonder [形容詞：～ (の状態) の] 答えは (wonderful)
＊「形容詞」とは、人・物・出来事などについて述べる言葉 (例：かわいい) です。
例：I have a cute cat. →cute が形容詞

---

1. work [名詞：～する人] ( ) [worker (workman, warker)] {workness}
2. art [名詞：～をする人] ( ) [artist]
3. destroy [名詞：～すること] ( ) [destruction] {destroyer}
4. criticize [名詞：～したもの] ( ) [criticism, (critism, critique)]
5. inform [名詞：～したもの] ( ) [information] {informer, informant}
6. comfortable [名詞：～なこと] ( ) [comfort, comfortableness] {comforter}
7. successful [名詞：～した状態のこと] ( ) [success, successfulness]
8. able [名詞：～こと] ( ) [ability]
9. invitation [動詞：～する] ( ) [invite]
10. supporter [動詞：～する] ( ) [support(s)]
11. argument [動詞：～する] ( ) [argue, argufy, argyfy]
12. emphasis [動詞：～する] ( ) [emphasize (emphasise)]
13. independence [形容詞：～ (の状態) の] ( ) [independent]
14. religion [形容詞：～ (の状態) の] ( ) [religious]
15. similarly [形容詞：～ (の状態) の] ( ) [similar]
16. origin [形容詞：～ (の状態) の] ( ) [original]
17. warmth [形容詞：～ (の状態) の] ( ) [warm]
18. specialist [形容詞：～ (の状態) の] ( ) [special, specialized,

specialized]
19. dirt [形容詞：～ (の状態) の] ( ) [dirty]
20. end [形容詞：～のない] ( ) [endless]

---

[3] 単語を書く問題 C
( ) の中に入る、反対(はんたい)の意味の英単語を 1 語書いてください。
例. good ( ) 答えの例は (bad)

---

1. open ( )
Example answers: closed, close, shut, covered, enclosed, sealed, fastened, locked, folded, restricted, bounded, protected, narrow, obstructed, crowded, cluttered, unavailable, inaccessible, introverted, reticent, reserved, prejudiced, biased, bigoted, subjective, unfair, unjust, stubborn, obdurate, lock, fasten, bar, seal, fold, block, obstruct, end, conclude, finish, terminate

2. high ( )
Example answers: low, short, stunted, dwarfed, moderate, mild, average, reasonable, routine, reduced, subdued, suppressed, restrained, low-ranking, lowly, unimportant, inconsequential, insignificant, undistinguished, secondary, common, menial, debased, degraded, ignoble, low-pitched, base, alto, deep, husky, gruff, hoarse, sad, cheerless, gloomy, joyless, depressed, dejected, melancholy, angry, mad, irritable, short, low

3. wet ( )
Example answers: dry, bone-dry, parched, dried, set, hardened, nonliquid, solid, dryness, parch, desiccated, dehydrate, evaporate, clear, sunny

4. alive ( )
Example answers: dead, deceased, expired, defunct, lifeless, departed, inanimate, unanimated, dispirited, spiritless, apathetic, inactive, unaware, extinct, inactive, inoperative, inoperable, gone, lost

5. independence ( )
Example answers: dependence, subordination, subjection, servitude, slavery, bondage, reliance, dependency

6. start ( )
Example answers: delay, stall, end, finish, terminate, stop, cease, finale, termination, windup, disadvantage, handicap

7. nonsense ( )
Example answers: sense, wisdom, fact, reality, truth, gravity, seriousness, reason

8. happily ( )
Example answers: sadly, unhappily, sorrily, sorrowfully, despondently, forlornly, miserably, gloomily glumly, melancholily, downcastly, joylessly, mournfully, somberly, unfortunately, unluckily, lucklessly, inauspiciously, unfittingly, unseasonably

9. like ( )
Example answers: unlike, dissimilar, different, diverse, divergent, dislike, hate, abhor, detest, loathe, abominate (Excluded for the analysis)

10. tragedy ( )
Example answers: blessing, boon, kindness, happiness, joy, pleasure, satisfaction, gratification, bliss, contentment, comedy

11. supporter ( )
Example answers: adversary, antagonist, opponent

12. able ( )
Example answers: unskillful, incapable, incompetent, inexpert, inept, inefficient, ineffective, amateurish, mediocre, indifferent, fair, unfit, inadequate, unqualified

13. export ( )
Example answers: import

14. buy ( )
Example answers: sell, vend, retail, auction, hawk, rent, lease

15. minor ( )
Example answers: major, greater, main, important, significant, adult, elder, grown-up

16. beautiful ( )
Example answers: ugly, unattractive, bad-looking, hideous, grotesque, unpleasant, bad, awful, disgusting, repulsive, repugnant, revolting, terrible, lousy, second-rate

17. quick ( )
Example answers: slow, long, lingering, gradual, deliberate, sluggish, lazy, lethargic, heavy, inert, inactive, dull, wearisome, unintelligent, stupid, unresponsive, slow-witted, inexpert, unskillfull, maladroit, calm, patient, restrained, temperate

18. hungry ( )
Example answers: sated, satisfied, full, fed

---

[4] 単語を書く問題 D
( ) の中に入る、適(てき)した英単語 (形容詞(けいようし)か名詞(めいし)) を 1 語書いてください。good, better, best, bad, nice, wonderful, great, big, small, old は書かないでください。同じ語が入ってもかまいません。
例. 形容詞が入る例：( ) air 答えの例は (clean) air
名詞が入る例：( ) cream 答えの例は (ice) cream

---

1. ( ) dog
Example answers: strong, stray, sleeping, guard , absolute, dirty, hot, top, watch

2. ( ) cancer
Example answers: lung, breast, stomach, skin

3. ( ) art
Example answers: modern, Indian, contemporary, fine, performance

4. ( ) birthday happy
Example answers: 74th, happy, last, next

5. ( ) tragedy great
Example answers: Shakespeare's, Greek, absolute, awful

6. ( ) religion
Example answers: different, famous, Islamic, world

[5]　単語を書く問題 E
(　　) の中に入る、適(てき)した英単語 (名詞(めいし))を 1 語書いてください。同じ語が入ってもかまいません。
例.　clean (　　)　答えの例は clean (air)

1.　perfect (　　)
Example answers: condition, teeth, performance, world, example

2.　wet (　　)
Example answers: towel, grass, weather, summer, hair

3.　minor (　　)
Example answers: league, change, error, road, role

4.　eager (　　) student
Example answers: student, beaver, crowds, volunteers

[6]　単語を書く問題 F
(　　) の中に入る、適(てき)した英単語 (名詞(めいし))を 1 語書いてください。同じ語が入ってもかまいません。
例. play (　　)　答えの例：play (baseball)
　注：(a/an/the) は、a, an, the がついても、つかなくてもいいという意味

1.　enjoy (a/an/the) (　　) tennis
Example answers: job, time, health, book

2.　buy (a/an/the) (　　) candy
Example answers: dress, drink, equipment, car

3.　destroy (a/an/the) (　　)　town
Example answers: environment, hope, faith, building

4.　wash (a/an/the) (　　)　dishes
Example answers: dishes, hair, hand, mouth, car

5.　cook (a/an/the) (　　)　food
Example answers: meal, dinner, breakfast, rice

6.　pay (a/an/the) (　　)
Example answers: cash, kids,　$5, bill, tax, attention

7.　ride (a/an/the) (　　)　bike
Example answers: bicycle, bus, elevator, wave, horse

8.　write (a/an/the) (　　)　letter
Example answers: books, ‘Harry Potter’, letters, music, programs, home

これで終わりです。ご協力ありがとうございました。

Appendix 7.7

Common words between the Size Section and the Depth Section

Size & Derivation	Size & Antonym	Size & Collocation
work inform invitation similarly origin warmth	open high alive start happily export	dog cancer birthday perfect eager enjoy wash cook pay ride write
supporter independence	supporter independence	
art destroy religion		art destroy religion
	wet minor tragedy buy	wet minor tragedy buy

Appendix 7.8

Words in the Derivation Subsection (20 items)

suffix	Level	Word to add suffix to	Answer	Word to take suffix from	Answer
-able	3			comfortable	comfort
-al	4	origin	original		
-ally	2				
-ance	2				
-ant	2				
-ation	4	inform	information	invitation	invite
-ction	2	destroy	destruction		
-er	3	work	worker	supporter	support
-ess	4				
-ful	4	wonder (例)	wonderful	successful	success
-ic	2				
-ical	2				
-ing	2				
-ion	2				
-ish	3				
-ism	4	criticize	criticism		
-ist	4	art	artist	specialist	special
-ition	2				
-ity	4	able	ability		
-ize	4	emphasis	emphasize		
-less	3	end	endless		
-ly	3			similarly	similar
-ment	4			argument	argue
-ness	3	kind (Example)	kindness		
-sive	2				
-th	3			warmth	warm
-tion	2			introduction (Example)	introduce
-ous	4	religion	religious		
-t	2	independence	independent		
-ure	2				
-y	3	dirt	dirty		

Note. The list is from Bauer and Nation (1993).

Appendix 7.9
Derivational prefix in the Antonym Subsection

Suffix	Level	Item	Answer (add)	Item	Answer (delete)
in-	4	export	import	independence	dependence
non-	3			nonsense	sense
un-	3	able	unable		
		beautiful	unbeautiful		
		happily	unhappily		
		like [deleted]	unlike		

Appendix 7.10
Speaking Test booklet (Version A)

スピーキングテスト

＊お願い：合図があるまでは、このテスト冊子を開かないでください。この冊子には、何も書き込まないでください。

<注意事項>

1 テスト中、機械の操作は先生が行います。機械にさわらないでください。

2 話すのは全て英語です。大きめの声で、マイクに向かって話してください。

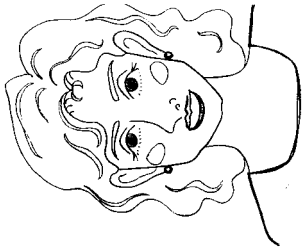
3 時間制限があるときもありますが、時間通りに話が終わらなくても、かまいません。できるだけ、文の形で、たくさん話してください。

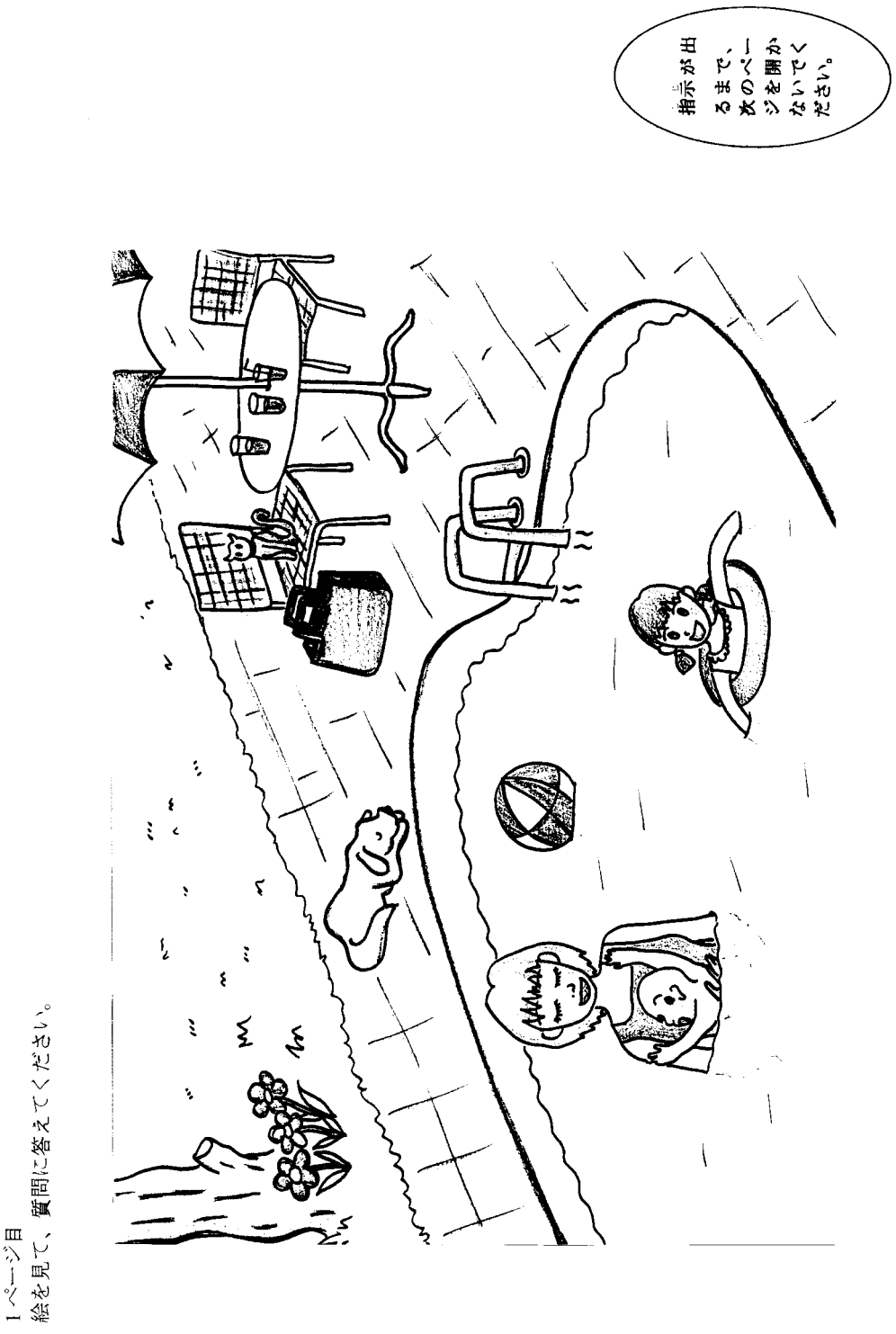
4 人によって問題の順番が異なっていますので、他の人は気にせず話してください。

5 テストのページ数があっているか、毎回確認してください。

<テストの設定>

皆さんの学校に、新しい外国人の先生 (Smith 先生) がやってきました。Smith 先生はあなたの英語の力を知りたいので、テストをすることになりました。Smith 先生と話しているつもりで、英語をたくさん話してください。考える時間はありませんので、Smith 先生が質問したり、Please start.と言ったりしたら、すぐに話し始めてください。





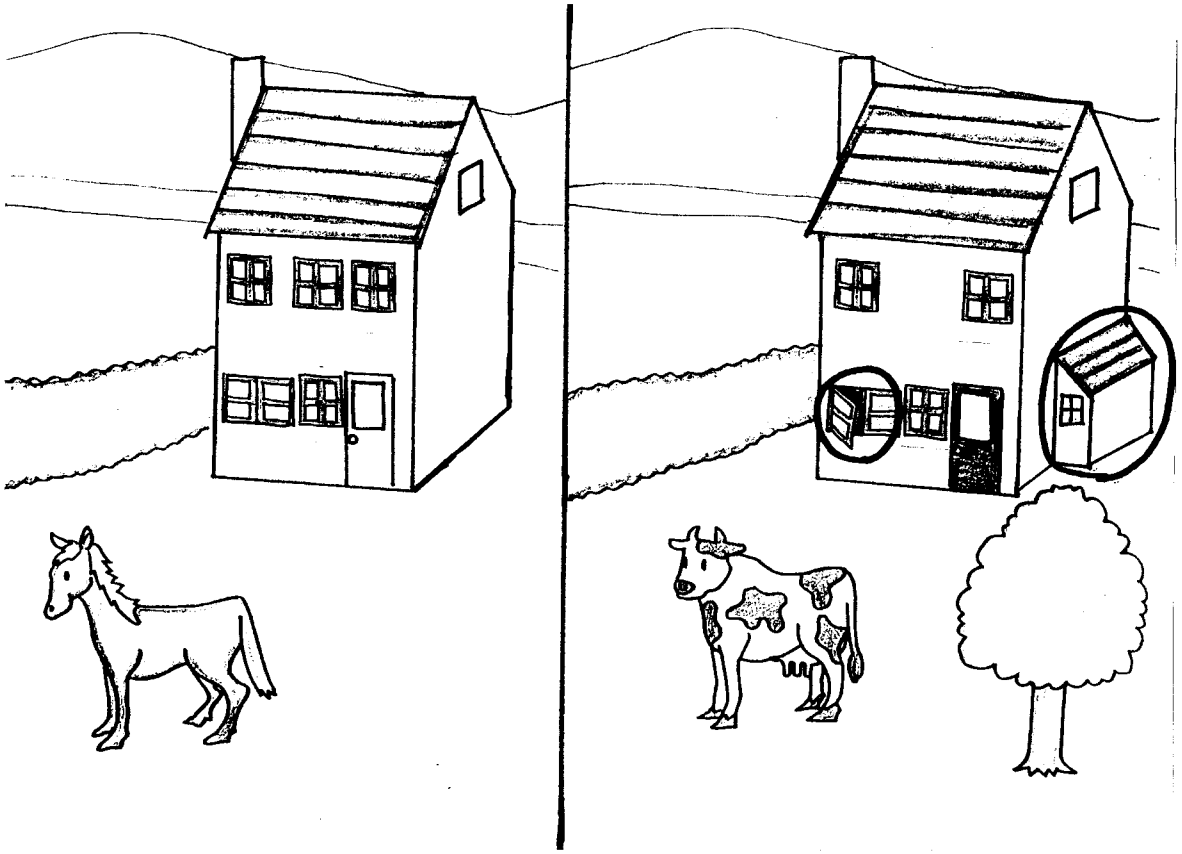
Note. Drawn based on University of Cambridge ESOL Examinations (n.d.-b, p. 62).

Task 1

2ページ目
これから先生に、あなたの自己紹介をしてください。自己紹介では最初に、**名前・家族・友達**のことを話してください。わからなかったら、他のこと (学校のこと、好きなことなど) をいろいろ話してください。話す時間は 60 秒間あります。Please start と Smith 先生が言ったら、必ず Yes と答えて、すぐに話し始めてください。

3 ページ目

2 枚の絵には違いがいくつかあります。違っている所を説明してください。印のついた物については最初に言ってください。わからなかったら飛ばして、他のことを話してもかまいません。話す時間は 60 秒間あります。Please start. と Smith 先生が言ったら、必ず Yes. と答えて、すぐに話し始めてください。



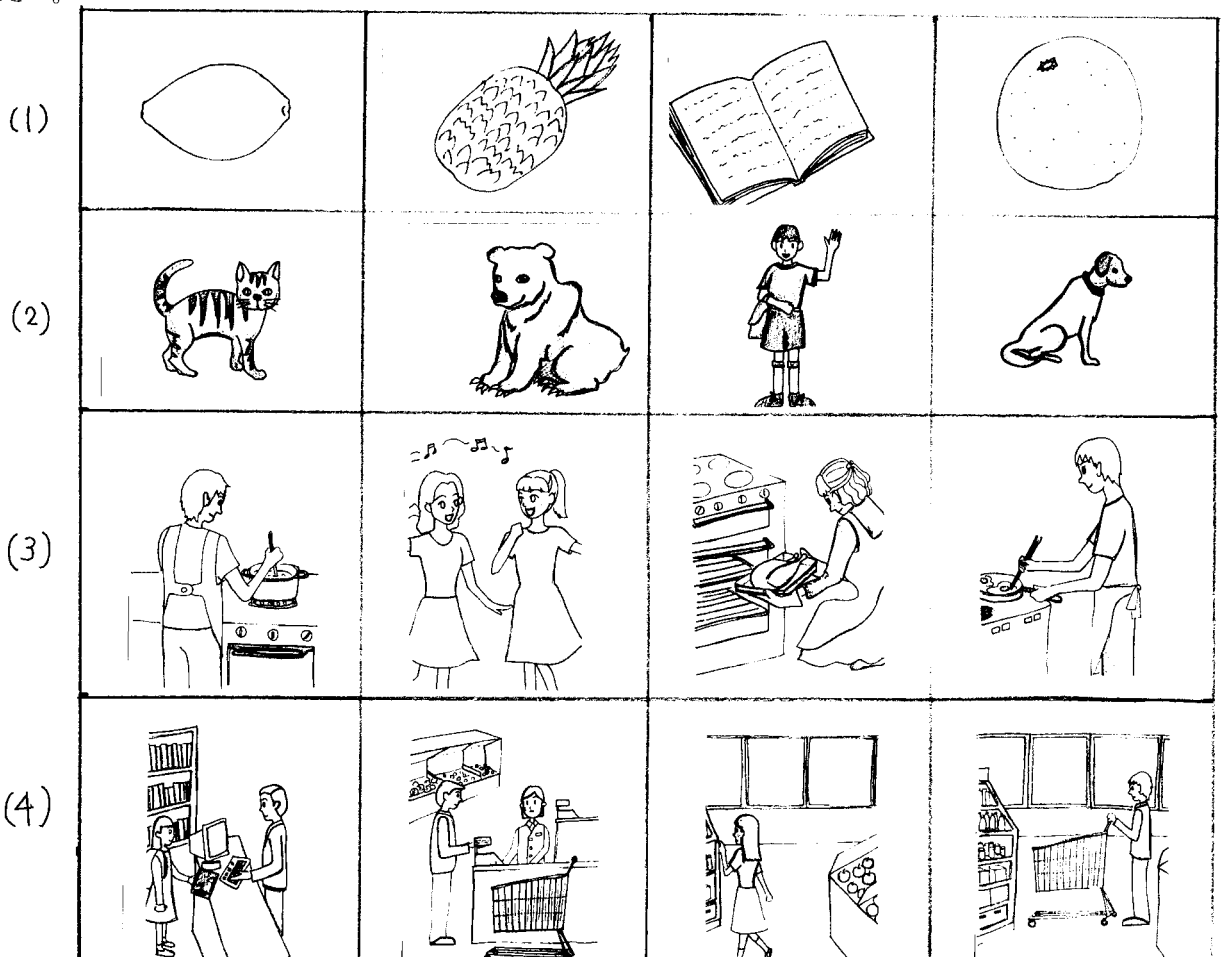
指示が出るまで、次のページを開かないでください。

Task 3

4ページ目
絵を見ていない先生にもどんな写真かわかるように、この写真に書いてあることをできるだけ詳しく言ってください。印のついた動作は最初に言ってください。わからなかったら飛ばして、他のことを話してもかまいません。話す時間は60秒間あります。Please start. と Smith 先生が言ったら、必ず Yes. と答えて、すぐに話し始めてください。
指示が出るまで、次のページを開かないでください。

5 ページ目

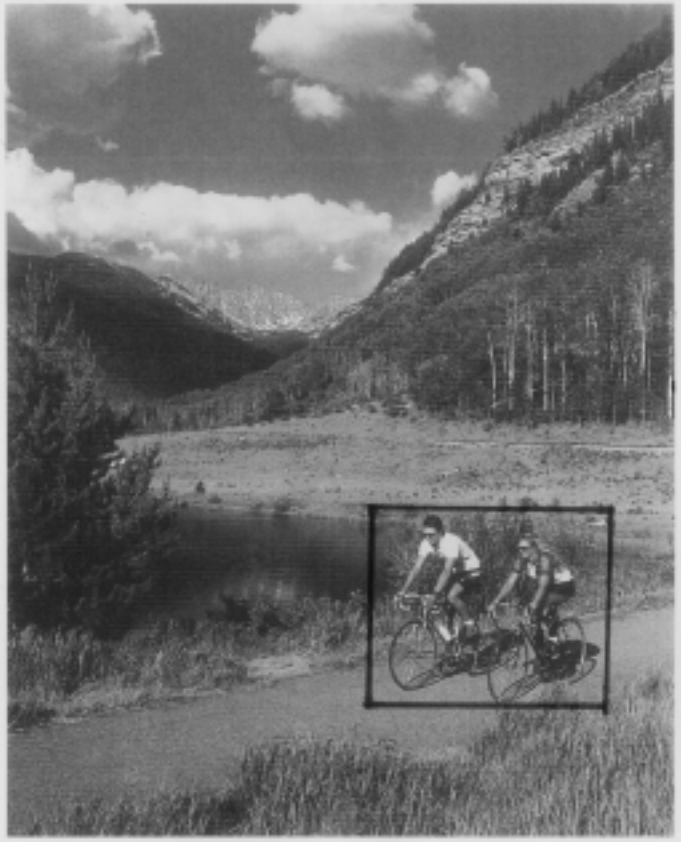
4 枚の絵の中には、1 枚違う種類のものが入っています。どれが違う種類か、その理由について話してください。わからなかったら飛ばして、パスと言って、次に進んでもかまいません。話す時間は 60 秒間あります。Please start.と Smith 先生が言ったら、必ず Yes.と答えて、すぐに話し始めてください。



指示が出るまで、次のページを開かないでください。

6 ページ目

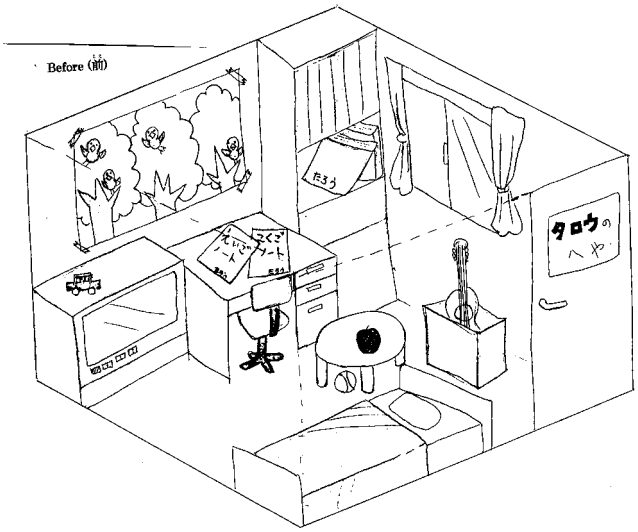
絵を見ていない先生にもどんな写真かわかるように、この写真に書いてあることをできるだけ詳しく言ってください。印のついた動作は最初に言ってください。わからなかったら飛ばして、他のことを話してもかまいません。話す時間は 60 秒間あります。Please start. と Smith 先生が言ったら、必ず Yes. と答えて、すぐに話し始めてください。

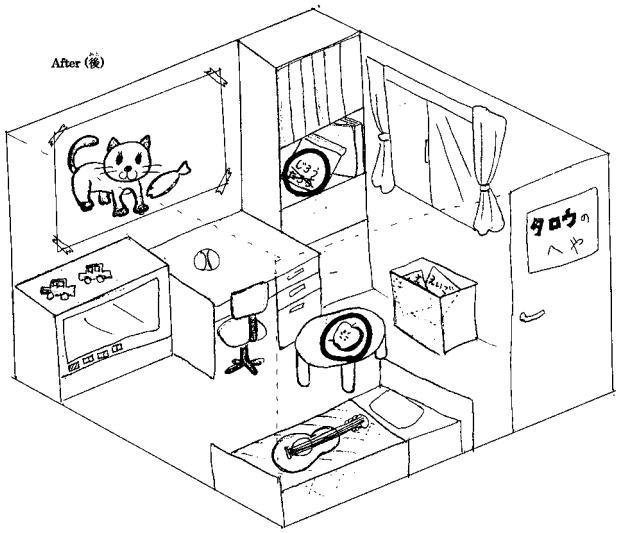


指示が出るまで、次のページを開かないでください。

Task 5

7ページ目
たたんである絵を開いてください。上と下に部屋の絵があります。
あなたの弟はとてもいたずら好きです。あなたが学校に行っている間に、あなたの部屋をいじってしまいました。「前はどうだったが、今はこうだ」というように、どこがどう違っているのかを言ってください。印のついた物は最初に言ってください。わからなかったら飛ばして、他のことを話してもかまいません。話す時間は60秒間あります。Please start.と Smith 先生が言ったら、必ず Yes.と答えて、すぐに話し始めてください。





指示が出るまで、次のページを開かないでください。

Appendix 7.11
Productive Vocabulary Knowledge Test Questionnaire

単語テストについてのアンケート

あまり深く考えずにお答えください。また、他の人のアンケートの答えは見ないで答えてください。

Part 1　[　　] に記入してください。 (同じことに以前答えた場合、答えなくて大丈夫です)

(1-1) いつから英語を勉強しはじめましたか。

[1. 小学校に入る前から　2. 小学生の時から　3. 中学校 1 年から]　　(1-1) [　　]

「1. 小学校に入る前から　2. 小学生の時から」と答えた人は、具体的にいつからですか？

[　　歳の時から、小学校　　年生から]

(1-2) 海外へ行ったことはありますか？あてはまる方に○をつけてください。

(1-2) [1. ある　2. ない]

「1. ある」と答えた人は、どこの国へ、どのくらいの期間行きましたか？

(何度も行った人は、全部書いてください)

[どこの国へ　　、どのくらいの期間　年　　ヶ月]

(1-3) 何か英語の資格を持っていますか？持っている人は、以下に書いてください。

(複数回答可)

[英検　　級、ケンブリッジ英検　　級、その他　　]

Part 2　あなたの気持ちに一番近いものを、下の 0〜5 から選んで、[　　] に書いてください。

0. わからない	1. 全然そうでない	2. ほとんどそうでない	3. どちらともいえない	4. 少しそうだ	5. とてもそうだ

(2-1) このテストで、あなたはよくできたと思いますか？　(2-1) [　　]

(2-2) このテストは英語の単語力をみていると思いますか？　(2-2) [　　]

(2-3) テストのやり方はよく分かりましたか？　(2-3) [　　]

(2-4) このテストを受けて、今後もっと英語の勉強をしようと思いましたか。　(2-4) [　　]

(2-5) テスト中、たくさん書こうと思いましたか？　(2-5) [　　]

(2-6) テスト中、不安だったりきんちょうしたりしましたか？　(2-6) [　　]

Part 3.　以下の質問に答えてください。

(3-1) テスト中に、何か困ったことはありましたか？

_____

_____

(3-2) テストを受けてみて、何か感想はありますか？

_____

_____

テストとアンケートに答えていただき、ありがとうございました。

Appendix 7.12
Speaking Test Questionnaire

スピーキングテストについてのアンケート　　名前＿＿＿＿＿＿

あまり深く考えずにお答えください。また、他の人のアンケートは見ないで答えてください。

Part 1　[　　] に記入してください。

(1-1) いつから英語を勉強しはじめましたか。
[1. 小学校に入る前から　2. 小学生の時から　3. 中学校 1 年から]　　(1-1) [　　]
(1 または 2 のとき、具体的にいつから？　　歳の時から、小学校　　年生から]

(1-2) 旅行も含めて、海外へ行ったことはありますか？あてはまる方に○をつけてください。
(1-2) [1. ある　2. ない]

「1. ある」と答えた人は、どこの国へ、どのくらいの期間行きましたか？
(何度も行った人は、全部書いてください)
[どこの国へ　　、どのくらいの期間　　年　　ヶ月]

(1-3) 何か英語の資格を持っていますか？持っている人は、以下に書いてください。
(複数回答可)
[英検　　級、ケンブリッジ英検　　級、その他　　]

Part 2　あなたの気持ちに一番近いものを、下の 0〜5 から選んで、[　　] に書いてください。

0. わから ない	1. 全然 そうでない	2. ほとんど そうでない	3. どちらとも いえない	4. 少し そうだ	5. とても そうだ

例：テレビゲームをよくしますか？　　→「全然しない」とき　　例 [　1　]

(2-1) 家族や友達と (日本語で) よく話しますか？　　(2-1) [　　]
(2-2) 授業中・授業外を含めて、英語をたくさん話したことはありますか？　　(2-2) [　　]
(2-3) 学校内、学校外を含めて、今まで英語を話すテストをたくさん受けたことがありますか？
(2-3) [　　]
(2-4) このテストで、あなたはよくできたと思いますか？　　(2-4) [　　]
(2-5) このテストは、英語の話す力をみていると思いますか？　　(2-5) [　　]
(2-6) テープの音は、よく聞こえましたか？　　(2-6) [　　]
(2-7) テストのやり方はよく分かりましたか？　　(2-7) [　　]
(2-8) 声を録音する方法は簡単でしたか？　　(2-8) [　　]
(2-9) テスト中、不安だったりきんちょうしたりしましたか？　　(2-9) [　　]
(2-10) テスト中、まわりの人や音が気になりましたか？　　(2-10) [　　]
(2-11) テスト中、たくさん話そうと思いましたか？　　(2-11) [　　]
(2-12) テスト中、今まで勉強した文法をたくさん使おうと思いましたか？　　(2-12) [　　]
(2-13) テスト中、今まで勉強した単語をたくさん使おうと思いましたか？　　(2-13) [　　]
(2-14) テスト中、あまり間違えないように話そうと思いましたか？　　(2-14) [　　]
(2-15) テスト中、すらすらと話そうと思いましたか？　　(2-15) [　　]
(2-16) テスト中、暗記した文をたくさん言おうと思いましたか？　　(2-16) [　　]
(2-17) このテストを受けて、今後もっと英語の勉強をしようと思いましたか。　　(2-17) [　　]
(2-18) このテストで出てきた絵を見たことがありましたか？　　(2-18) [　　]
(1 以外のとき、どの絵？　　)
(2-19) 先生と 1 対 1 で向かい合ってテストを受けたら、何か違いがあったと思いますか？
(2-19) [　　] (4 または 5 のとき、違いは何？＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿)

Part 3.　以下の質問に答えてください。

(3-1) このテストのやり方で何か困ったことがあったら、具体的に書いてください。(例：テープが聞こえない)
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

(3-2) このテストを受けているとき、難しかったのは何ですか？ (発音がわからないなど)
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

(3-3) このテスト中、言いたかったけれど、英語で言えなかった表現があったら、具体的に書いてください。
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

(3-4) 2 つのテストで共通していた単語を覚えていたら、具体的に書いてください。
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

(3-5) テストを受けてみて、何か感想はありますか？
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

テストとアンケートに答えていただき、ありがとうございました。テストの内容は友達に話さないでください。

Appendix 7.13
Instructions on the tape in the Speaking Test (Tape-Mediated AB Version)
Instructions in the tape (15 minutes per student) [Originally in Japanese]

Now we are going to begin the Speaking Test. Is the sound volume all right? If it is too loud or quiet, please tell the teacher. I am afraid that in some headphones, you can only hear from the left or right. Please put on the headphones and take the test. If you do not understand, call the teacher. Then read the instructions and test settings on the top page.

[Instructions on the cover page:
1. Do not touch the machines (e.g., tape cassette recorders). Only the teacher can touch them.
2. All you speak is English. Please speak into the microphone.
3. You do not need to finish your talk within the time even when there are time limits. Please speak English in sentences as much as you can.
4. The order of the tasks is different for each person. Do not be concerned about what others are doing.
5. Please make sure you turn to the correct page each time.

Test setting:
An assistant language teacher (Ms. Smith) is a new teacher. She wants to know your English ability, so you will have an English speaking test. Speak as if you were speaking to her. You do not have time to prepare. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.]

(10 seconds)

Now please speak English as if you were speaking to Ms. Smith. Do not cover the microphones with the test booklet. Now let's begin.

Ms. Smith (Teacher: T): Hello. My name is Mary Smith. May I have your name, please?
(5 second)
Expected responses from the student (S): Yes. My name is Taro Tsukuba.

T: Nice to meet you.
(5 second)
S: Nice to meet you, too.

T: Do you like English?
(3 second)
S: Yes./No.

T: I see. Thank you.

Please open the next page. This is page 1.
Please answer the questions.

(Warm-up: Pool picture: starters)

T: Where is the mother?
(6 seconds)
S: She's in the pool/water.

T: What color is the bag?

(6 seconds)
S: It's pink.

T: How many flowers are there near the tree?
(6 seconds)
S: There are three.

T: I see. Thank you.

Please open the next page. This is page 2.
Task 1
Please introduce yourself to Ms. Smith. Please talk about your name, family, and friends first. If you do not know what to say, please talk about anything (e.g., your school and favorite things). You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: Yes. My name is Kayoko Ogawa. My hobby is listening to music. There are four people in my family. …

(60 seconds)

T: I see. Thank you.

Please open the next page. This is page 3.
Task 2
There are differences between the two pictures. Please explain the differences. Please talk about the marked objects first. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: In the left picture, there is a horse. In the right picture, there is a cow. …

(60 seconds)

T: I see. Thank you.

Please open the next page. This is page 4.
Task 3
Describe this picture in as much detail as possible so that Ms. Smith, who is not looking at the picture, can understand what is in it. Please talk about the marked behaviors first. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: A mother and her daughter are washing dishes. They look happy. …

(60 seconds)

T: I see. Thank you.

Please open the next page. This is page 5.
Task 6
One picture out of the four is different. Please tell me what is different and why. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: Number one. A book is different because it's not a fruit. …

(60 seconds)

T: I see. Thank you.

Please open the next page. This is page 6.
Task 4
Describe this picture in as much detail as possible so that Ms. Smith, who is not looking at the picture, can understand what is in it. Please talk about the marked behaviors first. If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: A man and a woman are riding a bike. The sky is blue and beautiful.

(60 seconds)

T: I see. Thank you.

Please open the next page. This is page 7.
Task 5
Please unfold the folded picture. There are pictures above and below. Your brother (Jiro) is mischievous. While you were away at school, he scattered your belongings in your room. Say how and what in the room has changed by saying "something was something before, but now something is something." If you do not know what to say, please talk about anything. You have 60 seconds to speak. When Ms. Smith says "Please start" please say "Yes" and start speaking right away.

Now let's start.
T: Please start.

S: There was an apple on the table but there is half of the apple. … The ball was on the desk but now it is under the table..…

(60 seconds)

T: I see. Thank you.
This is the end of the test. Please tell the teacher you have finished.

Appendix 8.1
Instructions of the oral version of the Productive Vocabulary Knowledge Test

Instructions
1　Do not open this booklet until you hear the teacher say "Please start."
2　This test takes 10 minutes. There are many questions, so some people may not finish all. Please do not worry about not finishing the test. Think carefully without rushing, and answer the questions.
3　Please say as much as you can. If you do not know the answer or cannot remember the correct pronunciation, please say whatever you can remember.
4　Please give your response in a clear voice.
5　You are going to take a different test from other students, so do not be concerned about what others are doing.

6　Test items you are going to solve

Respond orally the English word that best corresponds to the Japanese meaning on your answer sheet. The first letter of the English word is already given.
Example: 町 (t_____)　The answer is *town* (*towns*).
ふれる、さわる (t　　)　The answer is (touch).

Both singular and plural forms are correct answers.

7　Answering method
Please say two things for each item.
(1) Item number: Say in Japanese. (because of time constraint)
(2) Answer: Say in English (with the best English pronunciation possible)
Example: 200 番, town; 201 番, touch; 203 番 …

8　When you do not know the answer, please say "Pass" and go on to the next item. Please do not go back to the previous items even when you have skipped items or made mistakes.

9　Please tell the teacher you have finished by raising your hand if you finish within 10 minutes.

Appendix 8.2
Oral version of the Productive Vocabulary Knowledge Test: The First Version starting with Item 1

単語を言う問題
日本語にあたる英単語を 1 語言ってください。始まりの文字は書いてあります。
例　町 (t_____)　答えは (town) または (towns)、
(名詞のときは、単数・複数形のどちらでも O.K.)
ふれる、さわる (t_____)　答えは (touch)

解答の方法
問題ごとに、2 つのことを言ってください。
(1) 問題番号：日本語で言う (時間の関係上)
(2) 答え：英語で言う (できるだけ英語らしい発音で)
例　200 番　town、201 番　touch、203 番　flower

注意：
わからなかったら、「パス」と言って次に行く
前には戻らない

1. 犬　(d_____)
2. ネズミ　(m_____)
3. くつした　(s_____)

4. たぶん (m_____)
5. 完全な、完ぺきな (p_____)
6. 解釈 (i_____)
7. ～を信じている, 信じる (b_____)
8. 橋 (b_____)
9. ～を洗う (w_____)
10. 彼は、彼が (h_____)
11. おば (a_____)
12. 動物 (a_____)
13. 運転する人、運転手 (d_____)
14. 独立、自立 (i_____)
15. 爆発する、～を爆発させる (e_____)
16. ～の後ろに (b_____)
17. 生きて、生きた状態で (a_____)
18. くだもの (f_____)
19. 集団、集まり (g_____)
20. ～を警告する、注意する (w_____)
21. 友達 (f_____)
22. 始まる、始める (s_____)
23. 外国に、外国へ、海外に、海外へ (a_____)
24. しきりに～したがっている (e_____)
25. 通常、ふつうは (u_____)
26. ～を曲げる (b_____)
27. 暖かいこと、暖かさ (w_____)
28. 女性、婦人 (l_____)
29. 3、三 (t_____)
30. 招待 (i_____)
31. 開いた、開いている (o_____)
32. 劇場 (t_____)
33. ～の価値がある、～に値する (d_____)
34. 上に、頭上に (a_____)
35. ～する余裕がある (a_____)
36. もう 1 つの, もう 1 人の (a_____)
37. ～を持ってくる (b_____)
38. 誕生日 (b_____)
39. 管弦楽団 (o_____)
40. 王、国王 (k_____)
41. 楽しみ、喜び (p_____)
42. 幸福に、楽しく (h_____)
43. 男, やつ (g_____)
44. 非常に広い、広大な (v_____)
45. 打ちとけた、形式ばらない (i_____)
46. じっと見つめる、～をじろじろ見る (s_____)
47. 内容、中身 (c_____)
48. 悲劇 (t_____)
49. 機会、好機 (o_____)
50. 起源 (o_____)
51. 支持者、味方 (s_____)
52. ～を必要とする (r_____)
53. ～に食べ物･えさを与える (f_____)
54. 所有すること、入手すること (p_____)
55. どんな～でも (w_____)
56. 客 (g_____)

57. 威厳 (d_____)
58. ～に乗る (r_____)
59. 卵 (e_____)
60. はう、腹ばいで進む (c_____)
61. 尊敬、敬意 (r_____)
62. リンゴ (a_____)
63. 財宝、宝物 (t_____)
64. ～を含む (i_____)
65. 生徒, 児童 (p_____)
66. タマネギ (o_____)
67. ～を楽しむ (e_____)
68. 湖 (l_____)
69. ゆうれい (g_____)
70. 日付 (d_____)
71. 労働、労力 (l_____)
72. それにもかかわらず (r_____)
73. 半分の (h_____)
74. 輸出 (e_____)
75. 糸 (t_____)
76. ～を買う (b_____)
77. ～を知らせる、通知する (i_____)
78. 優勝者 (c_____)
79. ～を支払う (p_____)
80. 軍人、兵隊 (s_____)
81. ～を破滅させる (r_____)
82. ～を料理する (c_____)
83. 講演、講義 (l_____)
84. 刑務所 (j_____)
85. ～を予期する、期待する (e_____)
86. 塔 (t_____)
87. 家族 (f_____)
88. 働く、仕事をする (w_____)
89. ～を破壊する (d_____)
90. 運河、人工水路 (c_____)
91. 島 (i_____)
92. 宗教 (r_____)
93. 類似して、同様に (s_____)
94. それの、その (i_____)
95. ぬれた、湿った (w_____)
96. 頼りになる、信頼できる (r_____)
97. 市長 (m_____)
98. ちり、ほこり (d_____)
99. より小さい、少数派の (m_____)
100. 芸術, 美術 (a_____)
101. どうぞ (p_____)
102. ～を書く (w_____)
103. だれの (w_____)
104. 恐怖、大あわて (p_____)
105. 作る人、製作者 (m_____)
106. 意識、正気 (c_____)
107. がん (c_____)
108. 波 (w_____)

これで終わりです。手をあげて、先生を呼んでください。

Appendix 8.3.
Summary of criteria derived from the six books
[α/χ] calm, (watch, yacht,) hot, (cotton)
[A] map, (bat, lamb,) plaid , aunt, (laugh)
[A/α:] fast, (half)
[α:ρ/α:] father, star, (par, car,) are, heart
[ς] love, (son,) blood, (flood,) country, (rough,) cup, (bus, lucky)
[ςρ] courage, current
[E] sofa, (ago,) moment, (bullet,) April, (competitive,) lemon, (collect, common,) famous, suppose, (circus, album)
[Eρ/E] calendar, paper, (over, sister,) circumference, doctor, (actor,) martyr, picture, (censure,) Saturday
[E:ρ/E:] early, (earth,) term, (her,) err, bird, (first,) works, turn, (nurse,) Byrd

[Y] many, aesthete, (aerial,) said, (again,) says, bed, (bet,) bread, (head,) their, (heir,) jeopardy, friend, Oedipus, bury

[ι] message, captain, roses, (basket,) been, (Greenwich,) forfeit, (foreign,) valley, hit, (bit,) parliament, (marriage,) lilies, (sieve,) women, busy, (business,) build, gym, (bicycle, cylinder)
[ι:] encyclopaedia, (Caesar,) picayune, complete, (eve,) sea, (beat,) tree, (bee,) people , conceit, (receive,) donkey, (key,) police, (ski,) piece, (field, babies,) fjord [fi:o:d, fjo:d, fi:o:rd, fjo:rd], amoeba, chamois, city, (baby)

[χ] dog, (often,) cough
[χ:] all, (ball,) taught, (caught, autumn,) law, bought, (four)
[χ:/χ] song
[χ:ρ/χ:] war, board, door, (floor,) morning, (or, store, bore, glory, corn)

[υ] wolf, (bosom,) book, (took,) would, (could; l is a silent letter), put, (beautiful)
[υ:] lieutenant, (lieu,) grew, Sioux, do, (woman,) shoe, (canoe,) school, (food,) soup, (you,) rule, (ruin,) fruit

[Oι] maestro, aye, height, (stein,) geyser, eye, like, (bite, sigh, right,) diamond, pie, (tie,) coyote, buy, try, (fly)
[Oυ] out, (mouth, bough,) cow, (owl)

[ει] make, (take,) Mae, rain, (sail,) gauge, day, (pay,) crepe, (suede,) break, (great,) matinee[nei], (negligee,) rein, (veil, weigh,) they, lingerie

[χι] choice, (oil, avoid,) toy, (boy)
[ου] chauffeur, beau, (plateau,) yeoman, sew, rope, (note,) road, (boat,) toe, (hoe,) brooch, soul, (though,) low

[OυEρ/OυE] flour, (our,) flower
[YEρ/YE] hair, (air,) parent, (various,) care, bear, (pear,) very
[ιEρ/ιE] deer, (beer,) fear, (hear,) serious, (cereal,) pierce, here
[υEρ/υE] poor, (boor,) tour, (tourist,) curious, (sure)

[β] bed, (job,) rabbit, (shabby)
[δ] dog, (bed,) sudden, loved
[δϑ] gradual, bridge, (judge,) soldier, adjective, procedure, (educate,) gem, (giant,) June, (jump,) pigeon, region

[δζ] records, (demands)
[ϕ] fat, (fifty,) effort, (staff,) laugh, (rough,) photograph, (phantom,) sapphire
[γ] go, (leg,) beggar
[ϑ] regime[reigi:m], (rouge,) jabot(foreign word), usual, vision, (division, fusion,) pleasure, (measure,) equation, seizure, azure, luxurious,
[γζ] executive, (exhibit)

[η] house, (high,) whole, (who)
[ιδ] graded
[φ] opinion, (onion,) hallelujah, yes, (yet)
[φυ] popular, (cure)
[φυ:] beauty, feud, new, (few,) lieu, (lieutenant,) cute, (excuse,) suit(sometimes u:), vacuum
[κ] camp, (copy,) accord, (accuse,) saccharin, school, (chemistry,) back, (neck,) lacquer, (racquet,) biscuit, (circuit,) kick, (king,) liquor, picturesque, (cheque)
[κφυ:] queue
[κσ] accent, (accept, accident,) six, (ax,) excite
[κΣ] anxious, (complexion)
[κω] acquire, queen, (square, queen)
[λ] lap, (sale,) fellow, (illumination)
[μ] man, (sum,) common, (hammer)
[ν] nap, (run,) connect, (cannon)
[N] uncle, (anger,) song, (singer, king, finger)
[Nγ] finger
[οε] Köchel(only for foreign words)
[π] pen, (cup,) Clapham, (shepherd,) happen, (happy)
[ρ] run, (rain,) marry, (merry,)
[σ] center, (city,) sun, (sister,) pass, (lesson)
[Σ] ocean, chef, (moustache,) speciality, (associate,) special, (precious,) sure, (sensuous, sugar,) conscience, (conscious,) ship, (fish,) cushion, (fashion,) mansion, (tension,) mission, (profession, permission,) pressure, (tissue,) censure, (cocksure,) station, (superstitious)
[τ] hoped, wrecked, tree, (set,) Thames, (Chatham,) letter, (button)
[T] think, (both)
[Δ] that, (other, bathe)
[τσ] tests, (students,) Nazi(loanword), pizza(loanword)
[τΣ] concerto, (cello,) chalk, (chart,) actual, (situation,) catch, (match,) question, picture, nature, (century)
[ϖ] Stephen([sti:vn]), vote, (service, have,) skivvy
[ζ] music, (husband, nose,) dessert, (possess,) xerox, (anxiety,) zoo, (lazy,) dazzle, (fuzz)

[ω] one, (everyone,) wall, (wait)
Note. These rules were written when at least one book presented both the rule and an example. Although they include some rules that look rather exceptional, what is treated as exceptional is different among the six books.

Appendix 8.4

Diversion of responses from the oral and written versions

No.	Answer	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	Total	Proportion of d)
1	dog	71	0	0	0	0	0	0	0	0	71	.00
2	mouse (mice)	33	0	0	31	2	0	2	0	2	70	.44
3	sock(s) (sox)	46	0	6	6	2	1	3	0	6	70	.09
4	maybe	39	1	0	9	6	0	1	14	0	70	0.13
5	perfect (parfect, perfitt, poifect)	55	1	2	7	0	0	1	0	4	70	0.10
6	interpretation	4	0	1	1	5	4	2	2	48	67	0.01
7	believe	24	0	0	26	1	2	4	0	13	70	0.37
8	bridge	50	0	1	13	1	2	1	0	1	69	0.19
9	wash	64	1	1	3	0	0	0	0	0	69	0.04
10	he	56	3	2	4	3	0	2	0	0	70	0.06
11	aunt	39	7	1	4	6	1	1	1	7	67	0.06
12	animal	68	0	0	0	0	0	0	0	0	68	0.00
13	driver	61	0	0	8	0	0	0	0	0	69	0.12
14	independence	15	0	0	1	8	3	2	0	39	68	0.01
15	explode	7	0	0	2	5	9	2	0	44	69	0.03
16	behind	21	2	0	2	22	8	0	4	10	69	0.03
17	alive	29	0	0	8	0	1	2	0	29	69	0.12
18	fruit(s)	19	0	0	40	0	1	9	0	0	69	0.58
19	group	32	0	2	19	0	4	4	0	8	69	0.28
20	warn	16	0	1	0	5	10	0	1	36	69	0.00
21	friend(s) (frend(s))	69	0	0	0	0	0	0	0	0	69	0.00
22	start	68	0	2	0	0	0	0	0	0	70	0.00
23	abroad	28	2	1	6	1	4	2	1	24	69	0.09
25	usually	36	3	3	20	6	0	0	0	2	70	0.29
26	bend	8	1	2	1	7	10	1	3	37	70	0.01
27	warmth	5	0	0	0	56	6	0	1	0	68	0.00
28	lady	33	1	0	16	0	1	4	1	14	70	0.23
29	three	57	4	0	5	2	0	0	0	0	68	0.07
30	invitation	3	3	0	2	38	8	1	1	15	71	0.03
31	open	67	0	0	0	2	0	1	0	0	70	0.00
32	theater (theatre)	15	0	3	4	0	7	3	0	37	69	0.06
33	deserve	1	0	1	0	0	4	2	0	61	69	0.00
34	ahead	8	1	0	0	7	11	1	0	42	70	0.00
35	afford	5	0	0	0	2	5	1	0	56	69	0.00
36	another	47	1	3	2	1	4	3	2	7	70	0.03
37	bring	53	1	1	2	1	2	3	0	6	69	0.03
38	birthday	67	1	0	3	0	0	0	0	0	71	0.04
39	orchestra	12	0	0	23	0	3	2	0	30	70	0.33
40	king	63	2	0	2	3	0	0	0	0	70	0.03
41	pleasure	13	0	0	6	1	4	2	1	42	69	0.09
42	happily	5	3	1	3	50	4	0	2	1	69	0.04
43	guy	18	0	0	10	4	3	3	1	30	69	0.14
44	vast	3	2	0	1	4	12	1	4	41	68	0.01
45	informal	11	0	2	0	1	3	1	0	51	69	0.00
46	stare	5	0	0	1	17	17	1	4	24	69	0.01
47	content	8	0	0	0	1	11	1	0	48	69	0.00
48	tragedy	3	1	0	1	2	5	1	0	55	68	0.01
50	origin	22	1	0	5	3	2	1	0	33	67	0.07
51	supporter	8	4	1	4	4	4	2	1	40	68	0.06
52	require	7	0	1	3	0	5	1	3	48	68	0.04
53	feed	21	3	0	2	12	4	2	12	12	68	0.03
54	possession	2	0	0	2	3	10	1	1	48	67	0.03
56	guest	16	0	1	10	0	4	2	1	33	67	0.15
57	dignity	4	0	0	0	0	5	1	0	57	67	0.00
58	ride	51	0	1	5	1	0	2	0	8	68	0.07
59	egg(s)	59	0	0	8	0	0	2	0	0	69	0.12
62	apple(s)	59	0	1	8	0	0	0	0	0	68	0.12
63	treasure	11	0	0	12	0	3	3	0	39	68	0.18
65	pupil	6	0	1	2	3	7	1	0	46	66	0.03
66	onion	55	0	0	1	1	1	4	0	6	68	0.01

(Appendix 8.4 continues)

(Appendix 8.4 continued)

No.	Answer	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	Total	Proportion of d)
67	enjoy	58	0	1	2	0	1	4	0	2	68	0.03
68	lake	57	2	0	7	0	0	2	0	0	68	0.10
69	ghost	8	0	0	36	1	4	4	2	11	66	0.55
70	date	32	1	0	7	17	2	1	3	5	68	0.10
71	labor(labour)	6	0	2	0	1	4	1	0	52	66	0.00
72	regardless	4	0	0	0	2	6	1	0	52	65	0.00
73	half (hafe, etc.)	37	1	2	3	1	1	4	0	18	67	0.04
75	thread	1	0	0	1	2	4	1	0	56	65	0.02
76	buy	63	3	0	3	0	0	0	0	0	69	0.04
77	inform	9	0	0	1	3	9	1	2	42	67	0.01
78	champion (champ, etc.)	20	0	2	19	1	4	4	0	18	68	0.28
79	pay	32	0	0	4	1	0	1	2	28	68	0.06
80	soldier	3	0	0	16	1	3	2	0	43	68	0.24
81	ruin	4	1	0	0	0	4	1	0	58	68	0.00
82	cook	64	3	2	0	1	0	0	0	0	70	0.00
83	lecture	15	0	0	0	1	3	1	0	51	71	0.00
84	jail	4	0	0	1	4	8	0	0	51	68	0.01
85	expect	17	0	4	0	0	3	1	2	44	71	0.00
86	tower	36	1	3	11	1	2	4	0	13	71	0.15
87	family (femily)	64	0	0	6	0	0	0	0	0	70	0.09
88	work (wark)	69	0	0	2	0	0	0	0	0	71	0.03
90	canal	2	0	0	2	2	4	1	1	56	68	0.03
91	island (iland, etc.)	34	4	0	12	0	1	5	0	15	71	0.17
92	religion	5	0	1	6	3	3	1	0	51	70	0.09
93	similarly	1	0	1	6	22	11	1	4	24	70	0.09
94	its	21	9	4	4	19	5	0	4	4	70	0.06
95	wet	25	2	3	6	3	6	1	9	16	71	0.08
97	mayor	9	0	1	5	2	5	3	0	44	69	0.07
99	minor	7	2	1	3	8	6	1	5	36	69	0.04
100	art	52	0	0	2	0	1	8	0	6	69	0.03
101	please	51	1	4	2	0	0	2	0	10	70	0.03
102	write	66	1	1	2	0	0	0	0	0	70	0.03
103	whose	39	7	2	8	13	1	0	0	0	70	0.11
106	consciousness	1	0	1	1	6	8	1	0	52	70	0.01
107	cancer	9	0	3	1	3	5	1	1	46	69	0.01
108	wave	29	0	4	15	4	3	4	0	11	70	0.21
	Total	2735	87	84	548	421	332	154	96	2155	6612	548/6612 = .08

Appendix 8.5

Analysis of responses that belonged to (d)

Answer	Response
mouse (mice) (*n* = 31) [/ˈmaʊs : ˈmaʊs/ mice /ˈmaɪs/ ; 5 mous · es/ -ɪz/]	Acceptable in terms of English spelling rules: mouce (*n* = 2), mous_ (*n* = 8), mause (*n* = 4), mauce (*n* = 1), maus_ (*n* = 2) Unacceptable: mius_ (*n* = 1, unacceptable: iu), mauth (*n* = 2, th) m_use (*n* = 2), mouth (*n* = 9): (c) same as muse and mouth
sock(s) (sox) (*n* = 6) [sɑk/ sɑks/]	Acceptable in terms of English spelling rules: so_ks (*n* = 1), soxe (*n* = 1) Unacceptable: shors (*n* = 1, hors) (same as shor), shoseunder (*n* = 1, hoseunder), soxth (*n* = 1, th) shoes (*n* = 1): (c) same as shoes
maybe (*n* = 9) [/ˈmeɪ bi(ː), ˈme bi/]	Acceptable in terms of English spelling rules: ma_by (*n* = 2), maybee (*n* = 1), meybe (*n* = 1), meibe (*n* = 1), maiby (*n* = 1) Unacceptable: mcybe (*n* = 1, c) mort_ (*n* = 1), may__ (*n* = 1): (c) same as mort and may

(Appendix 8.5 continues)

(Appendix 8.5 continued)

Answer	Response
perfect (parfect, perfect, perfitt, poifect) (*n* = 7) [pɑ͡ɹ fɪ kt]	Acceptable in terms of English spelling rules: porfect (*n* = 2), purfect (*n* = 1) Unacceptable: pa fect (*n* = 1), parphect (*n* = 1), peafect (*n* = 1, ea), purafect (*n* = 1, a)
interpretation (*n* = 1) [ˌɪntə͡ɹ prɪ tə͡ɪʃə n , -prɵ -ˌ]	Acceptable in terms of English spelling rules: interprɪtation (*n* = 1)
believe (*n* = 26) [ˌbɪ lɪ͡ɹ v , bɵ - , bˈɪ͡ɹ vˌ]	Acceptable in terms of English spelling rules: beli ve (*n* = 7), beleave (*n* = 3), bili ve (*n* = 1), b li ve (*n* = 1), bileave (*n* = 1), b leave (*n* = 1) Unacceptable: belieb_ (*n* = 1), b rieve (*n* = 1), beri ve (*n* = 2), b lli be (*n* = 1), b ri ve (*n* = 3), bell bu (*n* = 1, _), briabe (*n* = 1, ia), brrive (*n* = 1, rr) br eve (*n* = 1): (c) same as breve*
bridge (*n* = 13) [ˌbrɪ͡ dʒˌ]	Acceptable in terms of English spelling rules: bri ge (*n* = 6), brrid e (*n* = 1) Unacceptable: bridze (*n* = 1, z), brihdge (*n* = 1, h), bri gd (*n* = 1, d), bir dge (*n* = 1, i), bri gh (*n* = 1, gh) breech (*n* = 1): (c) same as breech
wash (*n* = 3) [ˌwɑ͡ʃ , wɔ͡ɹʃ , ((non-standard)) wɔ͡əʃ , wɑ͡əʃ ˈ wɔ͡ʃˌ]	Acceptable in terms of English spelling rules: washe (*n* = 2) Unacceptable: washu (*n* = 1)
he (*n* = 4) [ˌ((weak)) hi , iɹ , i; ((strong) hɪ͡ɹˌ]	Unacceptable: him (*n* = 2), his (*n* = 2): (c) same as him and his
aunt (*n* = 4) [ˌæ͡nt ˈ ɑ͡ɹˌ]	Acceptable in terms of English spelling rules: arnt (*n* = 1) Unacceptable: a nto (*n* = 1), a ncle (*n* = 2, ncle)
driver (*n* = 8) [ˌdrɑ͡ɪ vəˌ]	Unacceptable: drive_ (*n* = 8): (c) same as drive
independence (*n* = 1) [ˌɪ͡ndɪ pɛ͡ ndəns ˌ]	Unacceptable: independent_ (*n* = 1): (c) same as independent
explode (*n* = 2) [ˌɪksplɔ͡ʊ d, eks-ˌ]	Unacceptable: exprord_ (*n* = 1, r) explore (*n* = 1): (c) same as explore
behind (*n* = 2) [bɪ hɑ͡ɪ nd , bɵ -]	Unacceptable: behi de (*n* = 1, _) back (*n* = 1): (c) same as back
alive (*n* = 8) [ˌəlɑ͡ɪ vˌ]	Unacceptable: arive (*n* = 6), alivaly (*n* = 1, aly) arrive (*n* = 1): (c) same as arrive
fruit(s) (*n* = 40) [ˌfrʊ͡ɹ tˌ]	Acceptable in terms of English spelling rules: fruite (*n* = 3), fru t(s) (*n* = 3), fru te (*n* = 2), fro ts (*n* = 1), frout(s) (*n* = 2) Unacceptable: furuit (*n* = 1), fru tu (*n* = 2), furu ts (*n* = 1), frult (*n* = 1, l), fluit (*n* = 1, l), fru th (*n* = 1, th), flurt (*n* = 1, l, r), flu t (*n* = 2, l), fr__te (*n* = 1, _), floot(s) (*n* = 5, l), flowt (*n* = 1, l, ow), f ulusu (*n* = 1, l, su), f ulute (*n* = 1, l), fulu th (*n* = 1, l, th) flu te(s) (*n* = 8), flouts (*n* = 1): (c) same as flute and flout
group (*n* = 19) [ˌgrʊ͡ɹ pˌ]	Acceptable in terms of English spelling rules: groop (*n* = 1), gro p (*n* = 3), gr up (*n* = 4), gro pe (*n* = 2) Unacceptable: gur up (*n* = 1), growp (*n* = 3, ow), glo p (*n* = 1, l) gar__d (*n* = 1, a, _), gla pe (*n* = 1, l, a), guarp (*n* = 1, a, _), g__upe (*n* = 1, _)

(Appendix 8.5 continues)

(Appendix 8.5 continued)

Answer	Response
abroad (*n* = 6) [/əbrɔ̄ɹ d , ((米+))ə brɑ̄ɹ d/]	Unacceptable: abro d (*n* = 3), abroud (*n* = 2), abloud (*n* = 1)
eager (*n* = 1) [/ĭɹgə/]	Acceptable in terms of English spelling rules: e gar (*n* = 1)
usually (*n* = 20) [/jūɹʒ uə li , jūɹʒə li/]	Acceptable in terms of English spelling rules: usual y (*n* = 4), us elly (*n* = 1) Unacceptable: usuar y (*n* = 1), useally (*n* = 4, a), useually (*n* = 1, ua), usearly (*n* = 1, ar), useual y (*n* = 1, ua), userally (*n* = 1, ra), usear y (*n* = 2, a), useuar y (*n* = 1, ua), u aully (*n* = 1, _, au) usual__ (*n* = 1), useful_ (*n* = 1): (c) same as usual and useful
bend (*n* = 1) [/bē nd/]	Unacceptable: belong (*n* = 1): (c) same as belong
lady (*n* = 16) [/lēɪ di/]	Acceptable in terms of English spelling rules: ledy (*n* = 5), leddy (*n* =1) Unacceptable: ludy (*n* = 1, u), ldby (*n* = 1, d), labes (*n* = 1, d), leber (*n* = 1, b, er) le y (*n* = 1), leady (*n* = 3),* left (*n* = 1), life (*n* = 1): (c) same as ley, leady, left, and life
three (*n* = 5) [/θrĭɹ/]	Unacceptable: thiree (*n* = 1, i), tird_ (*n* = 1, ir, d) third (*n* = 3): (c) same as third
invitation (*n* = 2) [/ĭnvɪ tēɪʃə n/]	Unacceptable: invitertion (*n* = 1, er) invite____ (*n* = 1): (c) same as invite
theater (theatre) (*n* = 4) [/θĭɹə tə̄ | θĭətə , θ iē tə/]	Acceptable in terms of English spelling rules: thiater (*n* = 1), thioter (*n* = 1) Unacceptable: t eather (*n* = 1, _), th ater (*n* = 1, _)
another (*n* = 2) [/ənʌ̆ðə/]	Unacceptable: anyone_ (*n* = 1), a piece of (*n* = 1) : (c) same as anyone and a piece of
bring (*n* = 2) [/brĭŋ/]	Unacceptable: brive (*n* =2, ve)
birthday (*n* = 3) [/bə̆ɹθ dēɪ/]	Unacceptable: barthday (*n* = 2), bearsday (*n* = 1)
orchestra (*n* = 23) [/ɔ̄ɹkə strə , -kest- | ɔ̄ɹ-/]	Acceptable in terms of English spelling rules: o chestra (*n* = 2), orc estra (*n* = 3), orckestra (*n* = 1), or kestra (*n* = 4), ou kestra (*n* = 3), o__kestra (*n* = 2) Unacceptable: orchestora (*n* = 1), or kestora (*n* = 1), o__kestora (*n* = 2), orc estla (*n* = 1, l), orck stra (*n* = 1, _), or kestla (*n* = 1, l), oh kestr_ (*n* = 1, _)
king (*n* =2) [/kĭŋ/]	Unacceptable: kink (*n* = 1), kind (*n* = 1): (c) same as kink and kind
pleasure (*n* = 6) [/plē̆ʒə/]	Acceptable in terms of English spelling rules: ple sure (*n* = 1), ple ser_ (*n* = 1) Unacceptable: preasure (*n* = 3), pre sure (*n* = 1)
happily (*n* = 3) [/hæ̆ pɭ i | hæ̆ pɪ li/]	Unacceptable: happy__ (*n* = 3): (c) same as happy
guy (*n* = 10) [/gə̆ɪ/]	Unacceptable: gay (*n* = 10): (c) same as gay*
vast (*n* = 1) [/væ̆ st | vɑ̄ɹ st/]	Acceptable in terms of English spelling rules: varst (*n* = 1): acceptable
stare (*n* = 1) [/stēə/]	Unacceptable: stear_ (*n* = 1)): (c) same as stear (= stair)*
tragedy (*n* = 1) [/træ̆ dʒə di/]	Unacceptable: trag dy (*n* = 1, _)
origin (*n* = 5) [/ɔ̄ərə dʒɪ n, ɑ̄ə - , -dʒə n | ɔ̄r-/]	Unacceptable: orig n (*n* = 2) origen (*n* = 2),* orient (*n* = 1): (c) same as Origen and orient

(Appendix 8.5 continues)

(Appendix 8.5 continued)

Answer	Response
supporter (*n* = 4) [/sə pɔ́ə tə́ | -pɔ́ː -/]	Acceptable in terms of English spelling rules: sup_orter (*n* = 2) Unacceptable: stand____ (*n* = 1), side_____ (*n* = 1) : (c) same as stand and side
require (*n* =3) [/rɪ kwáɪə/]	Acceptable in terms of English spelling rules: recquire (*n* = 1) Unacceptable: requwire (*n* = 1), req_ire(*n* = 1, _)
feed (*n* = 2) [/fíː d/]	Acceptable in terms of English spelling rules: fead (*n* = 1) Unacceptable: fa_de (*n* = 1): (c) same as fade
possession (*n* = 2) [/pə zéʃə n/]	Acceptable in terms of English spelling rules: pos_ession (*n* = 1) Unacceptable: possess___ (*n* = 1): (c) same as possess
guest (*n* = 10) [/gést/]	Unacceptable: ga_st (*n* = 1),* g_est (*n* = 9)*: (c) same as gast and gest
ride (*n* = 5) [/ráɪ d/]	Unacceptable: raide (*n* = 1), rade (*n* = 1, a), riad_ (*n* = 1, a) rise (*n* =2): (c) same as rise
egg(s) (*n* = 8) [/ég, (American)éɪg/]	Acceptable in terms of English spelling rules: eagg (*n* = 1), ege (*n* = 6), eig (*n* = 1)
crawl (*n* = 3) [/krɔ́ːl , (American) krάːl/]	Unacceptable: crowl (*n* = 2) crow_ (*n* = 1): (c) same as crow
respect(s) (*n* = 5) [/rɪ spé kt/]	Acceptable in terms of English spelling rules: rispect (*n* =3) Unacceptable: r_spect (*n* =1, _) right__ (*n* =1): (c) same as right
apple(s) (*n* = 8) [/ǽp l/]	Acceptable in terms of English spelling rules: ap_le (*n* =1) Unacceptable: appul_ (*n* =3), appele (*n* =1, e), aprl_ (*n* =1, r), apor_ (*n* =1, o) ap___ (*n* =1): (c) same as ap
treasure (*n* = 12) [/trézə/]	Acceptable in terms of English spelling rules: treasur_ (*n* = 1), tre_sure (*n* = 1), treagure (*n* = 1), tr_ager_ (*n* = 1), tre_ger_ (*n* = 2) Unacceptable: tleasure (*n* = 1, l), tre_ture (*n* = 1, t), tare_ser_ (*n* = 1, a), tle_jar_ (*n* = 1, l), tr_agger (*n* = 1, g), tr_agver_ (*n* = 1, v)
pupil (*n* = 2) [/pjúː p l/]	Unacceptable: pepil (*n* = 1, e), purpil (*n* = 1, r)
onion (*n* = 1) [/ʌ́njə n/]	Acceptable in terms of English spelling rules: onione (*n* = 1)
enjoy (*n* = 2) [endʒɔ́ɪ , ɪ n- , ə n- | ɪn- , en- , ə n-/]	Unacceptable: excithing (*n* = 1, xcithing) exciting (*n* = 1): (c) same as exciting
lake (*n* = 7) [/lékɪ k/]	Unacceptable: leike (*n* = 1), laki (*n* = 1, i) late (*n* = 2), like (*n* = 2), leke (*n* = 1): (c) same as late, like, and leke (**leke**/lé k/ = like)
ghost (*n* = 36) [/góʊst/]	Acceptable in terms of English spelling rules: ghoast (*n* = 1), g_ost (*n* = 5), goost (*n* = 3), g_oast (*n* = 7), goust (*n* = 5) Unacceptable: gohst (*n* = 3, h), gorst (*n* = 11, r), gourst (*n* = 1, r)

(Appendix 8.5 continues)

(Appendix 8.5 continued)

Answer	Response
date (*n* = 7) [ˈdeɪ t]	Acceptable in terms of English spelling rules: dayte (*n* = 1), dete (*n* = 1) Unacceptable: day_ (*n* = 5): (c) same as day
half (**harf**/hɑːf f | hɑːɹ f/; **hafe**/hæf f/, haff, harf, haulf, hahf, halfe, haa(l)f, ha'f, hofe (*n* = 3) [ˈhæf f, hɑːɹ f | hɑːɹ f]	Unacceptable: herf (*n* = 2, er) harth (*n* = 1): (c) same as harth [**harth**/hɑːəθ | hɑːɹθ/= hearth], th)
thread (*n* = 1) [ˈθred d]	Unacceptable: threst (*n* = 1, st)
buy (*n* = 3) [ˈbaɪ]	Unacceptable: bay (*n* = 2), brow (*n* = 1): (c) same as bay and brow
inform (*n* = 1) [ˈɪnfɔːə m | -fɔːɹ m]	Unacceptable: imform (*n* = 1, m)
champion (*n* = 19) [ˈtʃæ mpi ən]	Unacceptable: chanpion (*n* = 10), chimpion (*n* = 1, i) c_ampion (*n* = 8): (c) same as **cam·pi·on**/kæ mpiə n/
pay (*n* = 4) [ˈpeɪ]	Unacceptable: pripade (*n* =1, ripade), pripeid (*n* =1, ripeid) pent (*n* =1), price (*n* =1): (c) same as pent and price
soldier (*n* = 16) [ˈsoʊ ldʒə]	Acceptable in terms of English spelling rules: soldior (*n* = 1), soldire (*n* = 1), solj er (*n* = 1), solg er (*n* = 1), soljure (*n* = 1) Unacceptable: sorujer (*n* = 1), sorud er (*n* = 1), solgeer (*n* = 1, e), solid er (*n* = 1, i), sails er (*n* = 1, ai, s), solsure (*n* = 1, s), sorsure (*n* = 1, s) sol____ (*n* = 1), sold er (*n* = 3)*: (c) same as sold and solder
jail (*n* = 1) [ˈdʒeɪ l, ((米+)) dʒeɪjəl]	Acceptable in terms of English spelling rules: jeil (*n* = 1)
tower (*n* = 11) [ˈtaʊə]	Acceptable in terms of English spelling rules: tau r (*n* = 1), towar (*n* = 1), tow_r (*n* = 4), towur (*n* = 1) Unacceptable: t_war (*n* = 1, _), torw_ (*n* = 1, r, _), trast (*n* = 1, rast) tawer (*n* = 1)*: (c) same as tawer
family (**fem - i·ly**/fɛ məli/) (*n* = 6) [ˈfæ m i]	Acceptable in terms of English spelling rules: familly (*n* = 2), fammily (*n* = 1) Unacceptable: famiry (*n* = 1), fumily (*n* = 1, u), frmily (*n* = 1, r)
work (**wark**/wɑːk wɔːə k wɔːə k | wɑːɹ k, wɔːɹ k/) (*n* = 2) [ˈwəːɹ k]	Unacceptable: worh (*n* = 1, h) whom (*n* = 1): (c) same as whom
canal (*n* = ?) [ˈkə næ l]	Unacceptable: canaw (*n* = 1, w) channel (*n* = 1): (c) same as channel
island (**i- land**/aɪlə nd/= island, i(s)lant, ile - dand, i(s)le) (*n* = 12) [ˈaɪlə nd]	Unacceptable: isrand (*n* = 1), i rand (*n* = 10), irland (*n* = 1, r)
religion (*n* = 6) [ˈrɪ lɪ dʒə n]	Acceptable in terms of English spelling rules: reliegion (*n* = 1) Unacceptable: rerigion (*n* = 2), rerigeon (*n* = 1, eo), resition (*n* = 1, s, t), reviel__ (*n* = 1, v, ie, l)

(Appendix 8.5 continues)

(Appendix 8.5 continued)

Answer	Response
similarly (*n* = 6) [/sɪ mɹ li/]	Acceptable in terms of English spelling rules: simila_ly (*n* = 1) Unacceptable: simira_ly (*n* = 1), s_e_milar__ (*n* = 1, _) similar__ (*n* = 3): (c) same as similar
its (*n* = 4) [/ɪts , ɪ ts/]	Unacceptable: it_ (*n* = 4): (c) same as it
wet (*n* = 6) [/wɛ t/]	Acceptable in terms of English spelling rules: we_a_t (*n* = 3) Unacceptable: we_e_ (*n* = 1), w_h_et (*n* = 1), w_a_t_e_ring (*n* = 1): (c) same as wee, whet, and watering
reliable (*n* = 2) [/rɪ laɪə b!/]	Unacceptable: rel_a_iable (*n* = 1), re_t_iable (*n* = 1, t)
mayor (*n* = 5) [/mɛɪɚ , (mɛɚ | mɛə/]	Acceptable in terms of English spelling rules: ma_je_r (*n* = 1) Unacceptable: may_e_r (*n* = 4): same as Mayer*
dust (*n* = 10) [/dʌ st/]	Unacceptable: d_a_st (*n* = 10): same as dast
minor (*n* = 3) [/maɪ nɚ/]	Acceptable in terms of English spelling rules: min_ne_r (*n* = 1) Unacceptable: mi_ni_mam (*n* = 1, imam) min_e_r (*n* = 1): (c) same as miner
art (*n* = 2) [/ɑɚt | ɑɹt/]	Unacceptable: art_i_st (*n* = 1), a_e_t (*n* = 1)*: (c) same as artist and aet.
please (*n* = 2) [/plɪɹ z/]	Unacceptable: p_r_ease (*n* = 2)
write (*n* = 2) [/raɪ t/]	Acceptable in terms of English spelling rules: wri_t_te (*n* =1) Unacceptable: wri_gh_t_ (*n* =1): same as wright
whose (*n* = 8) [/hɯɹ z; hʊ z, u(ɹ)z/]	Acceptable in terms of English spelling rules: who_e_s (*n* = 2) Unacceptable: who__ (*n* = 5), who_m_ (*n* = 1): (c) same as who and whom
panic (*n* = 5) [/pæ nɪ k/]	Acceptable in terms of English spelling rules: panic_k_ (*n* = 2) Unacceptable: puni_k_ (*n* = 2, u) pu_n_ic (*n* = 1): same as Punic
maker (*n* = 2) [/mɛɪ kɚ/]	Unacceptable: mader (*n* = 2, d)
consciousness (*n* = 1)[/kɑ nʃə snə s/]	Unacceptable: cons_ious____ (*n* = 1, _)
cancer (*n* = 1) [/kæ nsɚ/]	Acceptable in terms of English spelling rules: can_se_r (*n* = 1)
wave (*n* = 15) [/wɛɪ v/]	Unacceptable: wa_b_e (*n* = 1), we_a_b_e_ (*n* = 1), we_a_b (*n* = 2), we_e_b (*n* = 1), w_er_ve (*n* = 2, r), w_he_ve (*n* = 1, h), we_a_rb (*n* = 2, r), w_er_vu (*n* = 1, r) we_a_ve (*n* = 1),* we_b_ (*n* = 2), wi_d_e (*n* = 1): (c) same as weave, web, and wide

Note. [/phonetic symbol/] = From the *Taishukan's Unabridged Genius English-Japanese Dictionary* (Konishi & Minamide, 2002). *a response that would have been scored as correct if there were no other existent words. Originally another scoring method was made based on the Roman letter rules. However, it was not used in Study 2A because the criterion was considered too lenient. The correlation between the oral version and this method in the written version was .90.

Appendix 9.1 *All the responses in the Derivation Subsection from the participants*

2-1		2-2		2-3					
work		art		destroy					
worker		artist		destruction					
response	*n*	response	*n*	response	*n*	response	*n*	response	*n*
introduce	1	a	1	d	1	destroduction	1	desty	1
wark	1	airtist	2	danger	1	destroe	1	distrier	1
warkes	1	archtecture	1	de	2	destroyer	4	distroyment	1
woker	3	aretist	1	decsonry	1	destrogy	1	dstroyer	1
worcar	1	art	3	dededeking	1	destroy	2	dstruction	1
worce	1	artcodeneiter	1	des	1	destroial	2	play	1
worer	1	arter	56	describe	1	destroiance	1	story	1
work	8	artes	3	deseroying	1	destroid	8	stroy	1
worke	1	artess	3	dest	9	destroyed	5	Total	619
workes	4	artest	96	destance	1	destroiement	1		
workest	1	articet	1	destation	1	destroieness	1		
worket	1	article	1	desteny	1	destroyer	50		
working	9	articter	1	dester	1	destroiery	2		
workist	1	artiest	2	destiny	1	destroies	6		
worker	1	arting	2	destion	1	destroiess	3		
workness	7	artisit	1	destion	1	destroiest	2		
works	4	artiste	1	destioraction	1	destroily	1		
worlier	1	artistt	1	destiss	1	destroiment	4		
Total	47	artitect	1	destluction	1	destroiner	1		
		artker	1	destment	1	destroiness	6		
		artman	5	destner	1	destroing	8		
		artner	1	destnity	1	destroir	1		
		artness	7	desto	1	destrois	1		
		artress	1	destoction	1	destroiss	2		
		arts	3	destoial	1	destroist	2		
		artst	4	destoier	2	destroisy	1		
		artter	1	destor	1	destroit	7		
		arttist	1	destoraction	1	destroition	3		
		artuce	1	destores	1	destroization	1		
		drawer	1	destorier	1	destroize	1		
		introduce	1	destorist	1	destrolon	1		
		write	1	destoroies	1	destroly	2		
		Total	206	destoroite	1	destroment	2		
				destoroiyer	1	destron	3		
				destoroyer	3	destroness	2		
				destory	3	destronetion	1		
				destoryer	1	destroral	1		
				destoryess	1	destrose	1		
				destorying	2	destrot	1		
				destorytion	1	destrotion	6		
				destotion	1	destroy	34		
				destoyer	2	destroy piano player	1		
				destoytion	1	destroya	1		
				destr	9	destroyear	1		
				destract	1	destroyed	9		
				destraction	15	destroyer	198		
				destrance	1	destroyes	3		
				destration	3	destroyest	1		
				destrayer	1	destroyet	1		
				destre	2	destroygion	1		
				destreat	1	destroyin	1		
				destreption	1	destroying	40		
				destrer	1	destroyion	1		
				destres	1	destroyly	1		
				destretion	1	destroyment	15		
				destrey	1	destroyness	3		
				destrical	1	destroyor	1		
				destriction	3	destroyr	1		
				destried	1	destroys	5		
				destrier	1	destroyt	3		
				destries	1	destroytest	1		
				destriness	2	destroytion	2		
				destription	1	destrst	1		
				destrition	2	destrtion	1		
				destrize	1	destruct	3		
				destro	11	destructure	1		
				destroaisan	1	destrution	1		
				destroction	4	destry	3		
				destrodaction	1	destu	1		

Note. response = response scored as incorrect. *n* = The number of test takers who wrote the response.

2-4				2-5				2-6	
criticize				inform				comfortable	
criticism				information				comfort	
response	*n*			response	*n*			response	*n*
c	1	criticiziy	1	iformation	1	infrmer	1	ancomfortable	1
ciriticition	1	criticizly	1	imfomtion	1	inform	3	be comfortable	1
ciriticized	1	criticizm	18	imforess	1	infromation	1	cofortabling	1
ciriticizem	1	criticizment	1	imformation	4	infromed	1	cofortablly	1
ciriticizess	1	criticizn	2	imformed	2	infromer	1	com	1
ciriticizm	1	criticizness	1	imformer	6	infromess	1	comfart	1
ciriticzer	2	criticizst	1	imformment	1	infromotion	1	comfartablation	1
citizen	1	criticiztion	1	imformrcion	1	infromi	1	comfartablement	1
city	1	criticizton	1	imfortepu	1	infroms	1	comfartables	1
crasy	1	criticizum	1	infarm	2	infront	1	comfatation	1
cricizetion	1	criticizy	2	infarmation	1	inormetion	1	comfor	7
criti	2	criticne	1	infarmed	1	Total	313	comforeadle	1
critic	27	criticy	3	infomation	5	infrmer	1	comforl	1
critical	14	criticze	1	infomeation	1	inform	3	comform	3
criticate	1	criticzed	1	infomer	4	infromation	1	comforness	2
criticc	2	criticzeness	2	infomness	1	infromed	1	comforta	2
critice	2	criticzer	1	infoormation	1	infromer	1	comfortab	4
criticese	1	criticzern	1	infor	2	infromess	1	comfortabelepu	1
critici	3	criticzes	1	inforce	1	infromotion	1	comfortability	6
criticice	1	criticzey	1	inform	13	infromi	1	comfortabirity	1
criticicle	1	crition	2	informaction	1	infroms	1	comfortabition	1
criticilis	1	critique	1	informaison	1	infront	1	comfortabity	1
criticily	1	critisis	1	informait	1	inormetion	1	comfortablal	1
criticiness	1	critisism	1	informaition	2	Total	313	comfortablance	1
criticion	1	critition	1	informal	25			comfortablate	1
criticisation	1	critiz	1	informality	1			comfortablation	8
criticise	3	critization	1	informar	2			comfortable	19
criticision	2	critize	1	informat	1			comfortable to	1
criticit	1	critized	2	informate	5			comfortableat	1
criticition	6	critizem	1	informatison	1			comfortablece	1
criticity	1	critizer	1	informattel	1			comfortablect	1
criticive	1	critizetion	1	informce	1			comfortabled	5
criticiz	9	critizy	1	informd	3			comfortableed	1
criticizaition	1	crity	1	informdconcent	1			comfortableing	1
criticizate	1	cryer	1	informe	2			comfortablely	2
criticization	1	cuticized	1	informeation	1			comfortablement	1
criticization	36	I don't know	1	informecition	1			comfortablence	1
criticize	19	informer	1	informed	48			comfortabler	39
criticizer	75	sy	1	informeishon	1			comfortables	13
criticized	44	Total	453	informence	2			comfortabless	11
criticizeful	1			informeness	1			comfortablest	3
criticizeition	1			informent	11			comfortabletion	5
criticizeless	1			informer	17			comfortablety	4
criticizeliy	1			informertion	4			comfortabley	4
criticizely	2			informes	7			comfortablineition	1
criticizem	1			informeson	1			comfortabling	9
criticizement	5			informess	2			comfortablint	1
criticizen	20			informest	3			comfortablis	1
criticizence	3			informethion	1			comfortablitation	1
criticizeness	7			informetion	30			comfortablition	2
criticizent	4			informily	1			comfortablity	5
criticizer	5			informing	4			comfortabliy	1
criticizerly	1			informition	2			comfortablly	5
criticizes	10			informiton	1			comfortabls	1
criticizese	1			informity	39			comfortably	30
criticizeses	1			informly	4			comfortabness	1
criticizess	8			informnation	1			comfortain	1
criticizesson	1			informness	3			comfortalbles	1
criticizest	5			informor	1			comfortall	1
criticizetion	15			informs	1			comfortance	1
criticizeus	1			informth	2			comfortaness	1
criticizey	2			informtion	3			comfortast	1
criticizial	1			informus	1			comfortate	1
criticizical	1			informuse	1			comfortation	12
criticizily	1			informy	1			comfortblet	1
criticizing	10			infornt	3			comfortbling	1
criticizion	4			inforrmaition	1			comforte	5
criticizist	2			infort	2			comforter	2
criticizition	4			infrm	1			comfortest	1

2-6		2-7				2-8			
comfortable		successful				able			
comfort		success				ability			
response	*n*	response	*n*			response	*n*	response	*n*
comfortion	5	abling	1	successsfuler	1	ab	1	available	1
comfortive	1	saccessful	1	succesztion	2	abilty	1	avob	1
comfortly	1	scceessfuly	1	successy	2	abirty	1	bable	1
comfortment	3	seccess	1	succested	1	abitity	1	be	1
comfortnance	1	suc	1	succestuler	1	abl	3	be able	2
comfortness	4	succed	4	succeszed	1	ablable	1	be able to	2
comforto	1	succeded	1	successsfuler	1	ablation	2	can	25
comfortunary	1	succeed	51	succful	1	able	49	capable	5
comfot	1	succeeded	1	sucessfuly	1	able to	16	car	1
comfotably	1	succeful	1	sucseed	1	ablead	1	could	2
comfotation	1	succerssfuler	1	sucsseed	1	ableble	1	disable	1
comfote	1	succes	9	suker	1	abled	10	enable	16
comfratablation	1	succese	1	usccessed	1	ableed	1	invitationing	1
comfrtion	1	succeser	1	Total	246	ablees	1	lan	1
comftortalling	1	succesful	1			ablefor	1	nesesally	1
comfurt	1	succesfuler	1			ableful	4	possibl	1
compleet	1	succesfuly	1			ablege	1	possible	13
comportabless	1	succesion	2			ableing	1	thing able	1
confort	2	success	2			ableis	1	to able	2
confortableat	1	successaly	1			ablele	1	unable	1
confortabled	1	successary	1			ablell	1	valuable	1
confortablen	1	successation	1			ablelly	1	will	1
confortation	1	successding	1			ablely	1	write	1
confortion	1	successe	1			ableme	1	Total	335
cooker	1	successed	8			ablement	3		
fomforte	1	successeed	1			ablen	2		
imice	1	successer	4			ablence	1		
mcomfortable	1	successery	1			ablenes	1		
momfortable	1	successes	1			ableness	24		
succession	2	successess	1			ablenss	1		
successness	1	successest	1			ablent	1		
uncomfortable	1	successfce	1			abler	28		
Total	294	successfight	1			ableress	1		
		successfu	1			ablery	1		
		successfuit	1			ables	13		
		successful	10			able's	1		
		successfulate	1			abless	11		
		successfule	2			ablest	2		
		successfuled	2			ablet	3		
		successfuled	1			ableter	1		
		successfuler	17			ableth	1		
		successfules	3			abletick	1		
		successfuletion	1			abletion	2		
		successfuli	1			abletly	1		
		successfuling	4			ablety	2		
		successfulir	1			abley	7		
		successfulist	1			abli	1		
		successfulity	1			ablial	1		
		successfull	1			abliaty	1		
		successfulld	1			abligation	1		
		successfuller	1			ablily	1		
		successfully	31			ablina	1		
		successfuls	2			ablineition	1		
		successfulsion	1			abling	9		
		successfulution	1			ablity	9		
		successfuly	13			ably	10		
		successfund	1			about	1		
		succession	20			abrity	1		
		successive	2			alber	1		
		successly	1			anable	2		
		successment	1			anbless	1		
		successness	2			aple	1		
		successry	1			apple	1		

2-9		2-10		2-11				2-12	
invitation		supporter		argument				emphasis	
invite		support		argue				emphasize	
response	*n*	response	*n*	response	*n*			response	*n*
able to	1	argumenting	1	agree	1	arguy	2	amphas	1
do	1	arguming	1	angry	2	argy	1	amphasisy	1
in	1	arguy	1	angumen	1	arou	1	argum	1
innvitat	1	do	1	anguy	1	arouaotiong	1	argumenting	1
inter	1	introduce	1	ar	1	arugmate	1	do	1
introduce	1	invitationing	1	arg	2	augum	1	emblasize	1
invait	1	saprt	1	arge	3	do	1	embrece	1
invate	2	sport	3	argent	1	emphas	1	empahas	1
inveit	1	spport	1	arger	1	grgum	2	emph	6
invent	7	supart	1	argme	2	gument	1	empha	19
invente	8	suport	8	argmer	2	invitationing	1	emphace	1
invid	9	supot	1	argree	1	orgum	1	emphaes	1
invie	10	supper	1	argu	49	orgume	1	emphaese	1
invint	11	suppoortering	1	argud	1	supporting	1	emphair	2
invint	12	suppor	6	arguee	1	wereng	1	emphais	1
invisit	1	supporce	1	arguer	1	Total	410	emphaisliy	1
invit	127	supporice	1	argul	1			emphaiz	1
invita	19	supportce	1	argum	138			emphal	1
invitace	6	supporte	16	argumait	1			emphans	1
invitad	2	supporteing	1	argumation	1			emphanse	1
invitaed	1	supporter	8	argume	56			emphar	1
invitaes	1	supporterful	1	argumeiter	1			empharass	1
invital	1	supportering	3	argumen	16			emphart	1
invitaly	1	supporters	3	argumence	5			emphas	158
invitance	2	supportery	1	argumend	2			emphase	37
invitant	1	supportest	2	argumenge	1			emphases	2
invitar	3	supporting	9	argumennted	1			emphash	1
invitase	1	supportion	1	argumens	2			emphasi	2
invitast	1	supportment	1	argument	18			emphasible	2
invitat	39	supportted	1	argumentable	1			emphasicary	1
invitate	68	supporty	1	argumental	1			emphasiges	1
invitater	1	suppose	5	argumentary	1			emphasije	1
invitati	3	suppot	9	argumentce	2			emphasin	1
invitatice	3	suppote	1	argumente	1			emphasing	3
invitating	1	supprt	3	argumented	1			emphasis	11
invitatioer	1	suppurt	1	argumenter	14			emphasisable	1
invitation	8	Total	99	argumentes	1			emphasisation	1
invitational	3			argumentest	1			emphasisce	2
invitationary	2			argumentful	1			emphasised	1
invitationce	2			argumenting	11			emphasiser	10
invitationder	1			argumention	4			emphasising	6
invitationed	2			argumently	3			emphasision	1
invitationer	9			argumentns	1			emphasism	1
invitationes	2			argumentry	1			emphasisness	1
invitationing	8			arguments	2			emphasisry	1
invitationned	1			argumenty	2			emphasisser	1
invitationning	2			argumer	5			emphasist	5
invitations	3			argumese	1			emphasister	1
invitatly	1			argumest	1			emphasistion	1
invitatony	1			argumester	1			emphasisu	1
invitatve	1			argumet	2			emphasisum	1
invited	3			argumetily	1			emphasisy	3
inviting	1			argumeting	1			emphasite	1
invitry	1			argumt	1			emphasly	1
invituce	1			argumu	1			emphass	2
invitutison	1			argumy	4			emphast	2
invtat	1			argun	2			emphasy	31
itvite	1			argunt	1			emphasys	1
ivitations	1			argur	7			emphat	2
ivite	2			argure	1			emphath	1
supporting	1			argury	2			emphatise	1
surport	1			argust	1			emphatize	1
Total	410							emphay	2

2-12		2-13				2-14			
emphasis		independence				religion			
emphasize		independent				religious			
response	*n*	response	*n*			response	*n*		
emphe	2	amphasising	1	independencly	1	box	1	religionned	1
emphis	1	day	4	independencry	1	emphasising	1	religionning	1
emphissi	1	depend	1	independenction	1	re	1	religionoce	1
emphsise	1	do	1	independency	9	reale	1	religionous	3
emphy	1	eindependenceful	1	independend	1	realise	1	religionry	1
empla	1	emphasising	1	independender	1	region	1	religions	8
emplas	2	i	1	independenful	3	rel	1	religiont	4
emply	1	idependently	1	independenietd	1	relax	1	religiontion	2
empty	3	idpependent	1	independenly	2	relay	1	religionty	1
enemphasis	1	imdepend	1	independensful	1	release	1	religiony	9
enphas	1	imdependent	1	independenter	1	relgionful	1	religior	2
enphasism	1	indendenceful	2	independention	5	reliable	1	religiory	1
enpty	1	indenpendant	1	independently	3	relig	28	religiotion	2
eomphas	1	indenpendent	1	independenty	1	religal	1	religioust	1
imphas	1	indeped	1	independeny	2	religation	1	religioy	1
phasis	1	indepedencer	1	independer	1	relige	12	religis	1
supportering	1	indepedencing	1	independet	1	religeful	1	religit	1
Total	371	indepedention	1	independete	1	religer	5	religitial	1
		indepen	1	independful	2	religetation	1	religition	3
		indepence	5	independion	2	religetion	1	religitly	1
		independ	73	independment	1	religfull	1	religliorer	1
		independanceful	1	independncer	1	religh	1	religly	2
		independancely	1	independnt	1	relighenen	1	religonce	1
		independant	2	independtion	1	relight	3	religoner	1
		independe	13	independur	1	religi	6	religous	1
		independecefull	1	independy	2	religial	1	religry	1
		independecer	1	indepens	1	religic	1	religt	1
		independeceses	1	indepent	8	religie	1	religthion	1
		independecial	1	indepention	3	religiful	7	religtion	2
		independecly	1	indepnd	1	religiger	1	religy	12
		independed	1	indepndeful	1	religily	1	reliig	1
		independeful	3	indepndencely	1	religin	1	relio	1
		independely	1	indepndent	2	religination	1	relious	1
		independen	6	indespend	1	religinful	1	relise	1
		independenal	2	indipendent	1	religing	1	relition	1
		independenation	1	indpende	1	religinnful	1	relization	1
		independencal	2	intordence	1	religio	1	relize	4
		independencaly	1	Total	283	religioce	1	relizey	1
		independence	7			religioful	3	rely	3
		independenceally	1			religioly	3	resligion	1
		independenced	8			religion	22	riligious	1
		independenceday	3			religionable	3	rily	1
		independenceed	1			religional	50	rimit	1
		independenceful	16			religionally	1	sea	1
		independenceing	1			religionaly	18	sexy	1
		independencel	1			religionarity	1	the	1
		independencely	8			religionary	5	Total	359
		independencer	12			religionat	1		
		independencery	1			religionation	6		
		independences	2			religionay	2		
		independencese	1			religionce	1		
		independencess	1			religione	1		
		independencet	1			religioned	3		
		independencety	1			religioner	10		
		independenceu	1			religiones	2		
		independencful	2			religioness	3		
		independencfull	1			religioney	2		
		independencial	1			religionful	21		
		independencian	1			religionily	1		
		independencily	2			religioning	4		
		independencing	1			religionliy	1		
		independencion	1			religionly	13		

2-15				2-16		2-17		2-18	
similarly				origin		warmth		specialist	
similar				original		warm		special	
response	*n*			response	*n*	response	*n*	response	*n*
oh	1	simill	1	ariginal	1	aormthful	1	dirtd	1
origing	1	simillar	1	crisin	1	marmer	1	especially	1
originy	1	simlly	1	hise	1	originalty	1	especialy	2
simail	1	simily	17	onry	1	specialistest	1	spcialful	1
simaly	1	siming	1	opens	1	w	1	speal	1
simil	25	simiry	1	org	1	wamthion	1	specal	2
simila	5	simlar	1	orgenize	1	warker	1	spece	1
similaful	4	simle	1	orgin	1	warmal	4	specia	1
similal	11	simlilar	1	orginal	1	warmation	1	specialant	1
similality	1	simly	1	orginition	1	warmde	1	specialaze	1
similarable	1	simmilar	1	orig	9	warmer	23	speciale	3
similaral	1	simple	2	origan	2	warmful	12	specialer	2
similarce	1	simply	1	orige	1	warmfull	1	specialful	5
similare	1	siuily	1	origen	1	warming	4	specialing	1
similarful	9	smail	3	origi	1	warmless	1	specialism	2
similarier	1	smil	2	origiful	1	warmly	5	specialist	5
similariful	2	smilar	1	origilty	1	warmment	1	specialisten	2
similarity	4	smile	8	origiment	1	warmmer	1	specialistence	1
similarl	22	smille	1	origin	5	warmmy	1	specialister	12
similarlal	2	smily	4	originabul	1	warmn	1	specialistful	4
similarlat	1	Total	242	originale	2	warmnal	1	specialistic	1
similarlation	1	simle	1	originaler	1	warmness	2	specialistill	1
similarlful	2	simlilar	1	originality	5	warmress	1	specialistly	1
similarlical	1	simly	1	originally	10	warmt	1	specialistn	1
similarlieal	1	simmilar	1	originalrty	1	warmth	9	specialiston	1
similarlied	2	simple	2	originalty	3	warmthal	1	specialists	1
similarlier	2	simply	1	originaly	29	warmthce	1	specialisty	3
similarlies	2	siuily	1	originar	1	warmthed	1	specialiter	2
similarliess	1	smail	3	originariter	1	warmthely	1	speciality	13
similarliest	1	smil	2	originarity	6	warmther	12	specialize	1
similarliful	1	smilar	1	originartion	1	warmthese	1	speciall	1
similarlily	1	smile	8	originartty	1	warmthful	12	specially	18
similarling	2	smille	1	originary	22	warmthial	1	specialry	1
similarlison	1	smily	4	originate	29	warmthin	1	specialty	2
similarlition	1	Total	242	originathion	1	warmthing	2	specialy	36
similarlity	11			origination	5	warmthiy	1	specianist	1
similarll	1			origince	1	warmthly	6	speciarl	1
similarlly	2			origind	1	warmthness	1	speciary	1
similarlous	1			origine	1	warmtht	1	specicial	1
similarlry	1			origined	2	warmthy	9	specil	1
similarlty	2			originer	7	warmtion	2	speial	2
similarly	9			originful	10	warmtit	1	spenciale	1
similarlyed	2			origing	1	warmty	3	spestsxz	1
similarlyer	5	orugan	1	origingin	1	warmy	5	Total	141
similarlyes	1	orugun	1	origininal	1	warth	1		
similarlyful	7	similation	1	origining	1	worm	1		
similarlying	1	start	1	originll	1	wormely	1		
similarlyty	4	warmth	1	originly	6	wormer	1		
similart	2	Total	212	originoly	1	wormthest	1		
similarty	1			origins	3	wormthing	1		
similary	2			origintion	1	Total	145		
similaryed	1			originty	2				
similarzation	1			originul	1				
similasty	1			originy	6				
similate	1			origion	1				
similation	7			origiral	2				
similau	1			origist	1				
similay	1			orignalty	1				
simile	5			origun	1				
similer	3			oriiginal	1				
similful	3			oringe	1				
similing	1			orional	1				

2-19		2-20	
dirt		end	
dirty		endless	
response	*n*	response	*n*
air	1	and	1
dart	1	auf	1
darty	1	disend	1
dead	3	eddry	1
deat	1	empty	2
death	1	end	30
dent	1	endaly	1
der	2	enddtion	1
dey	1	ended	11
di	1	ender	20
die	3	enderful	1
dir	11	endery	1
dird	1	endes	1
dire	2	endess	3
direct	1	endest	2
dirfly	1	endful	5
dirion	1	endfull	1
dirss	1	endial	1
dirsty	1	ending	37
dirt	17	endist	1
dirtal	6	enditon	1
dirtd	1	endles	5
dirteful	1	endlily	1
dirtely	1	endlles	3
dirten	2	endlless	1
dirter	19	endly	17
dirtes	1	endnal	1
dirtess	2	endness	27
dirtest	6	endog	1
dirtful	30	endr	2
dirtfull	1	endral	1
dirtg	1	endres	15
dirth	2	endress	78
dirtial	1	endrest	1
dirtier	1	endroll	1
dirtiful	2	endrs	1
dirtil	1	endry	1
dirting	7	ends	4
dirtion	10	endth	1
dirtional	1	endty	1
dirtionaly	1	endur	1
dirtir	1	endure	2
dirtist	7	endy	2
dirtive	1	finish	1
dirtlly	1	forever	2
dirtly	14	imending	1
dirtment	1	last	1
dirtness	2	less	1
dirtry	1	lever	1
dirts	3	never	3
dirtst	1	neverend	1
dirttion	1	nevery	1
dirttly	1	next	1
dirtty	1	no	2
diry	1	noend	4
diy	1	not	3
drity	1	notend	1
endless	1	nothing	2
Total	188	of	1
		roll	1
		s	1
		the	4
Total	324	unend	2

Appendix 9.2

All the responses in the Antonym Subsection from the participants and whether they were scored as correct or incorrect

3-1			3-2			3-3					
open			high			wet					
closed	10	rater	low	10	rater	dry	10	rater		10	rater
cloce	0		bottom	0	1	bry	0		thirth	0	
cloes	0		boun	0		c	0		thithty	0	
cloese	0		dark	0	1	carm	0	1	warm	0	
clone	0		deep	1		cold	0		wash	0	
cloose	0		down	0	1	cool	0		wat	0	
clors	0		how	0		cozy	0	1	wear	0	
clorse	0		jnnre	0	1	cry	0		wettish	0	
close	1		junior	1		dead	0	1	what	0	
closed	1		law	1		ded	0		wont	0	
closs	0		lew	0	1	desert	0	1	yet	0	
cloth	0		light	0	1	desurt	0		wont	0	
clothe	0		loo	0		die	0		yet	0	
clould	0		lor	0		disert	0		wont	0	
clous	0		lose	0		dlai	0		yet	0	
clouse	0		lote	0		dlay	0		wont	0	
clows	0		lour	0		dly	0				
cluse	0		low	1		dorai	0				
cols	0		lowe	0		drai	0				
colse	0		megachanon	0		draw	0				
corse	0		onder	0	1	dray	0				
cous	0		raw	0	1	drey	0				
coze	0		rou	0		drg	0				
croose	0		row	0		dri	0				
crorse	0		shoot	0		dro	0				
cros	0		short	1		drul	0				
crose	0		shote	0		dry	1				
cross	0	1	shout	0		east	0				
crosu	0		showt	0		est	0				
crous	0		sitiher	0		fat	0				
crouse	0	1	smaill	0		get	0				
crouth	0		smale	0	1	go	0				
crows	0		small	0	1	hot	0	1			
crowse	0		smell	0		hyt	0				
get	0		smorl	0		imwet	0				
krose	0		t	0		late	0				
krousi	0		tall	0		lay	0				
kuods	0		under	0	1	light	0				
qurouse	0					m	0				
sarut	0					ment	0				
shatt	0					mild	0	1			
shut	1					mud	0	1			
shut down	1					not	0				
slose	0					out	0				
						pant	0				
						right	0				
						sat	0				
						sent	0				
						shat	0				
						smart	0	1			
						suit	0				
						swet	0				
						tears	0				
						tesh	0				
						tew	0				
						tharsty	0				
						thersty	0				
						therth	0				
						thirkly	0				
						thirs	0				
						thirst	0	1			
						thirsty	0	1			

Note. response = response from test takers. 10 = Whether it was scored as correct (1) or incorrect (0). rater = judged by the rater.

3-4			3-5			3-6			3-7		
alive, dead	10		independence, dependence	10	rater	start, end	10	rater	nonsense, sense	10	rater
adead	0		accord	0		alive	0		absense	0	
again	0		anviliebavor	0		arrive	0	1	anysone	0	
alife	0		attachment	1	1	back	0	1	common	0	
aliven	0		colonie	0		be	0		commonsense	1	1
aloo	0		colony	0	1	big	0		comonsense	0	
anow	0		conquer	0	1	cole	0		consense	0	
arrive	0		control	1	1	end	1		denger	0	
asleep	0		controled	0		final	0	1	easy	0	
bad	0		dangerous	0		finesh	0		efficient	0	
befor	0		day	0		finis	1		esnesnon	0	
believe	0		deindependence	0		finish	1		eye	0	
bie	0		dendence	0		finnish	0		fulsense	0	
breabe	0		denpendnce	0		fish	0		geneus	0	
brive	0		deny	0		fisish	0		good	0	1
come	0		depend	0		gale	0		good sence	0	
dai	0		dependant	0		gall	0		goodsense	0	1
daid	1		depende	0		galu	0		havesense	0	
day	0		dependence	1		garl	0		highsense	0	
deab	0		dependense	0		goal	0	1	humor	0	1
dead	1		dependent	0	1	gol	0		humoras	0	
death	0		destroy	0		gole	0		imo	0	
deatj	0		develop	0		goll	0		innonsense	0	
ded	0		disindependence	0		golole	0		inonsense	0	
dei	0		encaged	0		golu	0		inosence	0	
deid	1		end	0		gool	0		inosense	0	
deperture	0		exdependence	0		gorl	0		insense	0	
desalive	0		expensive	0		goru	0		interesting	0	1
deslive	0		free	0	1	goul	0		just	0	
dess	0		gather	0	1	last	0		mean	0	1
desth	0		govament	0		long	0		meaning	0	1
dete	0		groop	0	1	loss	0		meaningful	1	1
deth	0		happy	0		over	0		naysence	0	
did	0		important	0		quit	1	1	nice	0	
dide	0		indenendent	0		reach	0	1	nicesennse	0	
die	0	1	inko	0		smart	0		nicesense	0	
died	0	1	inpented	0		stay	0	1	nonno	0	
disappear	0	1	inper	0		stop	1	1	nonsens	0	
diy	0		insilens	0		stopee	0		not nonsense	0	
do	0		leave	0		stopp	0		nsense	0	
down	0	1	lestar	0					onani	0	
duy	0		linite	0					onsense	0	
erabur	0		no	0					pesose	0	
erila	0		nonindependence	1					ploblem	0	
get	0		ortmutic	0					possible	0	
go	0	1	outdependence	0					presense	0	
got	0		outpendence	0					problem	0	
inalive	0		pendence	0					profitable	0	
late	0	1	rely	0	1				puiet	0	
leave	0	1	rule	0	1				quicly	0	
left	0	1	ruled	0	1				reach	0	
lie	0	1	start	0					reason	1	
live	0		stop	0					reasonable	1	1
move	0		studio	0					right	0	1
not alive	0		stur	0					sadry	0	
olive	0		subject	0					sence	1	
reave	0		together	0	1				senceful	0	
sleep	0		undepedence	0					sencefully	0	
start	0		undependence	0					sensable	0	
stop	0	1	unindependence	0					sense	1	
vive	0		unit	0					senseable	0	
went	0	1	united	0	1				senseble	0	
wrong	0		wor	0					senseful	0	
									sensely	0	

3-7			3-8						3-10		
nonsense			happily						tragedy		
sense	10	rater	sadly, unhappily	10	rater				comedy	10	rater
sensflu	0		am happily	0		unbappily	0		comedy	1	
sensible	1	1	anger	0		unconfortable	0		comfortness	0	
sensintive	0		angly	0		unfartunately	0		comic	0	1
sensitive	0	1	angrey	0		unfortually	0		comical	0	
sensive	0		angrry	0		unfortunately	1		comicaly	0	
significant	0		angry	0		unfortunatly	0		commedy	0	
smart	0		anhappily	0		unfortunently	0		distragedy	0	
umor	0		anhappy	0		ungry	0		distrange	0	
upsense	0		annoy	0		unhaippily	0		doraemon	0	
usual	0		b	0		unhappily	1		emotion	0	
valid	1	1	bad	0	1	unhappiness	0		epicouse	0	
wonderful	0	1	baddly	0		unhappy	0		fortune	0	
worthful	0		badily	0		unluacky	0		happiness	1	
yesense	0		badlly	0		unlucky	0		imtragedy	0	
yumoa	0		badly	0	1	uphappy	0		intragedy	0	
			batlly	0		warm	0		keico	0	
			bed	0		worse	0		kind	0	
			blue	0		worst	0		lake	0	
			bluely	0	1	wrong	0		luckey	0	
			bully	0					main-stream	0	
			cay	0		sick	0		not tragedy	0	
			claud	0		sorrowly	0		optimism	0	1
			cly	0		sred	0		parody	0	
			cry	0		stamacece	0		peace	0	1
			darcky	0		suddenly	0		pleasant	0	
			dark	0		terrible	0	1	pleasure	1	1
			disappoint	0		unahappiy	0		pre	0	
			dishappily	0					preasure	0	
			dishappiy	0					saking	0	
			dusty	0					seer	0	
			ensiter	0					start	0	
			gaoly	0					strategy	0	
			happy	0					techolgy	0	
			happyness	0					tecnorogy	0	
			hardly	0					telivle	0	
			hevey	0					track	0	
			inhappily	0					traveluy	0	
			liaty	0					untragedy	0	
			nohappily	0					wonderful	0	
			nonhapily	0					ydegart	0	
			pay	0							
			poor	0							
			poorly	0	1						
			rad	0							
			sad	0							
			sadanly	0							
			saddenly	0							
			saddily	0							
			saddiny	0							
			saddly	0							
			sadenly	0							
			sadenry	0							
			sadilly	0							
			sadily	0							
			sadly	1							
			sadness	0							
			sadny	0							
			said	0							
			saidly	0							
			saidy	0							
			sarad	0							
			sard	0							
			sic	0							

3-11						3-12			3-13		
supporter						able			export		
opponent	10	rater				unskillful	10	rater	import	10	rater
adversary	1		objecter	0		able	0		albiver	0	
advice	0		offence	0		abler	0		amatuer	0	
against	0		offender	0	1	ablety	0		anexport	0	
againster	0		offer	0		amable	0		begginer	0	
alone	0	1	opponent	1		anable	0		begin	0	
anchi	0		oppornent	0		be not able	0		beginer	0	
antch	0		opportment	0		betu	0		beginner	0	
anti-supporter	0		opposer	1	1	binable	0		biginer	0	
ap	0		opposi	0		blew	0		can't	0	
asreat	0		opposider	0		can	0		disport	0	
athreat	0		opposite	0		cannot	0	1	earport	0	
attacker	1	1	oppositer	0		can't	0	1	easy	0	
attaker	0		opposition	0		did	0		enexport	0	
away	0		oppsiter	0		die	0		enport	0	
barrier	0		others	0		disable	0		eunexport	0	
booing	0		pappte	0		done	0		exam	0	
buinger	0		people	0		don't	0		exchange	0	1
ciritic	0		pepor	0		else	0		exit	0	
claimer	0	1	player	0	1	enable	0		experiment	0	
cocksucker	0		preventer	0	1	glad	0		express	0	
consumer	0		prodecer	0		have	0		exproiment	0	
corch	0		producer	0	1	hot	0		exqort	0	
counter	0		protestant	0	1	impossible	0	1	farster	0	
critcizer	0		pupil	0		inable	0		fort	0	
critic	1	1	reader	0		incapable	0	1	hear	0	
criticism	0		regedent	0		inpossible	0		heriport	0	
criticizer	0	1	regidenter	0		may	0		imexport	0	
custoner	0		registance	0		must	0		impart	0	1
decent	0		register	0		nable	0		import	1	1
destroyer	1	1	remember	0		never	0		include	0	
desupporter	0		sea	0		no able	0		inexport	0	
dissident	1	1	son	0		noable	0		inport	0	
dissuporter	0		supper	0		nonable	0		kit	0	
dissupporter	0		traugher	0		noshiled	0		level down	0	
effort	0		troubler	0	1	not	0		nonexport	0	
eimy	0		unsupport	0		not able	0		notexport	0	
eminem	0		unsupporter	0		notable	0		novice	0	
enemg	0		untisupporter	0		only	0		other	0	
enemy	0	1	walker	0		out	0		port	0	
enermy	0					poor	0	1	poseport	0	
fighter	0	1				possible	0		strange	0	
fooligan	0					sent	0		transportation	0	1
foolign	0					shoult	0		unexport	0	
fooligun	0					so	0		unko	0	
foorigun	0					to	0		xeport	0	
forligan	0					umable	0				
forllower	0					unable	1	1			
freeter	0					unabll	0				
fuligan	0					under	0				
furigah	0					vible	0				
furigunn	0					won't	0				
furiun	0										
furligan	0										
fuurigan	0		manager	0							
garter	0		myself	0							
gast	0		nonsupporter	1	1						
heavy	0		non-supporter	1							
help	0		nosupporter	0							
ignore	0		not	0							
irresponsibility	0		notsupporter	0							
keeper	0	1	manager	0							
legister	0										
manage	0										

3-14			3-15						3-16		
buy			minor						beautiful		
sell	10	rater	major	10	rater				ugly	10	rater
bought	0		adult	1		poplur	0		agly	0	
buy	0		agly	0		popular	0	1	agony	0	
cell	0		ask	0		populer	0		agray	0	
exchange	0	1	baka	0		populler	0		agree	0	
free	0		beginer	0		positiv	0		agry	0	
g	0		big	1	1	prior	0		agury	0	
girl	0		bigger	0		pro	0		angree	0	
give	0	1	billon	0		professional	0		angry	0	
have	0		citizen	0	1	prus	0		awful	1	
lent	0	1	common	0		right	0		bad	1	
nobuy	0		dirty	0		rinor	0		badful	0	
pay	0	1	disminor	0		ronin	0		badness	0	
pey	0		eager	0		sell	0		bairt	0	
puy	0		famous	0	1	special	0	1	bark	0	
reason	0		famouse	0		specialist	0		bary	0	
sail	0		good	0		street	0		beaty	0	
sale	0		huge	1	1	unminor	0		birty	0	
sall	0		inferior	0	1				bout	0	
say	0		junior	0	1				buss	0	
seal	0		large	1	1				clean	0	
seel	0		larger	0					cute	0	1
sel	0		lea	0					d	0	
seld	0		ma	0					dark	0	1
sele	0		mager	0					dart	0	
sell	1		magher	0					dartiful	0	
send	0	1	magor	0					darty	0	
sent	0		mainor	0					dast	0	
seru	0		majar	0					daty	0	
set	0		majer	0					dearty	0	
smell	0		majior	0					deatiy	0	
snall	0		major	1					deaty	0	
sold	0		majority	0					derty	0	
sole	0		majorty	0					diary	0	
solt	0		majour	0					diaty	0	
soud	0		manior	0					dirst	0	
sought	0		many	0	1				dirsty	0	
sould	0		maxium	0					dirstyful	0	
spend	0	1	mazo	0					dirt	0	
take	0	1	meajure	0					dirtful	0	
we	0		mean	0					dirtiful	0	
			measure	0					dirtir	0	
			meger	0					dirtty	0	
			megger	0					dirty	0	1
			megior	0					dity	0	
			megir	0					diuty	0	
			megour	0					dreadful	0	1
			meguer	0					drity	0	
			meijor	0					dud	0	
			mejar	0					dull	0	1
			mejer	0					dunty	0	
			mejor	0					durty	0	
			mejour	0					dusty	0	1
			mejur	0					dutiy	0	
najor	0		menen	0					duty	0	
people	0		mesar	0		invit	0		easy	0	
plass	0		mesur	0		kill	0		egly	0	
plus	0		mesure	0		kitanai	0		filthful	0	
pluser	0		milionere	0		kkussa	0		foul	1	1
poor	0	1	million	0		love	0		gimme	0	
poorer	0		mine	0		mess	0		grange	0	
			minolity	0		miror	0		ha	0	
			mixmor	0					hatelly	0	
			monter	0					horrible	1	1

3-16			3-17			3-18					
beautiful			quick			hungry					
ugly	10	rater	slow	10	rater	sated	10	rater			
miserable	0	1	answer	0		abundant	0	1	serfty	0	
nonbeautiful	0		blue	0		angry	0		sick	0	
not beautiful	0		cate	0		anpty	0		sleepy	0	
peaple	0		dark	0		beg	0		small	0	
pity	0		deray	0		comfort	0		smily	0	
poor	0	1	difficult	0		content	1	1	stuffed	1	1
simpul	0		dificult	0		contented	1	1	sufficient	0	1
slow	0		dull	1		deshungry	0		surtisfid	0	
smell	0		ealey	0		eager	0		sutisfying	0	
spupid	0		fast	0		efficient	0		temple	0	
strange	0		hard	0		embarrassed	0		thirsty	0	1
stupic	0		j	0		emper	0		thirthry	0	
talivle	0		late	1	1	empty	0		ungry	0	
terreble	0		lately	0		enmpy	0		unhungry	0	
terrible	1		later	0		enogh	0		up	0	
thouhigy	0		law	0		enoght	0		wealth	0	
u	0		loose	1	1	enougf	0				
ugly	1		lose	0		enough	0	1			
ugree	0		low	0		enptly	0				
ugry	0		lrou	0		enpty	0				
umbeatiful	0		noise	0		fall	0				
unbeautiful	1		noisy	0		fat	0				
undeautiful	0		nose	0		fay	0				
ungly	0		pull	0		fell	0				
ungry	0		push	0		few	0				
urge	0		quickless	0		fill	0	1			
wonderful	0		quiet	0		filled	0	1			
			rebirth	0		fli	0				
			sell	0		food	0				
			shroow	0		fool	0				
			slaw	0		for	0				
			slew	0		fore	0				
			sllory	0		ful	0				
			sllow	0		full	1				
			sllwor	0		fully	0				
			slooly	0		furt	0				
			sloory	0		fut	0				
			sloou	0		fuu	0				
			slor	0		glad	0				
			slorly	0		good	0				
			slory	0		happiy	0				
			slou	0		happy	0	1			
			slouwy	0		hull	0				
			slow	1		implace	0				
			slowe	0		interestinged	0				
			slowl	0		ippai	0				
			slowliy	0		manhappy	0				
			slowlly	0		manpuku	0				
			slowly	1	1	many	0	1			
			slowry	0		meet	0				
			slowy	0		met	0				
			slwol	0		much	0	1			
			snowly	0		nonhungry	0				
			sroly	0		not hungry	0		prenty	0	
			srorry	0		occupied	0	1	pull	0	
			srory	0		p	0		pure	0	
			sroue	0		panion	0		satisfactory	1	1
			srouly	0		pl	0		satisfied	1	
			srow	0		planty	0		satisfy	0	
			srowly	0		plenty	0	1	sentury	0	
throw	0		stowry	0		poor	0				
uma	0		streat	0		pow	0				
unquick	0		surow	0		pranty	0				

Appendix 9.3

All the responses in the Collocation Subsection from the participants and whether they were scored as correct or incorrect

4-1									4-2		
() dog	10	rater							() cancer	10	rater
a	0		Japanese	1	1	wild	1		aeveloped	0	
angry	1	1	killed	0	1	wise	1	1	aha	0	
awful	1	1	kind	1	1	write	0		alive	0	1
baby	0	1	large	1	1	yellow	1	1	all	0	1
bad(×)	0		learge	0		young	1	1	animal	0	1
bark	0	1	like	0	1	your	0		anti	0	
barking	1	1	little	1	1				awful	0	1
beautifle	0		lovely	1	1				bad(×)	0	
beautifu	0		mad	1					badly	0	1
beautiful	1	1	male	1	1				be	0	
beautyful	0		mud	0	1				beautiful	0	1
biantiful	0		my(×)	0					become	0	
big(×)	0		new	1	1				big(×)	0	
black	1	1	nice(×)	0					bitterly	0	
blind	1	1	noisy	1					black	0	1
blue	0	1	noizy	0					blood	1	1
body	0	1	one	1					botom	0	
brack	0	1	onother	0					brain	1	1
brave	1	1	our	0					building	0	1
broun	0		parson	0	1				but	0	
brown	1	1	perfect	1	1				came	0	
brute	0	1	perty	0					car	0	
bull	1	1	pick	0					care	0	1
buy	0		pink	0	1				carefully	0	
cate	0		pleatty	0					catch	0	1
child	0	1	pletty	0					cause	0	1
cite	0		plety	0					chest	1	1
clean	1	1	plitty	0					cleam	0	
cleaver	0		poor	1	1				clean	0	1
clever	1	1	powerful	1	1				clearn	0	1
cool	1	1	prally	0					coll	0	
crazg	0		preaty	0					cool	0	1
crazy	1	1	preety	0					cute	0	1
curt	0	1	prettly	0					dadly	0	
cut	0	1	pretty	1	1				dangeous	1	1
cute	1	1	prety	0					danger	0	1
cutie	0	1	pritty	1	1				dangerours	0	
darty	0		prity	0					dangerous	1	1
dead	1		prytty	0					dead	0	1
dengerous	0		puritty	0					deadful	0	
died	0	1	real	1	1				deep	0	1
dig	0	1	red	1	1				denger	0	
dirty	1		resaroir	0					dengerous	0	
domestic	1		robot	1	1				dergerous	0	
faithful	1		run	0					desease	0	
famous	1	1	sanlty	0					desperate	0	1
fast	1	1	sarage	0					died	0	1
fat	1	1	scared	1	1				different	1	1
favorite	1	1	scary	1	1				difficult	0	1
fierce	1		sitting	1	1				difficulty	0	1
fired	0	1	sleepy	1	1				dirty	0	1
first	1	1	small(×)	0					disavoidable	0	
fool	0	1	smart	1	1				disease	0	1
foolish	1	1	special	1	1				dog	0	1
ful	0		sport	0	1				dreadful	1	1
funky	1	1	spotted	1	1				eager	0	1
gay	0	1	sriping	0					ear	1	1
golden	1	1	strange	1	1				early	0	1
great(×)	0		strong	1	1				easy	0	1
happily	0		super	0	1				end	0	1
happy	1	1	sweet	1	1				endless	0	1

(Appendix 9.3 continues)

(Appendix 9.3 continued)

4-1									4-2		
() dog	10	rater							() cancer	10	rater
harmless	1	1	terrible	1	1				famous	1	1
heavy	1	1	this	0					fatal	1	1
honest	0	1	Tom's	1	1				father's	0	1
hot	1		unhappy	1	1				fatul	0	
hot/curt	1		useful	1	1				favorite	0	1
human	0	1	wanderful	0					fearful	1	1
hungry	1	1	warm	0	1				find	0	1
impossible	0	1	weak	1	1				first	0	1
intensive	0	1	wet	1	1				from	0	
interesting	1	1	white	1	1				fund	0	1

Note. response = response from test takers. 10 = Whether it was scored as correct (1) or incorrect (0). rater = judged by the rater.
(×) = Incorrect because the instruction said, "Do not write *good, better, best, bad, nice, wonderful, great, big, small,* or *old*."

4-2						4-3					
() cancer	10	rater				() art	10	rater			
good(×)	0		slight	0	1	abstract	1		happy	0	1
great(×)	0		small(×)	0		abstuct	0		heill	0	
gurde	0		smell	0	1	air	0	1	home	0	1
happy	0	1	soon	0		amimal	0	1	huge	1	1
hard	0	1	sorry	0	1	an	0		ice	0	1
hardly	0		sounds	0	1	anciant	0		imperishable	1	1
harmful	1	1	special	0	1	ascie	0		important	1	
harmfull	0		sport	0	1	ascli	0		impressive	1	1
hate	0	1	sports	0	1	bad(×)	0		industrial	1	1
have	0	1	sppedy	0		bag	0	1	interested	0	
haveay	0		stomac	0		beatiful	0		interesting	1	1
havy	0		stomach	1		beauteful	0		international	1	1
head	1	1	stomache	0		beautiful	1	1	Japanese	1	
heart	1	1	stomuch	0		beautifule	0		just	1	1
heavy	0	1	strange	1	1	beauty	0	1	large	0	1
help	0	1	strong	0	1	beawtiful	0		liberal	1	
helpless	1	1	sun's	0	1	beutiflu	0		life	0	1
high	0	1	super	0	1	beutifu	0		like	0	1
home	0	1	terebble	0		beutiful	0		line	1	1
hopeless	1	1	terible	0		big(×)	0		maddan	0	
horrible	1	1	terminal	1		body	1	1	moden	0	
huge	0	1	terrible	1	1	box	0	1	modern	1	1
human	0	1	tongue	1	1	brack	0	1	morden	0	
hurt	0	1	trade	0	1	burtiful	0		Mozart's	1	
ill	0	1	tragic	1	1	butiful	0		music	0	1
important	0	1	tragical	0	1	butifulu	0		my()	0	
kidny	0		tribble	0	1	buttful	0		nail	1	1
kind	0	1	unuseful	0		c	0		neil	0	
lamb	0	1	very	0	1	car	0	1	neile	0	
large	0	1	vit	0		cat's	0	1	new	1	
last	0	1	warm	0	1	cheap	1	1	nice(×)	0	
laung	0		weak	0	1	cheep	0		oil	0	1
learge	0		who	0		classical	1		old(×)	0	
left	0	1	wide	0	1	clean	0	1	original	1	
light	0	1	wonderful(×)	0		clese	0		outstanding	1	1
little	0	1	world	0	1	conseptual	0		paint	0	1
love	0	1	worse	0	1	cool	0		paper	1	1
lumb	0		worst	0	1	cultural	0	1	pen	1	1
lung	1		worth	0	1	cute	1	1	perfect	1	1
maqueike	0		wrong	0	1	dark	0	1	physical	1	1
miclo	0					deautiful	0		picter	0	1
misarable	0	1				des	1	1	picture	0	1
more	1	1				design	0	1	political	1	1
motion	0	1				difficult	1		poor	0	1
movable	0	1				digital	1	1	pop	1	

(Appendix 9.3 continues)

(Appendix 9.3 continued)

4-2						4-3					
() cancer	10	rater				() art	10	rater			
my()	0					dirty	1	1	popular	1	
native	0	1				dorowing	0		prease	0	
new	1	1				dream	0	1	pretty	1	1
nice(×)	0					enjoy	0		priceless	1	1
north	0	1				excellent	1	1	progressive	1	1
office	0	1				expensive	1	1	prtty	0	
operated	0	1				exserent	0		recent	1	1
perfect	0	1				famous	1	1	religious	1	
poor	0	1				famouse	0		rock	1	1
pretty	0	1				fantasic	0		runesans	0	
quick	0	1				fantastic	1	1	sadly	0	
red	0	1				faverit	0		school	1	1
remarkable	0	1				favorite	1	1	sclipt	0	
remediable	1	1				fine	1		simple	1	1
right	0	1				first	0	1	single	0	1
round	0	1				flower	1	1	sky	0	1
s	0					fun	1	1	smile	0	1
sad	0	1				funny	1	1	smole	0	
searios	0					funtastic	0		snow	1	1
serious	1	1				good(×)	0		sock	0	1
serius	0					graphic	0		special	1	1
severe	1	1				gray	0	1	specioul	0	
she	0					great(×)	0		speical	0	
sick	0	1				H	0		spical	0	
silious	0					hair	0	1	spritual	0	
skin	1					happiy	0		st	0	

4-3			4-4						4-5		
() art	10	rater	() birthday	10	rater				() tragedy	10	rater
stone	1	1	16th	1		today is	0		air	1	1
strange	1	1	17th	1		tomorrow	0	1	all	0	1
study	0	1	air	0	1	Tom's	1		ancient	1	1
super	0	1	annual	0	1	un (= one)	1		another	1	1
take	0		baby's	1	1	uncle	0		attructive	0	
technical	1		bad(×)	0		unforgettable	1	1	awful	1	
terrible	1	1	beautiful	0	1	unhappy	1	1	bad(×)	0	
this	0		cat	0	1	very	0	1	beautiful	0	1
traditional	1	1	Christ's	1		wet	0	1	big(×)	0	
tric	0		cristan	0		whose	0	1	body	0	1
trick	0	1	cute	0	1	wonderful(×)			Bohemian	1	1
uniqu	0		destiny	0	1	worst	1		book	0	1
veautiful	0		dusy	0		your()	1		boring	0	1
violence	0	1	early	0	1				brother's	1	
visual	1		Empelar's	0					build	0	1
wall	1	1	emperor	0	1				careful	0	1
water	0	1	enjoy	0					cat	1	1
wintter	0		fairst	0					clean	0	1
wonderful(×)	0		famous	0	1				cluely	0	
wonderfull	0		farst	0					comical	0	1
wood	1	1	father's	1					cool	0	1
worest	0		fifth	1					crazy	0	1
worst	1	1	first	1					cute	0	1
write	0		friend's	1	1				danger	0	1
			fun	1	1				dangerous	0	1
			funny	1	1				deep	1	1
			Goethe's	1					desperate	1	1
			good(×)	0					difficult	0	1
			greatful	0	1				don't	0	
			happily	0					drak	0	
			happiy	0					dreadful	1	1
			happly	0					eager	0	1
			happy	1					enjoy	0	1

(Appendix 9.3 continues)

(Appendix 9.3 continued)

4-3			4-4						4-5		
() art	10	rater	() birthday	10	rater				() tragedy	10	rater
			hateful	0	1				expected	1	1
			her	0					famous	1	1
			his	0					fatal	1	1
			huppy	0					fearful	0	1
			impressive	1	1				first	1	
			interesting	0	1				fool	0	1
			Kim's	1					for	0	1
			Kiyomi's	1					foreign	1	1
			last	1					fuge	0	
			lucky	0	1				fun	0	1
			memorial	1	1				glad	0	1
			mother's	1					global	1	1
			my father	0					good(×)	0	
			my(×)	0					great(×)	0	
			new	0	1				happy	0	1
			next	1					heart-break	0	1
			nice(×)	0					heavy	0	1
			ninth	1					high	0	1
			on	0	1				ice	0	1
			only	0	1				inco	0	
			panic	0	1				interesting	1	1
			pleasant	1	1				invitation	0	1
			rainy	0	1				Japanese	1	1
			rice	0	1				large	0	1
			royal	1	1				last	1	1
			sad	1	1				light	0	1
			sadly	0					lion	0	1
			same	1	1				little	0	1
			school	0	1				local	1	1
			second	1					lonely	0	1
			sixty	0	1				long	0	1
			special	1	1				lucky	0	1
			speciality	0	1				magnificent	0	1
			special	0					minor	1	
			tenth	1					miseralbe	0	
			terrible	1	1				more	0	1
			their	0					my()	0	
			third	1					mysterious	1	1

4-5			4-6								
() tragedy	10	rater	() religion	10	rater						
new	1	1	admirable	1	1	majority	1		various	0	1
nice(×)	0		air	0	1	make	0	1	very	0	1
night	1	1	ancient	1	1	many	0	1	very cute	0	1
nomal	0		another	1	1	mejor	0		vice	0	1
nonfiction	1	1	Arabian	1	1	midium	0		wet	0	1
nonsense	0	1	bad(×)	0		minar	0		wide-spread	1	1
old(×)	0		beautiful	1	1	minor	1	1	winter	0	1
one	1		blue	0	1	most	0	1	wonderful()	0	
original	1	1	ceparate	0		music	0	1	world	1	
outstanding	0	1	chaming	0		my()	0		worldwide	1	1
perfect	1	1	Christ	0	1	mysterious	1	1	worst	1	1
poor	0	1	christ	0		national	1	1	wrong	0	1
popular	1	1	Christ's	1	1	natural	1		young	0	1
powerful	0	1	clean	0	1	new	1	1			
pretty	0	1	close	0	1	nice(×)	0				
prety	0		cold	0	1	nomal	0				
quick	0	1	common	1	1	none	0	1			
real	1		crazy	1	1	nonsense	0	1			
red	0	1	critism	0		odd	1	1			
sad	1	1	cross	0		on	0	1			

(Appendix 9.3 continues)

(Appendix 9.3 continued)

4-5			4-6								
() tragedy	10	rater	() religion	10	rater						
science	0	1	cute	0	1	oriental	1	1			
sentimental	0	1	dande	0		origin	0	1			
serious	1	1	dangerous	1	1	original	1	1			
Shakepeare's	1		dark	0	1	other	1	1			
Shakespere's	0		de	0		our	0				
shocking	1	1	dead	1	1	own	1				
short	0	1	deep	1	1	paysoba	0				
sorrowful	1	1	deeply	0		peaceful	1	1			
sorry	0	1	different	1		percect	0				
special	0	1	difficult	1	1	perfect	1	1			
strange	1	1	dog	0	1	pious	1	1			
strong	0	1	doubtful	1	1	polite	0	1			
successive	0	1	dreamy	0	1	popular	1	1			
sudden	1		easy	0	1	powerful	1	1			
super	0	1	end	0	1	pretty	0	1			
terrible	1		ethnic	1	1	primitive	1				
to	0		excite	0		pure	1	1			
traditional	1	1	faithful	0	1	raion	0				
train	1	1	famous	1	1	rediculous	0				
tremble	0	1	final	0	1	respectable	1	1			
tremendous	1	1	first	1		respective	0	1			
tsunami	1	1	fool	0	1	ridiculous	1	1			
ugly	0	1	foolish	1	1	right	0	1			
un	1		freedom	0	1	rigter	0				
unbelievable	1	1	friend	0	1	sad	0	1			
unexpected	1	1	global	1	1	same	1	1			
unexpressive	0	1	go	0	1	serious	1	1			
unforgatable	1		good(×)	0		sillent	0				
unfortunately	0		great(×)	0		silous	0				
unfortune	0	1	greed	0	1	sister's	1				
unhappy	0	1	happy	0	1	small(×)	0				
unknown	1	1	heat	0	1	some	1	1			
unlucky	0	1	high	0	1	son	0				
vast	1	1	holy	1	1	special	1	1			
very	0	1	horrible	1	1	standard	1	1			
warm	0	1	huge	0	1	start	0	1			
well	0	1	hungry	0	1	strange	1	1			
white	0	1	i	0		strong	0	1			
wild	0	1	important	1	1	super	0	1			
worse	1	1	incomprehensable	0		superb	0	1			
worst	1	1	incomprehensive	1	1	superior	1	1			
new	1	1	independent	0	1	tall	0	1			
			interesting	1	1	terrible	1	1			
			Islam	1	1	tradisional	0				
			Japanese	1	1	tragic	0	1			
			kind	0	1	un	1				
			large	0	1	unbelievable	1	1			
			local	1	1	unbelivable	0				
			mad	0	1	uncomprehensible	0	1			
			majn	0		understandable	1	1			
			major	1		unfamiliar	1	1			

5-1									5-2		
perfect ()	10	rater							wet ()	10	rater
ability	1	1	house	1	1	sence	1	1	air	0	1
action	1	1	human	1	1	shcool	0		bady	0	
address	1	1	humans	1	1	shoes	1	1	bath	0	1
air	0	1	hypothesis	1	1	shoot	0	1	bed	1	
ancer	0		idea	1	1	shot	1	1	bet	0	1
animal	1	1	imo	0		shoto	0		blanket	1	1
anser	0		information	1	1	singer	1	1	body	1	1
ansewr	0		job	1	1	size	1	1	bog	1	1
ansur	0		kichev	0		skill	1		book	1	1
answer	1	1	ladg	0		sky	0	1	box	1	1
answers	1	1	lady	1	1	smile	1	1	boy	1	1
anwer	0		life	1	1	snap	1	1	bsut	0	
art	1	1	line	1	1	snow	1	1	cat	1	1
ate	0		looks	1	1	song	1	1	climinat	0	
baby	1	1	love	1	1	speaking	0	1	closeth	0	
ball	1	1	machine	1	1	sports	0	1	cloth	1	
baseball	0	1	man	1	1	spots	1	1	clothe	1	1
berabera	0		map	1	1	square	1	1	clothes	1	
blind	0	1	mark	1	1	srim	0		clothing	1	1
blue	0	1	marks	1	1	stail	0		cloths	1	
body	1	1	men	1	1	stile	0	1	coat	1	1
book	1	1	method	1	1	stlye	0		condition	1	1
boy	1	1	mission	1	1	stoam	0		dag	0	
brain	0	1	mode	1	1	stome	0		dance	0	1
brother	1	1	mother	1	1	storm	1	1	day	1	
car	1	1	movement	1	1	story	1	1	desk	1	1
carculate	0		music	1	1	stream	1	1	dessert	0	1
cat	1	1	noise	1	1	student	1	1	dir	0	
choice	1	1	oil	1	1	students	1	1	dish	1	1
circle	1	1	oile	0		study	1	1	dog	1	1
claim	1	1	para	0		style	1	1	doth	0	
clime	0	1	parson	1	1	sudent	0		dream	1	1
computer	1	1	partner	1	1	sun	0	1	dress	1	1
crime	1		pen	1	1	superman	0	1	dust	0	1
cuple	0		people	1	1	supperman	0		ear	1	1
dark	0	1	perfect	0		tast	0		earth	1	1
day	1		performance	1		taste	1	1	experiment	0	1
days	1		person	1		teacher	1	1	eye	1	1
defence	1	1	personality	1	1	team	1	1	eyes	1	1
desk	1	1	piching	0		technology	1	1	face	1	1
diamond	1	1	picing	0		test	1	1	fall	0	1
dictionary	1	1	picking	0	1	the	0		fish	1	1
dog	1	1	pictur	0		thing	1	1	floor	1	1
door	1	1	picture	1	1	tragedy	0	1	flower	1	1
dream	1	1	pitch	1	1	tuis	0		food	1	1
dreamer	1	1	pitching	1	1	TV	0	1	fur	1	1
English	1		pitting	0	1	understanding	1	1	girl	1	1
er	0		place	1		victory	1	1	go	0	1
exam	1	1	plan	1	1	vision	1	1	grass	1	
exsam	0		play	1	1	voice	1	1	green	0	1
face	1	1	player	1	1	wave	1	1	grip	0	1
fan	1	1	point	1	1	wife	1	1	ground	1	1
fice	0	1	points	1	1	win	1	1	gruss	0	
food	1	1	prize	1	1	winner	1	1	hair	1	
form	1	1	problem	0	1	woman	1	1	hairs	1	
freeze	0	1	qu	0		words	1	1	hand	1	1
from	0		queen	1	1	work	1	1	handkerchief	1	1
game	1	1	question	1	1	world	1		hands	1	1
gard	0		reason	1	1				hankerchief	0	
garm	0		record	1	1				hat	1	1
girl	1	1	report	1	1				head	1	1
girls	1	1	research	1	1				hear	0	1
god	1	1	result	1	1				hiar	0	

(Appendix 9.3 continues)

(Appendix 9.3 continued)

5-1									5-2		
perfect ()	10	rater							wet ()	10	rater
grade	1	1	robot	1	1				homework	0	1
guide	1	1	room	1	1				house	1	1
guir	0		run	1	1				human	0	1
guy	1	1	saibogu	0					island	1	1
hair	1	1	same	0	1				kami	0	1
hoip	0		scheam	0					land	1	1
home	1	1	scientist	1	1				lip	1	1
homework	1	1	score	1	1				load	0	1

5-2						5-3					
wet ()	10	rater	time	0	1	minor ()	10	rater			
man	1	1	tire	1	1	accident	1	1	lerg	0	
men	1	1	tish	0		aer	0	1	leter	0	
monstar	0		tishes	0		air	0	1	level	1	1
nose	1	1	tishu	0		animal	1	1	liege	0	1
paper	1	1	tissh	0		answer	1	1	line	0	1
pen	1	1	tisshes	0		ant	0	1	magazine	0	1
pepar	0		tisshu	0		ants	0	1	man	1	1
peper	0		tissue	1	1	artist	1	1	maner	0	
place	1	1	tisue	0		attitude	1	1	map	1	1
pussy	1	1	toiret	0		bag	0	1	milion	0	
rain	0	1	Tokyo	0	1	ball	0	1	mind	1	1
religion	0	1	towel	1	1	band	1	1	minor	0	
rider	0	1	towl	0		bank	1	1	mistake	1	1
road	1		tree	1	1	bard	0	1	momor	0	
room	1	1	T-shirt	1	1	baseball	0	1	moncky	0	
sand	1	1	TV	0	1	bird	1	1	monny	0	
seart	0		twish	0		boll	0	1	monter	0	
season	1	1	umbrella	1	1	bomber	0	1	mouse	0	1
seat	1	1	uniform	1	1	book	1	1	movie	1	1
seet	0	1	up	0	1	boy	0	1	music	1	1
shart	0		washlet	0		bus	0	1	musician	0	1
shat	0		wat	0		c	0		name	0	1
sheet	1	1	wear	0	1	camera	0	1	ok	0	1
shirt	1	1	weather	1	1	car	0	1	opinion	0	1
shit	1	1	wet	1		cat	0	1	paper	0	1
shite	1	1	wether	0	1	CD	0	1	parson	1	1
shoe	1	1	wind	0	1	cencer	0		part	1	1
shoes	1	1	window	1	1	change	1		party	0	1
shooes	0		woman	1	1	children	0	1	pencil	0	1
short	0	1	year	1	1	choice	1	1	people	0	1
shose	0					chord	1	1	person	1	1
shout	0	1				city	1	1	place	1	1
shuot	0					class	1	1	play	0	1
shurt	1	1				code	1	1	player	1	1
shurts	1	1				computer	0	1	program	1	1
shut	1	1				cord	0	1	proman	0	
skin	1	1				cort	0		question	0	1
socks	1	1				country	1	1	race	1	1
soil	1	1				creat	0		rat	0	1
sokes	0	1				culture	1	1	reag	0	
solid	0	1				desk	0	1	reage	0	
sox	1	1				dog	0	1	reague	0	
sponge	1					door	0	1	rearge	0	
store	1	1				driver	0	1	red	0	1
storit	0					dust	0	1	region	0	1
suirts	0					egg	0	1	relgion	0	
suit	1	1				face	0	1	religion	1	1
suits	1	1				fashion	1	1	remark	0	1
sulth	0					fish	0	1	report	1	1

(Appendix 9.3 continues)

(Appendix 9.3 continued)

5-2						5-3					
wet ()	10	rater	time	0	1	minor ()	10	rater			
surt	0					food	0	1	route	1	1
surts	0					fox	0	1	rule	1	1
susesen	0					game	1	1	scientist	0	1
sute	0					ge	0		seven	0	1
sutu	0					girl	0	1	shop	0	1
sweater	1	1				glass	0	1	singer	1	1
swimsouts	0					go	0	1	song	1	1
swwetsu	0					group	1	1	songer	0	
tai	0					hobby	0	1	sport	1	1
taol	0					house	0	1	sports	1	1
taole	0					human	0	1	station	0	1
tawel	0					idea	1	1	subject	1	1
tawol	0					information	0	1	sum	0	1
teash	0					key	0	1	table	0	1
teesh	0					knowlege	0		team	1	1
teish	0					language	1	1	teem	0	
ter	0					leag	0		theory	0	1
tesh	0					leage	0		thing	0	1
teshu	0					leager	0		things	0	1
thing	1	1				leagu	0		threat	1	1
thish	0					league	1	1	tolk	0	
tichsu	0					leng	0		tork	0	

5-3			5-4						6-1		
minor ()	10	rater	eager ()	10	rater				enjoy ()	10	rater
town	0	1	a	0		oil	0	1	art	1	1
TV show	0	1	air	0	1	old	0	1	bacation	0	
ty report	0		animal	1	1	olugan	0		bareball	0	
unakiken	0		apple	0	1	opinion	0	1	baseball	1	1
wall	0	1	bag	0	1	orange	0	1	basketball	1	1
way	0	1	ball	0	1	pen	0	1	bemari	0	
woman	0	1	bike	0	1	people	1	1	birthday	1	1
word	0	1	bird	1	1	person	1	1	book	1	
work	0	1	blue	0	1	place	0	1	boy	0	1
writer	0	1	bog	0	1	play	0	1	brive	0	
			book	0	1	player	1	1	class	1	1
			box	0	1	police	1	1	cook	0	1
			boy	1	1	pople	0		cooking	1	1
			bud	0	1	practice	0	1	dance	1	1
			cancer	0	1	present	0	1	dancing	1	1
			car	0	1	problem	0	1	day	1	1
			card	0	1	question	0	1	days	1	1
			cat	1	1	rabbit	1	1	dinner	1	1
			chair	0	1	reading	0	1	do	0	
			chance	0	1	requirement	0	1	drive	1	1
			coat	0	1	rice	0	1	driving	1	1
			collector	1	1	robot	1	1	dunce	0	1
			cong	0		scholar	1	1	earty	0	
			den	0	1	school	0	1	eat	0	
			desire	0	1	sea	0	1	eating	1	1
			dog	1	1	sense	0	1	English	1	1
			doilling	0		side	0	1	festival	1	1
			dream	0	1	site	0	1	festivel	0	
			due	0	1	sport	0	1	film	1	1
			eagor	0		stant	0		fishing	1	1
			eater	1	1	state	0	1	football	1	1
			edge	0	1	student	1		friend	0	1
			educationist	1	1	supporter	1	1	frinds	0	
			effect	0	1	table	0	1	game	1	1
			egg	0	1	taxation	0	1	geam	0	
			elephant	1	1	teacher	1	1	girl	0	1

(Appendix 9.3 continues)

(Appendix 9.3 continued)

5-3			5-4						6-1		
minor ()	10	rater	eager ()	10	rater				enjoy ()	10	rater
			en	0		team	1	1	gluf	0	
			exam	0	1	test	0	1	golf	1	1
			exame	0		tests	0	1	guitar	1	1
			expectation	0	1	than	0	1	guiter	0	
			explore	0	1	thing	0	1	handball	1	1
			face	1	1	thinking	0	1	hiking	1	1
			family	1	1	thought	0	1	holiday	1	1
			fighter	1	1	tiger	1	1	horrday	0	
			film	0	1	to	0		Ieyasu	0	
			flower	0	1	to eat	0		karaoke	1	1
			friend	1	1	tragedy	0	1	life	1	1
			funs	0	1	TV	0	1	listening	1	1
			game	0	1	voice	1	1	lunch	1	1
			girl	1	1	way	0	1	magazin	0	
			guy	1	1	window	0	1	mail	0	1
			hand	0	1	women	1	1	move	0	1
			her	0	1	work	0	1	movie	1	1
			hobby	0	1	worker	1	1	movies	1	1
			hope	1	1	writer	1	1	music	1	1
			horse	1	1	young	0	1	musick	0	
			house	0	1				myself	1	
			human	0	1				oneself	1	
			land	0	1				park	1	1
			letter	0	1				partey	0	
			lion	1	1				partty	0	
			lump	0	1				party	1	1
			lunch	0	1				pary	0	
			man	1	1				perty	0	
			marrige	0					piano	1	1
			match	0	1				picnic	1	1
			ment	0	1				pirty	0	
			monkey	1	1				play	1	1
			mouth	0	1				play tennis	0	
			neck	0	1				playing	1	1
			nose	0	1				playing the piano	1	1

6-1			6-2						6-3		
enjoy ()	10	rater	buy ()	10	rater				destroy ()	10	rater
pretty	0	1	anything	1	1	hut	0	1	an	0	
read	0	1	apple	1	1	icecleam	0		animal	1	1
reading	1	1	apples	1	1	item	1	1	anya	0	
running	1	1	bag	1	1	juice	1	1	apartment	1	1
same	0	1	ball	1	1	magazine	1	1	apple	0	1
school	1	1	base	1	1	meat	1	1	arth	0	
shopping	1	1	baseball	0	1	meet	0	1	bag	0	1
sing	0	1	basketball	0	1	milk	1	1	ball	1	1
singing	1	1	bed	1	1	neckress	0		baseball	1	1
skating	1	1	ben	0	1	nootbook	0		bell	1	1
ski	0	1	bicycle	1	1	note	1	1	bike	1	1
skiing	1	1	bike	1	1	notebook	1	1	bill	0	1
sking	0		boat	1	1	orange	1	1	billding	0	
sleep	0	1	boll	0	1	orenge	0		biulding	0	
soccer	1	1	boo	0	1	paper	1	1	bod	1	1
soccerball	0	1	book	1	1	pen	1	1	bog	0	1
soccrball	0		books	1	1	penchil	0		book	1	1
softball	1	1	box	1	1	pencil	1	1	bottle	1	1
spccer	0		breads	1	1	pet	1	1	box	1	1
speaking	1	1	bug	0	1	piano	1	1	boy	1	1
sport	1	1	buuk	0		plane	1	1	bridge	1	1
sports	1	1	bycicle	0		potate	0	1	brock	0	1

(Appendix 9.3 continues)

(Appendix 9.3 continued)

spotu	0		cake	1	1	preasent	0		buil	0	
stakeboard	0	1	camera	1	1	present	0	1	building	1	
stay	1	1	cameras	1	1	salt	0	1	buildings	1	
storl	0		candies	1	1	school	0	1	built	0	
story	1	1	candy	1	1	shirt	1	1	bulding	0	
study	0	1	cap	1	1	shoes	1	1	cake	1	1
studying	1	1	car	1		shoose	0		car	1	1
succer	0		card	1	1	skart	1	1	card	1	1
summe	0		cat	1	1	slave	1	1	castle	1	1
summer	1	1	CD	1	1	soccer ball	0		cat	1	1
summer vacation	1	1	CDs	1	1	something	1	1	CD	1	1
swim	0	1	cer	0		soup	1	1	chair	1	1
swimming	1	1	chair	1	1	suit	1	1	chear	0	
talk	0	1	chocolate	1	1	sweater	1	1	chest	0	1
talking	1	1	clock	1	1	swet	0		city	1	1
tark	0		cloth	1	1	sweter	0		computer	1	1
television	1	1	clothes	1	1	table	1	1	construction	0	1
tenis	0		coat	1		the way	0	1	country	1	1
tennice	0		coffee	1	1	thing	1	1	creature	1	1
tennis	1	1	comic	1	1	ticket	1	1	cup	1	1
tennis game	1	1	computer	1	1	tickets	1	1	cut	0	1
tenniss	0		cooki	0		time	1		day	0	1
test	1	1	cooler	1	1	tomate	0	1	desk	1	1
time	1	1	cort	0		toy	1	1	dictionary	1	1
tomorrow	0	1	cup	1	1	train	0	1	dish	0	1
tour	1	1	cut	0	1	T-shat	0		dog	1	1
travel	1	1	desk	1	1	T-shirt	1	1	door	1	1
traveling	1	1	dictionary	1	1	T-shout	0		dream	0	1
trip	1	1	dog	1	1	TV	1	1	earth	1	1
TV	1	1	doll	1	1	vegetable	1	1	elephant	1	1
vacation	1	1	dress	1		video	1	1	enemy	1	1
valleyball	0		drink	1		wallet	1	1	enemy building	1	1
view	1	1	egg	1	1	watch	1	1	enjel	0	
visiting	1	1	eraser	1	1	water	1	1	envilonment	0	
volleyball	1	1	fish	1	1	wax	1	1	er	0	
watch	0	1	flower	1		way	0	1	factory	1	1
watching	1	1	food	1	1	wear	0	1	flower	1	1
weekend	1		foods	1	1	woman	1	1	forest	1	1
work	1	1	fruit	1	1				forests	1	1
working	1	1	game	1	1				girl	1	1
writing	1	1	ghost	0	1				glass	0	1
writting	0		goods	1	1				glasses	0	1
xmas	0		guitar	1	1				goast	0	
yourself	1		guiter	0					gost	0	
			guitter	0					guitar	1	1
			gun	1	1				guiter	0	
			hat	1	1				heart	0	1
			hobby	0	1				house	1	1
			house	1	1				houses	1	1

6-3						6-4					
destroy ()	10	rater				wash ()	10	rater			
human	1	1	waste	0	1	apple	1	1	towel	1	1
incect	0		water	0	1	bas	0	1	water	0	1
it	0		way	0	1	basketball	0	1	wear	0	1
kid	1	1	window	1	1	bath	0	1	window	1	1
king	0	1	wood	1	1	bathroom	1	1	windows	1	1
lake	1	1	world	1	1	bike	1	1	wish	0	1
land	1	1				body	1	1			
life	1	1				bog	0	1			
ling	0	1				bus	1	1			
man	1	1				car	1	1			
mather	1	1				cat	1	1			

(Appendix 9.3 continues)

(Appendix 9.3 continued)

6-3						6-4					
destroy ()	10	rater				wash ()	10	rater			
mind	1	1				cer	0	1			
mirror	1	1				chopstick	1	1			
monstar	0					cloth	1	1			
mood	0	1				clothes	1	1			
mother	1	1				cook	0	1			
mountain	1	1				cop	0	1			
movie	0	1				crosses	1	1			
natur	1	1				cup	1	1			
nature	1	1				cups	1	1			
neus	0					cut	0	1			
notebook	1	1				deash	0				
order	0	1				desh	0				
para	0					deshes	0				
park	1	1				dinner	0	1			
parson	1	1				dirty	0				
patty	0	1				dise	0				
peace	1	1				dises	0				
pen	1	1				dish	1				
pencil	1	1				dishees	0				
people	1	1				disher	0				
person	1	1				dishes	1				
piano	1	1				dishese	0				
pictire	0					dishies	0				
picture	1	1				disk	0	1			
plant	1	1				dodey	0				
poison	0	1				dog	1	1			
pollution	0	1				dream	0	1			
road	1	1				eyes	1	1			
robot	1	1				face	1				
room	1	1				finger	1	1			
school	1	1				fuse	0	1			
sentence	0	1				glass	1	1			
ship	1	1				hair	1				
soccer ball	1	1				hand	1				
star	1	1				hands	1				
stone	1	1				harnd	0				
store	1	1				haus	0				
story	0	1				head	1	1			
strey	0					house	1	1			
structure	1	1				hund	0				
supermarket	1	1				hunds	0				
table	1	1				jeacket	0				
tawn	1	1				nife	0				
team	0	1				pan	1	1			
the	0					paper	0	1			
thurre	0					plate	1	1			
Tokyo	1	1				ret	1	1			
torch	1	1				shate	0				
tour	0	1				shirt	1	1			
tower	1	1				shit	0	1			
town	1	1				shoes	1	1			
toy	1	1				shoses	0				
trees	1	1				shouse	0	1			
TV	1	1				sox	1	1			
twon	0					teeth	0	1			
vase	1	1				television	0	1			
video	1	1				three	0				
village	1	1				tish	0				
wall	1	1				toilet	1	1			
war	0	1				tooth	0	1			

6-5						6-6					
cook ()	10	rater				pay ()	10	rater			
apple	1	1	meat	1	1	100 yen	1		letter	0	
barger	0		meet	0		10000000 yen	1		lunch	0	
beaf	0		miso soup	0		3 yen	1		maner	0	
beef	1	1	mother	0		apple	0		maney	0	
binner	0		natto	0		artist	0	1	many	0	
bird	1	1	okonomiyaki	0		atensstion	0		meal	0	
breadkfast	0		omlet	1	1	atenthion	0		meney	0	
breakfast	1		omnlet	0		atention	0		mind	0	1
cake	1	1	onigeri	0		attension	0		momey	0	
caraaze	0		onion	1	1	attention	1		moner	0	
cary	0		pasta	1	1	back	0		money	1	
cat	0	1	peach	1	1	bag	0		monie	0	
chichen	0		piano	0		ball	0		monney	0	
chickin	0		pie	1	1	bill	1		monny	0	
chiken	0		piezz	0		bills	1		montr	0	
chikin	0		piza	0		bit	0	1	mony	0	
chocolate	1	1	pizza	1	1	book	0		mouth	0	
cikin	0		plate	0	1	books	0		notebook	0	
cittin	0		poison	0	1	box	0		off	0	1
cokiy	0		potato	1	1	boy	0	1	on day	0	
cook	0		poteto	0		buy	0		pay	0	
cooki	0		rice	1		cake	0		payment	1	1
cookie	0	1	rise	0		camk	0		pecure	0	
cookies	0	1	rowfish	0		car	0		pen	0	
cookires	0		runch	0	1	card	0	1	pencil	0	
cooky	0	1	salad	0	1	cash	1		phone	1	1
coop	0		salada	0		cat	0	1	piano	0	
cuck	0		samdwich	0		CD	0		pizza	0	
curry	1	1	sarad	0		change	0	1	place	0	
curry and rice	1	1	shef	0		check	0		plain	0	
cut	0	1	soba	1	1	chip	0	1	price	1	
dan	0		something	1	1	chips	0	1	respect	1	
denar	0		soop	0		close	0		rice	0	
deneer	0		soup	0		cloth	0		shirt	0	
dener	0		spergetty	0		coat	0		shot	0	1
denner	0		stake	0		coin	1	1	soccer	0	
desert	0		steak	1	1	coins	1	1	soccer ball	0	
diner	0		sukiyaki	1	1	con	0		something	1	1
dinery	0		supe	0		cook	0		son	0	1
dinnar	0		sure	0		cost	0	1	T shirt	0	
dinner	1		sushi	0	1	cregit	0		tantenichiyon	0	
dinnie	0		susi	0		cut	1	1	tantion	0	
dish	1	1	takoyaki	0		daller	0		tax	1	
dishes	1	1	tamagoyaki	0		day	0	1	tee	0	
dog	0	1	tomato	1	1	debt	1		tennis	0	
donutu	0		tometoes	0		dog	0	1	tension	0	1
dudu	0		udo	0	1	doll	0		tention	1	
eeg	0		vegetable	1	1	dollar	1		thing	0	1
egg	1	1	yakisoba	0		doller	0		ticket	0	
eggs	1	1				dress	0		tiket	0	
eninem	0					egg	0		time	0	1
fish	1	1				era	0		toy	0	
fith	0					eye	0		T-shirt	0	
flied egg	0					far	0		TV	0	
food	1	1				fare	1	1	wage	1	1
frinch	0					fee	1	1	yen	1	
funch	0					fine	1				
furuit	0					floor	0				
hambarg	0					flower	0				
hambarger	0					focent	0				
hambuger	0					food	0				
him	0					fruit	0				
hotcake	1	1				gold	0	1			
humberg	0					guitar	0				

(Appendix 9.3 continues)

(Appendix 9.3 continued)

6-5						6-6					
cook ()	10	rater				pay ()	10	rater			
ice	0	1				hat	0				
kitchen	0					house	0				
lanch	0					in	0				
lice	0					interest	1	1			
lunch	1					juice	0				
lunth	0					law	0				
meal	1					lawyer	1	1			

6-7			6-8								
ride ()	10	rater	write ()	10	rater						
airplane	0	1	answer	1							
away	0		apul	0							
bas	0		art	0	1						
beycelce	0		article	1	1						
bice	0		book	1							
bicecly	0		books	1							
bicicle	0		bycicle	0							
bick	0		capital	0	1						
bicycke	0		car	0	1						
bicycle	1		cat	0	1						
bikck	0		composition	1	1						
bike	1		cut	0	1						
bive	0		dialy	0							
boat	1	1	diary	1	1						
book	0		dog	0	1						
bort	0		down	0							
bus	1		English	1	1						
buy	0		essay	1							
buyscle	0		girl	0	1						
bycical	0		hand	0	1						
bycicle	0		house	0	1						
bycircle	0		Japanese	1	1						
bycycle	0		language	0	1						
byke	0		lateer	0							
bysecle	0		later	0							
car	1	1	latter	0	1						
carriage	1	1	letar	0							
cat	0	1	leter	0							
cow	1	1	leterr	0							
cut	0		lettar	0							
dicycle	0		lette	0							
dike	0		letter	1							
dog	0	1	letters	1							
ear	0		lettre	1	1						
elephant	1	1	litter	0	1						
fish	0	1	littey	0							
girl	0	1	little	0	1						
horse	1		name	1							
horth	0		noot	0							
hose	0		note	1	1						
hoss	1		notebook	0	1						
hourse	0		novel	1							
hous	0		number	1	1						
house	0	1	paiper	0							
hurse	0		paper	1	1						
jet corster	0		pen	0	1						
maginger Z	0		people	0	1						
man	0	1	pepar	0							
mountain	0	1	picture	0	1						
news	0	1	poem	1							
on	0		report	1	1						

(Appendix 9.3 continues)

(Appendix 9.3 continued)

6-7			6-8								
ride ()	10	rater	write ()	10	rater						
on time	0		repote	0							
parson	0	1	reter	0							
rollarcoaster	0		retter	0							
sky	0	1	script	1	1						
strider	0	1	sentence	1	1						
sun	0	1	sign	0	1						
taun	0		skin	0							
time	0	1	story	1							
train	1	1	TV	0	1						
way	0	1	ward	0							
			wieter	0							
			word	1	1						
			words	1	1						
			wreteer	0							

Appendix 9.4

Raters' linguistic background history (Raters who evaluated the Antonym Subsection and Collocation Subsection)

Question	Rater A	Rater B	Rater C	Rater D	Rater E
Is English your first language (L1)?	Yes.	Yes.	No.	No.	No, but I can speak English better than L1.
(For a non-native speaker) Which level of English are you at broadly, advanced, intermediate, novice?			Advanced	Advanced overall.	
What other languages have you learned? (For each language) Which level are you at broadly, advanced, intermediate, novice?	French, Italian, Spanish, & Japanese (advanced), Portuguese (intermediate), Bahasa & Indonesia (novice).	Thai & Computer languages (novice).	None.	German (novice).	Thai & Spanish (novice).
(For a native speaker) How long have you been in Japan? OR How long were you in Japan?	I was for 6 years in Japan plus visits twice a year for four years after that, ten days each.	I have been here for 15 years.			Since last June (for eight months).
(For a non-native speaker) How long have you been in countries where English is spoken?			For about 4 years.	For 9 months.	

Note. Rater E was considered a native speaker of English because she started to acquire English before puberty and in childhood (based on Davies, 2004, p. 435; see Davies, 2004, for other five definitions of the native speaker).

Appendix 9.5

Item measurement report of the Productive Vocabulary Knowledge Test (all the Size [second analysis] and Depth Sections)

No.	Answer	*M*		InfitMS	InfitZ			*M*		InfitMS	InfitZ
	Size						Derivation				
1	dog	-7.08	0.39	0.95	-0.13	1	work	-4.53	0.15	1.11	1.04
2	mouse (mice)	-1.42	0.08	1.19	5.16	2	art	-2.45	0.10	1.05	1.02
3	sock(s)	-2.61	0.09	1.15	3.33	3	destroy	3.98	0.19	0.81	-1.59
4	maybe	-1.21	0.08	1.22	5.77	4	criticize	1.94	0.12	1.41	5.00
5	perfect	-2.63	0.09	1.07	1.55	5	inform	0.15	0.09	1.00	0.10
6	interpretation	3.75	0.20	0.93	-0.50	6	comfortable	1.03	0.10	1.13	2.22
7	believe	-0.90	0.08	0.89	-2.98	7	successful	-0.54	0.08	1.13	3.36
8	bridge	-1.26	0.08	0.96	-1.12	8	able	2.79	0.14	0.79	-2.50
10	he	-3.95	0.12	1.19	2.40	9	invitation	-0.33	0.09	1.07	1.69
13	driver	-3.87	0.11	0.86	-1.97	10	supporter	-2.77	0.10	0.94	-1.02
14	independence	1.23	0.11	1.14	2.09	11	argument	1.69	0.11	0.71	-4.65
15	explode	2.49	0.14	0.82	-2.14	12	emphasis	2.66	0.14	0.76	-2.90
16	behind	-0.87	0.08	1.06	1.60	13	independence	-0.46	0.08	0.87	-3.46
19	group	-1.32	0.08	1.23	6.08	14	religion	2.47	0.13	0.70	-4.00
20	warn	0.85	0.10	0.72	-5.25	15	similarly	-0.43	0.08	0.83	-4.74
22	start	-4.76	0.15	1.03	0.26	16	origin	-1.28	0.09	1.17	3.97
23	abroad	-1.45	0.08	0.83	-4.97	17	warmth	-1.26	0.09	0.93	-1.85
25	usually	-1.35	0.08	1.30	7.84	18	specialist	-1.92	0.09	0.91	-1.95
26	bend	0.96	0.10	0.69	-5.80	19	dirt	-0.45	0.08	0.94	-1.73
27	warmth	3.26	0.17	0.92	-0.70	20	end	-0.32	0.09	1.02	0.57
28	lady	-1.76	0.08	0.89	-3.07		M	0.00	0.11	1.00	-0.60
30	invitation	1.91	0.12	0.86	-1.94		SD	2.08	0.03	2.00	2.80
32	theater(theatre)	-0.08	0.09	0.82	-4.06		Antonyms				
33	deserve	3.67	0.20	0.77	-1.75	1	open	-4.09	0.13	0.98	-0.23
34	above	0.94	0.10	0.97	-0.56	2	high	-2.97	0.1	0.97	-0.55
38	birthday	-3.59	0.11	0.97	-0.44	3	wet	-0.58	0.08	0.97	-0.78
39	orchestra	1.51	0.11	0.92	-1.11	4	alive	-0.33	0.08	0.92	-2.13
40	king	-4.14	0.12	1.34	3.76	5	independence	0.93	0.1	0.86	-2.74
41	pleasure	1.39	0.11	0.88	-1.87	6	start	-2.78	0.1	1.39	6.36
42	happily	1.74	0.12	0.90	-1.49	7	nonsense	0.12	0.09	1.16	3.5
43	guy	-0.63	0.08	1.14	3.44	8	happily	-0.02	0.09	0.90	-2.45
44	vast	2.31	0.13	0.85	-1.90	9	like	excluded			
45	informal	2.89	0.16	0.91	-0.85	10	tragedy	2.74	0.14	0.71	-3.44
46	stare	2.08	0.13	0.99	-0.16	11	supporter	3.69	0.19	0.93	-0.49
47	content	1.75	0.12	0.97	-0.42	12	able	1.26	0.1	1.13	2.21
48	tragedy	1.70	0.12	0.63	-5.94	13	export	1.98	0.12	0.87	-1.93
50	origin	0.43	0.09	0.95	-0.96	14	buy	-2	0.09	0.97	-0.76
51	supporter	1.97	0.12	0.90	-1.34	15	minor	1.2	0.1	0.76	-4.59
52	require	2.70	0.15	0.76	-2.73	16	beautiful	2.25	0.12	1.09	1.1
53	feed	0.09	0.09	0.81	-4.26	17	quick	-1.87	0.09	0.87	-3.18
54	possession	4.78	0.29	0.95	-0.21	18	hungry	0.49	0.09	1.27	5.31
56	guest	-0.11	0.09	0.97	-0.62		M	0.00	0.11	1.00	-0.30
57	dignity	3.63	0.19	0.81	-1.45		SD	2.11	0.03	0.20	3.00
59	egg(s)	-4.44	0.14	0.98	-0.16		Collocation				
62	apple(s)	-4.26	0.13	0.89	-1.36	1	dog	-1.06	0.09	1.05	1.05
63	treasure	1.18	0.11	0.84	-2.63	2	cancer	3.00	0.11	0.90	-1.48
66	onion	-3.51	0.10	1.21	3.16	3	art	-0.53	0.08	0.97	-0.85
67	enjoy	-3.80	0.11	1.11	1.57	4	birthday	-1.25	0.09	1.15	2.80
68	lake	-3.52	0.10	0.90	-1.62	5	tragedy	3.36	0.12	0.80	-2.80
69	ghost	0.89	0.10	1.06	0.92	6	religion	2.85	0.10	0.78	-3.78
70	date	-0.68	0.08	1.21	5.07	7	perfect	-1.31	0.10	0.92	-1.60
71	labor (labour)	1.68	0.12	0.62	-6.13	8	wet	0.48	0.08	0.92	-2.37
72	regardless	4.29	0.24	0.93	-0.39	9	minor	2.14	0.09	0.95	-1.01
75	thread	4.23	0.24	1.22	1.12	10	eager	3.08	0.11	1.02	0.33
76	buy	-4.20	0.13	1.07	0.82	11	enjoy	-2.17	0.12	1.07	0.84
77	inform	1.71	0.12	0.78	-3.34	12	buy	-2.46	0.13	1.03	0.37
78	champion	-0.18	0.09	1.18	3.75	13	destroy	0.34	0.08	0.97	-0.93
79	pay	-2.48	0.09	0.85	-3.82	14	wash	-2.20	0.12	1.00	0.06
80	soldier	1.52	0.11	0.89	-1.63	15	cook	-0.88	0.09	1.18	3.88

(Appendix 9.5 continues)

(Appendix 9.5 continued)

No.	Answer	*M*		InfitMS	InfitZ			*M*		InfitMS	InfitZ
	Size						Derivation				
81	ruin	3.39	0.18	0.84	-1.30	16	pay	0.22	0.08	0.93	-2.26
83	lecture	1.60	0.11	0.73	-4.24	17	ride	-1.46	0.10	0.93	-1.22
84	jail	2.23	0.13	1.19	2.16	18	write	-2.15	0.12	0.93	-0.87
85	expect	0.20	0.09	0.82	-3.95		M	0.00	0.10	1.00	-0.50
86	tower	-1.58	0.08	1.02	0.57		SD	1.98	0.02	0.10	1.90
88	work	-5.16	0.17	0.98	-0.11						
90	canal	3.97	0.22	0.96	-0.25						
91	island	-2.91	0.09	0.92	-1.83						
92	religion	1.72	0.12	0.77	-3.47						
93	similarly	4.23	0.24	0.98	-0.11						
94	its	-0.17	0.09	1.21	4.32						
95	wet	-1.31	0.08	0.98	-0.60						
97	mayor	1.33	0.11	0.84	-2.52						
99	minor	2.20	0.13	0.83	-2.19						
100	art	-3.07	0.10	0.94	-1.15						
103	whose	-1.18	0.08	1.23	5.88						
106	consciousness	4.70	0.28	0.94	-0.25						
107	cancer	-0.04	0.09	1.05	1.03						
108	wave	-1.63	0.08	0.90	-2.90						
M		0.00	0.13	1.00	-0.40						
SD		2.72	0.06	0.20	2.90						

Note. *M* = Item difficulty; *SE* = Standard Error; MS = Mean Square statistic; Z = Mean Square "statistic standardized toward a unit-normal distribution";

Appendix 9.6

Estimated vocabulary size of junior and senior high school students

	n	*M*	*SD*	Minimum	Maximum	α	CI
Junior high	327	1032.72	414.84	38	2231	.93	817.60 to 1247.84
Senior high	583	1222.06	575.58	231	2962	.96	996.43 to 1447.69

Note. CI = 95% confidence interval = calculated using $M \pm 1.96 \times (SD \times \sqrt{(1-\alpha)}$ (

Appendix 9.7

Estimated vocabulary size for each school

	n	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Overall	913	1162.48	538.29	0.74	0.06	34.00	2966.00
JHS A	14	859.57	272.08	-0.47	-0.17	307.00	1261.00
JHS B	57	969.54	261.48	-0.09	1.00	239.00	1705.00
JHS C	26	760.46	330.96	0.80	-0.05	307.00	1534.00
JHS D	27	484.74	228.16	0.76	0.75	136.00	1091.00
JHS E	33	761.30	189.76	0.12	-0.43	375.00	1125.00
JHS F	36	1109.89	320.01	0.65	-0.70	682.00	1807.00
JHS G	137	1289.23	422.81	-0.44	0.04	34.00	2216.00
SHS A	32	723.41	273.60	1.50	3.57	273.00	1670.00
SHS B	136	889.36	352.55	0.31	-0.25	205.00	1841.00
SHS C	36	1052.97	279.32	0.61	0.59	443.00	1773.00
SHS D	57	1010.11	245.46	0.75	-0.10	716.00	1705.00
SHS E	84	912.36	311.36	0.72	1.10	239.00	1943.00
SHS F	46	943.39	256.75	0.15	-0.51	477.00	1534.00
SHS G	61	1458.10	307.95	0.40	-0.40	955.00	2182.00
SHS H	131	2068.08	377.19	-0.52	0.39	955.00	2966.00

Appendix 9.8

One-way Analyses of Variance for effects of the STEP Test grade on the Size and Depth Sections

Variable	Source	*df*	*SS*	*MS*	*F*	p^a	*df*	p^b	η_G^2
Size	Between	4	55590653.76	13897663.44	98.10	0.00	4	0.00	0.40
	Within	584	82733313.07	141666.63			120.41		
	Total	588	138323966.83						
Derivation	Between	4	3581.26	895.32	63.16	0.00	4	0.00	0.30
	Within	584	8278.51	14.18			146.30		
	Total	588	11859.77						
Antonym	Between	4	2079.64	519.91	73.20	0.00	4	0.00	0.33
	Within	584	4147.93	7.10			127.64		
	Total	588	6227.56						
Collocation	Between	4	1254.18	313.55	46.77	0.00	4	0.00	0.24
	Within	584	3914.73	6.70			68.07		
	Total	588	5168.92						

Note. SS = Sum of Squares; MS = Mean Squares. [a]Not corrected by Brown-Forsythe method; [b]Corrected by Brown-Forsythe method. Between = Between groups.

Appendix 10.1

Words that belong to 1 to 200 words in Ota et al. (2004) and the JACET8000

	JACET	Ota et al						
a	1-100	1-100	helps	101-200	1-100	place	101-200	No
about	1-100	1-100	her	1-100	101-200	play	101-200	1-100
after	1-100	1-100	here	101-200	101-200	put	101-200	No
all	1-100	1-100	high	101-200	No	reads	No	1-100
also	1-100	101-200	him	1-100	101-200	really	101-200	101-200
am	1-100	1-100	his	1-100	101-200	right	101-200	101-200
an	1-100	1-100	home	101-200	101-200	saw	1-100	No
and	1-100	1-100	house	101-200	101-200	school	101-200	101-200
another	101-200	No	how	1-100	1-100	see	1-100	1-100
any	101-200	1-100	however	101-200	No	seeing	1-100	1-100
are	1-100	1-100	i	1-100	1-100	seem	101-200	No
around	101-200	No	if	1-100	1-100	seems	101-200	No
as	1-100	1-100	important	101-200	No	she	1-100	1-100
at	1-100	1-100	in	1-100	1-100	small	101-200	No
back	1-100	101-200	into	1-100	No	so	1-100	1-100
be	1-100	1-100	is	1-100	1-100	some	1-100	1-100
beautiful	No	101-200	isn't	1-100	1-100	still	101-200	No
because	1-100	1-100	it	1-100	1-100	student	No	101-200
become	101-200	No	its	101-200	No	studies	No	1-100
before	101-200	1-100	know	1-100	1-100	study	No	1-100
best	No	101-200	larger	101-200	No	such	101-200	No
big	101-200	101-200	let	101-200	No	take	1-100	1-100
book	No	1-100	like	1-100	1-100	teacher	No	101-200
books	No	1-100	likes	1-100	1-100	than	1-100	101-200
both	101-200	No	little	1-100	101-200	that	1-100	1-100
bought	No	1-100	live	101-200	1-100	the	1-100	1-100
boy	101-200	No	lives	101-200	1-100	their	1-100	101-200
but	1-100	1-100	long	101-200	101-200	them	1-100	101-200
by	1-100	1-100	look	1-100	1-100	then	1-100	1-100
called	101-200	1-100	looking	1-100	1-100	there	1-100	1-100
came	1-100	101-200	looks	1-100	1-100	these	1-100	101-200
can	1-100	1-100	lot	No	101-200	they	1-100	1-100
change	101-200	No	love	101-200	1-100	think	1-100	1-100
changed	101-200	No	man	1-100	101-200	this	1-100	1-100
child	No	101-200	many	1-100	101-200	thought	1-100	No
class	No	101-200	may	101-200	No	three	No	1-100
come	1-100	1-100	me	1-100	101-200	time	1-100	101-200
could	1-100	No	more	1-100	1-100	times	1-100	101-200
country	101-200	101-200	mother	101-200	101-200	to	1-100	1-100
day	1-100	101-200	mothers	101-200	101-200	today	No	101-200
days	1-100	101-200	movies	No	101-200	too	101-200	101-200
different	101-200	No	much	101-200	1-100	try	101-200	No
dog	No	101-200	music	No	101-200	tv	No	101-200
doing	1-100	1-100	my	1-100	101-200	two	No	101-200
don't	1-100	1-100	name	101-200	101-200	up	1-100	101-200
down	101-200	No	named	101-200	101-200	use	1-100	1-100
english	No	101-200	next	101-200	101-200	uses	1-100	1-100
enjoy	No	1-100	nice	No	101-200	very	1-100	1-100
enjoying	No	1-100	no	1-100	1-100	want	1-100	1-100
every	101-200	101-200	not	1-100	1-100	was	1-100	1-100
families	101-200	101-200	now	1-100	1-100	watch	No	1-100
family	101-200	101-200	of	1-100	1-100	water	101-200	No
father	No	101-200	often	101-200	No	way	1-100	101-200
few	101-200	No	old	101-200	101-200	we	1-100	1-100
first	No	101-200	older	101-200	101-200	well	1-100	101-200
for	1-100	1-100	on	1-100	1-100	went	1-100	1-100
friend	101-200	101-200	one	No	1-100	were	1-100	1-100
friends	101-200	101-200	only	1-100	101-200	where	101-200	1-100
from	1-100	1-100	open	101-200	No	which	1-100	1-100
get	1-100	1-100	opened	101-200	No	who	1-100	1-100
go	1-100	1-100	or	1-100	101-200	will	1-100	1-100
going	1-100	1-100	other	1-100	1-100	with	1-100	1-100

(Appendix 10.1 continues)

(Appendix 10.1 continued)

	JACET	Ota et al						
a	1-100	1-100	helps	101-200	1-100	place	101-200	No
good	1-100	101-200	our	1-100	101-200	woman	101-200	No
great	101-200	101-200	out	1-100	101-200	women	101-200	No
had	1-100	1-100	over	1-100	101-200	work	1-100	1-100
hand	101-200	No	own	101-200	No	years	1-100	101-200
happy	No	101-200	part	101-200	No	you	1-100	1-100
has	1-100	101-200	parts	101-200	No	young	101-200	No
have	1-100	1-100	people	1-100	101-200	younger	101-200	No
he	1-100	1-100	peoples	1-100	101-200	your	1-100	101-200
help	101-200	1-100	picture	No	101-200			
helping	101-200	1-100	pictures	No	101-200			

Note. No = Not in the list.

Appendix 11.1

Examples of utterances from a student and calculation method

Task 1 (Self-introduction):

|my name is * * [Name].| |i have one sister.| {she is sorri} |i was in baseball club last year. (ERROR)| |and i like sport very much.| |and in those days, i'm interested in badminton.| |and i sometimes plays. (ERROR)| |it is very interesting sports (ERROR) i think.| |and i also like japanese sports, for example, judo and|

8 AS-units, 9 clauses, 3 dysfluency markers, 3 clauses with errors (6 error-free clauses) , 3 AS-units with errors (6 error-free AS-units)

1 subordinate clause (it is very interesting sports :: I think)

51 tokens, 36 types

11 grammatical word types (and, for, have, i, in, is, it, my, one, those, was) (36 – 11 = 25 lexical word types),

1 sophisticated word type (J1001) (baseball),

16 sophisticated word type (J101) (baseball, club, example, interested, interesting, japanese, last, much, name, one, plays, sister, sometimes, sport, sports, those)

F1	51/45*60 = 68.00
F2	9/45*60 = 12.00
F3	3/45*60 = 4.00
F4	3/51 = 0.06
F5	3/8 = 0.38
A1	6/9 = 0.67
A2	5/8 = 0.63
SC1	9/8 = 1.13
SC2	1/8 = 0.13
SC3	1/9 = 0.11
SC4	51/8 = 6.38

LC1	$36/\sqrt{51} = 5.04$
LC2	36/8 = 4.50
LC3	36/9 = 4.00
LC4	25/51 = 0.49
LC5	$25/\sqrt{51} = 3.50$
LC6	25/8 = 3.13
LC7	25/9 = 2.78
LC8	25/36 = 0.69
J1001	
LC11	1/51 = 0.02
LC12	$1/\sqrt{51} = 0.14$
LC13	1/8 = 0.13
LC14	1/9 = 0.11
LC15	1/36 = 0.03
J101	
LC11	16/51 = 0.31
LC12	$16/\sqrt{51} = 2.24$
LC13	16/8 = 2.00
LC14	16/9 = 1.78
LC15	16/36 = 0.44

Appendix 11.2
Other examples of utterances from students
Note. * = Student said the name; { } = repetitions, self-corrections, and other functionless words uttered. Words in { } and in Japanese (unless the Japanese words have become English words) were ignored in accuracy rating. The utterances for 45 seconds were transcribed.

■Self-introduction
my name is * *. my family is four people. i like baseball. and i like listen to music, too.

■Comparison of two pictures: e.g., a horse and a cow
right house is open the window. right the two house right door is blue. window two

■Picture description: Washing dishes
child is washing the cup in the mother. and

■Picture description: Cycling
there is cycling is mountain with bike of two in her eyes. there is {in the} in the cycling. sunny day.

■Comparison of two pictures: Before and after
before notebook name is taro. after notebook is name taro don't jiro. and before apple after it's an apple. team in the desk
---
■Self-introduction
my name is * *. and my house three and four families. and i like basketball. i am member of basketball club. and {my friend} my friend name is *. * likes play basketball, too. {high school} high school is very beautiful and very cruel.

■Comparison of two pictures: e.g., a horse and a cow
window is different from before window. window is opening. and {door's} door's color is blue, different from before. i don't know. please pass.

■Picture description: Washing dishes
children aids washing dishes and {help my} help your mother. and they're washing dishes. pass

■Picture description: Cycling
they're cycling. and {there} there many tree forest {and} and {house hou} house mountain. pass

■Comparison of two pictures: Before and after
apple is ate {half} half apples. pass

■Self-introduction
my name is * *. {i have family is} my family is to six. i have friend i'm don't know.

■Comparison of two pictures: e.g., a horse and a cow
open the window. don't horse. this is cow. not tree. this is a tree.

■Picture description: Washing dishes
the girl washing the cup in kitchen. her mother help the girl. she is washing

■Picture description: Cycling
three are recycling the road. there is many tree. sky is blue. this is beautiful.

■Comparison of two pictures: Before and after
my brother eat an apple. my brother he wrote jiro.

■Self-introduction
my name is * *. my family is six people, my grandfather and my grandmother and father and mother and old brother. my friend is very kind. i'm from *. i like listen to music. i play the basketball. my hobby is shopping and sleeping. my part time job in *.

■Comparison of two pictures: e.g., a horse and a cow
breaking door and house. right picture is left is picture cow in tree. right picture, second, three window. left picture is two window.

■Picture description: Washing dishes
my young sister is washing cup. {my mother is} my young sister is washing cup. my mother is very kind. two people is enjoy speaking about my dream.

■Picture description: Cycling
two people is riding the bike. it's a sun. i see sea. i see beautiful air.

■Comparison of two pictures: Before and after
dictionary and apple. picture is change. guitar change two books. table and two books change to ball. テレビ on the car change to テレビ on the two car.

■Self-introduction
my name is * *. i have a sister. her name is *. my parents {are in} are normal. my friends are many in my school.

■Comparison of two pictures: e.g., a horse and a cow
the windows is opened. the door's color is blue. {there are} there is a cow. there is a tree is around. there are four windows at the house.

■Picture description: Washing dishes
a girl is washing a cup in the kitchen. the woman help the girl to washing. there are many books on

■Picture description: Cycling
a man and a girl is riding a bike by the lake. there are many trees by the lake. the weather is very good.

■Comparison of two pictures: Before and after
i think the apple before is one. but after, the apple is half. and {the} the book is mine before. but after, the book is name jiro. another, there

Appendix 11.3 *Grammatical words that the participants used*

	frequency	judgement						
a	1332	gra	into	11	gra	twenty	15	context
about	29	gra	is	3651	gra	two	324	context
across	1	gra	isn't	32	gra	under	22	gra
after	376	gra	it	230	gra	us	1	gra
all	19	context	its	8	gra	(i')ve	2	gra
almost	4	gra	many	266	gra	wasn't	1	gra
am	15	gra	may	2	gra	we	24	gra
an	149	gra	me	77	gra	well	23	context

(Appendix 11.3 continues)

(Appendix 11.3 continued)

	frequency	judgement						
and	2800	gra	mine	2	gra	were	26	gra
another	39	gra	more	26	context	what	1	gra
any	6	gra	most	2	context	when	8	gra
anyway	1	gra	much	89	context	where	1	gra
are	1112	gra	must	2	gra	which	18	gra
aren	5	gra	my	1137	gra	who	13	gra
around	24	context	near	117	gra	whose	8	gra
as	4	gra	nearer	1	gra	why	1	gra
at	45	gra	never	1	gra	will	8	gra
be	10	gra	next	16	gra	with	170	gra
because	32	gra	no	43	gra	would	1	gra
been	4	gra	nobody	1	gra	yeah	3	gra
before	427	gra	not	123	gra	you	18	gra
behind	22	context	nothing	13	gra	your	4	gra
beside	14	gra	now	103	gra	Total	24976	
besides	2	gra	of	301	gra			
between	1	gra	on	610	gra			
but	787	gra	once	1	context			
by	85	gra	one	298	context			
can	44	gra	onto	2	gra			
couldn't	1	gra	or	37	gra			
did	1	gra	our	5	gra			
didn't	2	gra	out	5	gra			
do	10	context	outside	1	context			
doing	7	context	over	11	context			
don't	48	gra	quarter	1	context			
down	6	context	(they')re	64	gra			
each	3	context	really	9	gra			
eight	8	context	right	525	context			
eighteen	3	context	seem	1	gra			
eleven	1	context	seven	22	context			
every	19	gra	seventeen	2	context			
few	6	gra	several	1	gra			
fifteen	6	context	she	214	gra			
fifty	6	context	should	1	gra			
first	36	context	since	2	gra			
firstly	2	gra	six	41	context			
five	64	context	sixteen	4	context			
for	36	gra	sixth	1	gra			
forty	6	context	so	120	context			
four	99	context	some	35	gra			
fourteen	2	context	someone	2	gra			
fourth	5	context	something	7	gra			
from	32	gra	somewhere	2	gra			
had	4	gra	such	2	gra			
hardly	1	gra	ten	3	context			
has	124	gra	than	11	gra			
hasn	1	gra	that	32	context			
have	324	gra	the	2809	gra			
haven	1	gra	their	20	gra			
having	2	gra	them	83	gra			
he	71	gra	then	15	gra			
hello	5	gra	there	1479	gra			
her	370	gra	these	6	context			
here	11	gra	they	440	gra			
hers	6	gra	third	16	gra			
hi	1	gra	thirteen	3	context			
him	9	gra	thirty	1	context			
his	52	gra	this	120	context			
how	22	gra	those	2	context			
however	2	gra	three	69	context			
i	1037	gra	thus	1	gra			
if	2	gra	to	313	gra			
in	735	gra	twelve	3	context			

Note. gra = grammatical words; context = word that is either grammatical or lexical depending on the context.

Appendix 11.4

Raters' linguistic background history (Raters who evaluated accuracy of speaking performance)

	Rater A	Rater C	Rater D
Is English your first language (L1)?	No.	Yes.	Yes.
(For a non-native speaker) Which level of English are you at broadly, advanced, intermediate, novice?	Advanced.		
What other languages have you learned? (For each language) Which level are you at broadly, advanced, intermediate, novice?	German (novice), Chinese (novice), Japanese sign language (novice).	Japanese (advanced), Chinese (intermediate), French (intermediate), Vetnamese (novice), Tamil (novice).	French (advanced), Japanese (intermediate).
(For a native speaker) How long have you been in Japan? OR How long were you in Japan?		For 14 years. I have worked as an English teacher.	For 12 years. I have worked as an English teacher.
(For a non-native speaker) How long have you been in countries where English is spoken?	I went there for a few weeks five times.		

Note. Rater B is the same person of Rater C in Appendix 9.4. Rater E is the same person of Rater D in Appendix 9.4.

Appendix 11.5

Manual for raters who evaluated accuracy

Please underline the part where there is an error. Even when there are some errors in one words, just underline once.

e.g., There are two people. They are playing bike. Girl is very old. Nice view. This sky is blue and white. Lake is very beautiful.

■Error guidelines

Errors includes morphological, syntactic, lexical, and discourse errors. Look at the task pictures when you evaluate errors.

Criteria:

Whether an utterance is correct as English (NOT Whether an utterance is acceptable)

Whether an utterance corresponds to task contents

(If you understand when you see a task picture, please consider it correct.)

1) Do not count spelling errors or typing errors (including word changes like "there/their" and punctuations, such as comma errors).

2) Be conservative about counting tense errors on preceding discourse. But do not look at the sentence in isolation.

Correct: A man and a woman are cycling in the mountain. *The woman* has long hair.

3) Be lenient about oral expressions that occur even among proficient speakers often (see examples

in the rectangles). (base on McCarthy & Carter, 2001)

· further specification of the house is given in posthead appositional items
Correct: Yeah it's a big house, *six bedrooms* (p. 55)
Correct: It's a large house, *lovely, just right* (p. 55)

· the use of progressive forms with verbs considered to be unamenable to progressive concetext (e.g., want, like, have to). adopting an indirect or nonassertive stance
Reason : One … feature of the real-listener relationship is tentativeness and indirectness, a politeness strategy that minimizes imposition and threat to face (Brown & Levinson, 1987). This often manifests itself in tense and aspect choices
Correct: [Telephone inquiry to travel agent]
Customer: Oh, hello, my husband and I *are wanting* to go to the Hook of Holland next weekend. (p. 58)

· in oral narrative, … speakers exercise considerable liberty in tense and aspect choice for the dramatization of events, or for their foregrounding and backgrounding. the variation and rate of change from one form to another tends to be more intense in spoken narratives [than written narratives]. (pp. 58-59)
Correct: [Speaker 1 is telling a story about how difficult it was to buy his favorite ice cream, called *Magnum*, in a small, provincial English town]
Speaker 1: So *we're looking* in there and *we can't find* any Magnums so *we turn round* and *he actually interrupts* his phone call to say you know what you looking for and *we said* have you got any Magnums [Speaker 2: Mm] and *he sort of shook his head* in a way as to say no you know we don't get such things it was a complete rejection [Speaker 3: Yeah] and we, *we sort of took a step* back from the thing and *there it was* labeled Magnum.
Correct: A girl *wash* the cup. She *is helping* her mother.
Correct: Before I *went* to school, an apple *was* on the desk, but after I *come* back school, my apple *is* eaten by Jiro.

· Some warns against incorrect placement of adverbials between verb and direct object (e.g., **She speaks very well English.*). However, in certain spoken and written registers, most notably journalism, this … "rule" is regularly contradicted (pp. 60-61)
Correct: Mr. [name] said he will fight *vigorously* attemps to extradite him to Britain.

· Adverbials may occur after tags, and adverbs not normally considered amenable to final placement in written text regularly occur clause-finally (p. 61)
Correct: Spanish is more widely used isn't it *outside of Europe?*
Correct: You know which one I mean *probably*.

· fronted objects (p. 61)
Correct: *Those pipes* he said he's already disconnected; *the others* he's going to disconnect.

· content matter is placed outside of the core clausal positions, in the form of what have traditionally been determined left- and right-displaced or left- and right-dislocated elements, or pre-posted elements. (pp. 61-62)
Correct: *Paul*, in this job that he's got now, when *he* goes into the office *he*'s never quite sure where he's going to be sent. [content for the subject]
Correct: *A friend of mine, his* uncle had the taxi firm when we had the wedding. [attribute of the subject]
Correct: Well, *this little story I was going to tell you about,* I was on holiday with an elderly friend of mine in Butlins, Barry Island, South Wales, as you know, and she asked me … [a broad topical framework not necessarily repeated in any subsequent element]

·after conventional clause elements have been exhausted, further linguistic matter may arise on the record (p. 62)
Correct: An *he*'s quite a comic *the fellow*, you know.

·double negative (p. 65)
Correct: It should fit there, cos it's *not* that big I *don't* think.

·conditional clause complexes that challenge the rule that excludes a modal verb from the conditional clause (p. 65)
Correct: If I'*d have* stopped I probably would have wondered what she was going to say. (Instead of *if I had stopped* …)

McCarthy, M., & Carter, R. (2001). Ten criteria for a spoken grammar. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 51-75). Mahwah, NJ: Lawrence Erlbaum.

4) Be lenient about article errors from translations of proper nouns.

5) Don't count register errors related to lexical choices (e.g., lots, kids).

6) Disregard an unfinished sentence at the end of the utterances.

Appendix 11.6
Correlations between 29 Speaking Performance Measures of five tasks

	T1 F3	T1 F4	T1 F5	T1 F6	T1 A1	T1 A2	T1 SC1	T1 SC2	T1 SC3	T1 SC4	T1 LC1	T1 LC2	T1 LC3	T1 LC4
T1F1	.88	.27	-.13	-.05	.06	.01	.25	.22	.23	.40	.75	.02	-.16	-.51
T1F3	--	.18	-.16	-.14	.12	.03	.45	.42	.42	.07	.65	-.24	-.52	-.47
T1F4		--	.83	.88	-.02	-.02	-.05	-.08	-.07	.11	.22	.01	.04	-.20
T1F5			--	.96	-.10	-.09	-.08	-.10	-.10	-.06	-.12	-.02	.03	-.01
T1F6				--	-.10	-.10	-.05	-.07	-.07	.12	-.04	.11	.14	-.04
T1A1					--	.97	.23	.19	.18	.02	.16	.10	-.06	-.19
T1A2						--	.08	.05	.03	-.03	.11	.07	.01	-.15
T1SC1							--	.95	.95	.33	.28	.27	-.35	-.11
T1SC2								--	.99	.29	.27	.25	-.32	-.06
T1SC3									--	.29	.28	.25	-.32	-.04
T1SC4										--	.42	.83	.59	-.14
T1LC1											--	.40	.19	-.15
T1LC2												--	.80	.20
T1LC3													--	.27
T1LC4														--
T1LC5														
T1LC6														
T1LC7														

Note. n = 225. when r = | .13| to | .16|, $p < .05$; when r = | .17| or more, $p < .01$.

	T1 LC5	T1 LC6	T1 LC7	T1 LC8	T1 L11X	T1 L12X	T1 L13X	T1 L14X	T1 L15X	T1 L11Y	T1 L12Y	T1 L13Y	T1 L14Y	T1 L15Y
T1F1	.62	.00	-.16	-.05	-.07	.15	.03	-.01	.02	-.31	.53	.05	-.08	.06
T1F3	.53	-.25	-.49	-.05	-.02	.17	-.01	-.08	.06	-.32	.43	-.18	-.39	.01
T1F4	.11	-.04	-.01	-.11	-.11	-.06	-.08	-.08	-.09	-.16	.08	-.03	-.01	-.08
T1F5	-.16	-.05	.00	-.08	-.07	-.11	-.09	-.09	-.09	-.09	-.17	-.09	-.04	-.15
T1F6	-.09	.07	.10	-.07	-.08	-.09	-.05	-.05	-.09	-.06	-.08	.04	.07	-.09
T1A1	-.12	-.10	-.25	-.38	-.02	-.01	-.02	-.07	-.05	-.12	-.04	-.05	-.17	-.22
T1A2	-.15	-.12	-.17	-.34	-.05	-.05	-.06	-.08	-.08	-.11	-.08	-.08	-.13	-.23
T1SC1	.17	.20	-.37	-.11	.15	.21	.22	.04	.16	.00	.23	.22	-.24	.04
T1SC2	.18	.20	-.33	-.08	.18	.23	.25	.08	.19	.05	.24	.23	-.20	.06
T1SC3	.20	.21	-.32	-.08	.20	.25	.26	.09	.20	.06	.26	.23	-.20	.07
T1SC4	.32	.74	.50	-.07	.01	.12	.28	.22	.04	.11	.44	.75	.58	.21
T1LC1	.72	.27	.07	-.22	.07	.25	.18	.13	.06	.06	.69	.32	.17	.03
T1LC2	.22	.84	.63	-.19	.10	.13	.32	.27	.04	.38	.36	.81	.68	.10
T1LC3	.08	.68	.84	-.13	.01	.00	.18	.25	-.06	.37	.19	.65	.81	.08
T1LC4	.32	.55	.59	.63	.41	.29	.35	.37	.31	.61	.12	.31	.36	.28
T1LC5	--	.50	.36	.51	.28	.43	.36	.33	.30	.25	.75	.38	.27	.37
T1LC6		--	.82	.37	.28	.30	.48	.44	.24	.51	.46	.84	.73	.37
T1LC7			--	.41	.19	.16	.33	.40	.14	.48	.29	.67	.84	.33
T1LC8				--	.32	.30	.30	.31	.37	.27	.20	.13	.17	.49
T1LC11X					--	.96	.94	.94	.97	.45	.32	.30	.23	.38
T1LC12X						--	.95	.93	.97	.39	.47	.34	.24	.41
T1LC13X							--	.97	.94	.46	.44	.50	.40	.43
T1LC14X								--	.93	.47	.40	.47	.46	.43
T1LC15X									--	.37	.34	.27	.20	.42
T1LC11Y										--	.62	.73	.73	.78
T1LC12Y											--	.70	.57	.73
T1LC13Y												--	.88	.65
T1LC14Y													--	.63
T1LC15Y														--
T2F1														
T2F3														
T2F4														
T2F5														
T2F6														
T2A1														

	T2 F1	T2 F3	T2 F4	T2 F5	T2 F6	T2 A1	T2 A2
T1F1	.63	.50	.35	.10	.19	-.13	-.14
T1F3	.54	.44	.25	.03	.10	-.09	-.11
T1F4	.38	.32	.43	.26	.30	-.02	.02
T1F5	.09	.11	.22	.18	.18	.04	.06
T1F6	.17	.14	.30	.23	.25	.03	.06
T1A1	.15	.15	.09	.04	.09	.23	.22
T1A2	.13	.15	.08	.03	.08	.22	.23
T1SC1	.10	.08	.08	.02	.08	.08	.04
T1SC2	.09	.06	.08	.02	.07	.06	.03
T1SC3	.08	.06	.07	.02	.07	.08	.04
T1SC4	.27	.18	.26	.14	.23	-.06	-.05
T1LC1	.49	.39	.23	.04	.14	-.02	-.05
T1LC2	.05	.01	.09	.06	.14	.05	.05
T1LC3	-.01	-.04	.04	.03	.08	.00	.03
T1LC4	-.44	-.39	-.20	.00	-.07	.05	.09
T1LC5	.26	.17	.21	.13	.17	-.11	-.09
T1LC6	-.07	-.12	.08	.12	.16	-.03	.01
T1LC7	-.13	-.16	.03	.09	.09	-.07	-.01
T1LC8	-.22	-.25	-.01	.12	.05	-.15	-.08
T1LC11X	-.08	-.10	-.05	.03	-.02	-.11	-.03
T1LC12X	.07	.01	.01	.03	.01	-.12	-.06
T1LC13X	-.01	-.05	.01	.07	.04	-.12	-.04
T1LC14X	-.02	-.06	.01	.06	.03	-.14	-.06
T1LC15X	-.02	-.07	-.01	.05	.00	-.13	-.05
T1LC11Y	-.21	-.17	-.12	-.05	-.08	-.03	.00
T1LC12Y	.31	.24	.19	.07	.11	-.12	-.10
T1LC13Y	.03	.01	.09	.06	.10	-.06	-.03
T1LC14Y	-.01	-.03	.04	.05	.06	-.09	-.05
T1LC15Y	-.02	-.02	.04	.05	.01	-.17	-.11
T2F1	--	.79	.41	-.01	.16	-.14	-.14
T2F3		--	.31	.00	.02	-.07	-.08
T2F4			--	.84	.90	.00	.04
T2F5				--	.91	-.01	.06
T2F6					--	.01	.06
T2A1						--	.84

	T2 SC1	T2 SC2	T2 SC3	T2 SC4	T2 LC1	T2 LC2	T2 LC3	T2 LC4	T2 LC5	T2 LC6	T2 LC7	T2 LC8	T2 L11X	T2 L12X	T2 L13X	T2 L14X
T1F1	.15	.	.	.33	.26	-.03	-.13	-.39	.08	-.08	-.14	-.12	-.18	.00	-.04	-.12
T1F3	.11	.	.	.25	.25	-.03	-.10	-.31	.10	-.06	-.10	-.08	-.09	.08	.04	-.06
T1F4	.15	.	.	.20	.10	-.07	-.12	-.36	-.18	-.23	-.20	-.30	-.09	-.01	-.05	-.11
T1F5	.06	.	.	.03	.05	-.05	-.06	-.15	-.15	-.16	-.13	-.21	-.02	-.01	-.03	-.07
T1F6	.09	.	.	.11	.04	-.03	-.08	-.22	-.18	-.18	-.15	-.25	-.07	-.04	-.07	-.11
T1A1	.29	.	.	.19	.01	.08	-.13	-.20	-.14	-.03	-.18	-.18	.05	.12	.13	.01
T1A2	.29	.	.	.16	-.01	.06	-.14	-.18	-.15	-.05	-.19	-.17	.06	.12	.13	.02
T1SC1	.12	.	.	.12	.03	.07	-.10	-.08	.00	.05	-.10	-.02	.01	.03	.05	-.09
T1SC2	.09	.	.	.10	.06	.07	-.07	-.03	.05	.09	-.06	.02	.05	.08	.09	-.01
T1SC3	.10	.	.	.10	.05	.08	-.07	-.02	.05	.09	-.06	.02	.06	.08	.10	-.01
T1SC4	.16	.	.	.26	.02	.07	-.12	-.23	-.05	.01	-.15	-.06	-.16	-.11	-.11	-.16
T1LC1	.17	.	.	.29	.21	.01	-.16	-.30	.08	-.04	-.17	-.09	-.10	.04	.01	-.09
T1LC2	.14	.	.	.16	-.05	.09	-.11	-.08	-.04	.06	-.12	-.01	-.08	-.08	-.07	-.11
T1LC3	.05	.	.	.08	-.07	.05	-.04	-.02	-.03	.03	-.05	.02	-.07	-.09	-.08	-.04
T1LC4	-.23	.	.	-.27	-.15	-.02	.11	.40	.09	.11	.16	.24	.11	-.03	-.03	.08
T1LC5	-.04	.	.	.12	.14	-.03	-.04	-.07	.16	.02	-.01	.07	-.08	-.01	-.06	-.06
T1LC6	-.01	.	.	.04	-.10	.04	-.03	.08	.03	.09	-.02	.12	-.06	-.11	-.11	-.09
T1LC7	-.10	.	.	-.04	-.11	.00	.03	.13	.03	.06	.05	.14	-.05	-.12	-.13	-.01
T1LC8	-.27	.	.	-.19	-.09	-.07	.15	.26	.11	.06	.20	.21	.01	-.07	-.10	.03
T1LC11X	-.13	.	.	-.12	-.02	-.09	.07	.13	.11	-.02	.09	.13	.09	.05	.02	.04
T1LC12X	-.09	.	.	-.04	.05	-.09	.03	.03	.12	-.03	.06	.10	.04	.05	.01	.02
T1LC13X	-.07	.	.	-.06	-.02	-.08	.02	.06	.08	-.02	.04	.10	.05	.03	.00	.01
T1LC14X	-.11	.	.	-.08	-.02	-.09	.05	.08	.09	-.02	.07	.12	.06	.04	.00	.04
T1LC15X	-.15	.	.	-.10	.01	-.10	.08	.08	.10	-.04	.10	.10	.07	.05	.01	.05
T1LC11Y	-.11	.	.	-.19	-.11	-.10	-.05	.22	.11	.01	.01	.21	.19	.15	.13	.14
T1LC12Y	.04	.	.	.11	.13	-.08	-.14	-.12	.16	-.04	-.11	.08	.02	.13	.09	.02
T1LC13Y	.04	.	.	.04	-.05	-.02	-.11	.00	.05	.02	-.09	.10	.02	.03	.02	-.02
T1LC14Y	-.02	.	.	-.01	-.08	-.05	-.07	.05	.05	.00	-.04	.12	.03	.02	.01	.04
T1LC15Y	-.10	.	.	-.11	-.04	-.14	-.06	.10	.13	-.04	.00	.18	.13	.15	.11	.12
T2F1	.30	.	.	.58	.30	-.05	-.23	-.68	.01	-.17	-.29	-.22	-.27	.02	-.01	-.18
T2F3	.59	.	.	.17	.18	-.36	-.63	-.58	-.02	-.40	-.64	-.16	-.10	.15	.05	-.27
T2F4	.18	.	.	.32	.13	.04	-.09	-.39	-.14	-.12	-.17	-.27	-.23	-.12	-.13	-.18
T2F5	.10	.	.	.05	-.02	.04	-.04	-.13	-.17	-.07	-.09	-.18	-.16	-.16	-.15	-.17
T2F6	.16	.	.	.35	.03	.24	.01	-.26	-.18	.07	-.07	-.23	-.26	-.22	-.18	-.21
T2A1	.01	.	.	-.13	.01	-.04	-.06	.07	-.08	-.12	-.07	-.13	.17	.12	.11	.12
T2A2	.01	.	.	-.10	-.03	-.04	-.05	.02	-.16	-.16	-.10	-.19	.14	.08	.08	.06
T2SC1	--	.	.	.33	-.09	.10	-.79	-.41	-.23	-.01	-.81	-.20	-.11	.01	.08	-.46
T2SC2		--	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3			--	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4				--	.15	.66	.07	-.51	-.11	.44	-.03	-.24	-.42	-.24	-.07	-.27
T2LC1					--	.42	.27	.10	.55	.28	.20	-.15	-.01	.09	.07	.09
T2LC2						--	.39	.09	.15	.80	.30	-.15	-.23	-.25	-.04	-.13
T2LC3							--	.33	.23	.37	.95	.06	-.01	-.10	-.08	.39
T2LC4								--	.64	.47	.52	.68	.35	.15	.11	.33
T2LC5									--	.57	.44	.74	.17	.24	.16	.24
T2LC6										--	.47	.45	-.12	-.13	.02	-.03
T2LC7											--	.35	.05	-.04	-.04	.44
T2LC8												--	.17	.18	.10	.18
T2L11X													--	.92	.88	.79
T2L12X														--	.94	.75
T2L13X															--	.73
T2L14X																--
T2L15X																

	T2 L15X		T2 L11Y	T2 L12Y	T2 L13Y	T2 L14Y	T2 L15Y	T3 F1	T3 F3	T3 F4	T3 F5	T3 F6	T3 A1	T3 A2	T3 SC1	T3 SC2
T1F1	-.03		-.37	.13	-.05	-.12	-.07	.62	.62	.31	.06	.10	.19	.15	.08	.08
T1F3	.05		-.29	.15	-.04	-.08	-.04	.52	.55	.25	.03	.05	.21	.17	.07	.08
T1F4	-.02		-.35	-.14	-.20	-.17	-.28	.34	.28	.47	.30	.34	-.07	-.06	-.07	-.08
T1F5	-.01		-.16	-.14	-.15	-.11	-.23	.07	.03	.25	.22	.23	-.17	-.16	-.09	-.10
T1F6	-.05		-.22	-.16	-.16	-.14	-.25	.14	.08	.32	.24	.27	-.15	-.14	-.09	-.10
T1A1	.12		-.21	-.13	-.03	-.17	-.19	.18	.16	-.02	-.10	-.10	.33	.29	.09	.09
T1A2	.12		-.19	-.15	-.06	-.17	-.20	.16	.13	-.03	-.10	-.10	.29	.25	.08	.07
T1SC1	.05		-.10	-.02	.03	-.11	-.05	.15	.17	.03	-.05	-.05	.17	.16	.08	.10
T1SC2	.09		-.05	.03	.05	-.07	-.01	.12	.16	.01	-.05	-.07	.14	.13	.09	.10
T1SC3	.10		-.04	.03	.06	-.07	-.01	.11	.15	.01	-.05	-.07	.14	.14	.08	.09
T1SC4	-.10		-.25	-.05	.00	-.14	-.08	.34	.28	.17	.00	.06	.06	.08	.08	.09
T1LC1	.03		-.31	.10	-.04	-.16	-.08	.53	.45	.20	-.02	.03	.20	.19	.03	.07
T1LC2	-.06		-.11	-.08	.03	-.13	-.06	.13	.03	.01	-.07	-.01	.00	.04	.04	.08
T1LC3	-.08		-.04	-.07	.02	-.06	-.03	.03	-.08	-.01	-.03	.01	-.11	-.06	-.01	.01
T1LC4	-.02		.38	.03	.06	.13	.19	-.34	-.38	-.20	-.05	-.07	-.21	-.14	-.13	-.09
T1LC5	-.03		-.07	.16	.00	-.01	.09	.37	.32	.17	.03	.05	.04	.06	-.04	-.01
T1LC6	-.09		.04	-.03	.04	-.03	.06	.05	-.02	.00	-.03	.00	-.10	-.04	-.03	.01
T1LC7	-.11		.11	-.01	.03	.04	.10	-.05	-.12	-.02	.00	.03	-.20	-.13	-.08	-.06
T1LC8	-.07		.27	.10	.04	.19	.23	-.14	-.11	-.01	.06	.03	-.20	-.14	-.11	-.12
T1LC11X	.04		.11	.08	-.04	.09	.12	.01	.00	-.05	-.03	-.06	.01	.03	-.08	-.03
T1LC12X	.03		.02	.11	-.04	.06	.10	.16	.14	.00	-.03	-.05	.04	.05	-.04	.00
T1LC13X	.03		.04	.05	-.05	.03	.08	.11	.07	-.02	-.04	-.06	.01	.04	-.07	-.02
T1LC14X	.03		.07	.07	-.04	.06	.11	.09	.05	-.01	-.02	-.04	-.02	.02	-.09	-.04
T1LC15X	.04		.08	.09	-.05	.10	.11	.06	.06	-.02	.00	-.04	.02	.04	-.06	-.02
T1LC11Y	.15		.20	.06	-.02	-.02	.18	-.13	-.17	-.15	-.11	-.11	-.12	-.08	.01	.07
T1LC12Y	.12		-.12	.16	-.04	-.11	.10	.40	.35	.14	-.03	.00	.07	.08	.07	.11
T1LC13Y	.04		-.02	.02	.00	-.10	.08	.14	.07	.01	-.07	-.04	-.04	.01	.06	.10
T1LC14Y	.03		.03	.02	-.01	-.05	.10	.05	-.02	-.01	-.04	-.01	-.13	-.08	.01	.05
T1LC15Y	.14		.11	.12	-.04	-.01	.21	.05	.06	.00	-.03	-.03	-.09	-.06	.04	.07
T2F1	-.03		-.67	.08	-.13	-.25	-.15	.70	.66	.41	.13	.19	.17	.12	.12	.10
T2F3	.12		-.59	.01	-.39	-.61	-.15	.57	.59	.32	.10	.12	.10	.06	.06	.05
T2F4	-.14		-.41	-.14	-.13	-.15	-.29	.44	.37	.49	.28	.33	.11	.10	.01	-.03
T2F5	-.16		-.15	-.20	-.10	-.09	-.23	.19	.14	.29	.22	.25	.05	.06	-.05	-.08
T2F6	-.22		-.29	-.21	.03	-.07	-.29	.30	.20	.33	.20	.25	.12	.12	-.02	-.06
T2A1	.12		.05	-.13	-.15	-.09	-.19	-.13	-.16	-.07	-.06	-.06	.23	.26	.00	-.04
T2A2	.09		-.01	-.20	-.17	-.11	-.25	-.09	-.13	-.05	-.06	-.05	.22	.26	.02	.01
T2SC1	.03		-.44	-.25	-.03	-.80	-.25	.23	.24	.08	-.02	.00	.13	.10	.09	.09
T2SC2	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	-.26		-.51	-.07	.46	-.02	-.23	.41	.30	.22	.03	.12	.19	.16	.16	.15
T2LC1	-.08		.18	.66	.37	.24	-.01	.33	.34	.11	-.01	.00	.07	.02	.16	.13
T2LC2	-.31		.13	.20	.84	.30	-.11	.04	-.02	-.07	-.10	-.06	.12	.10	.18	.17
T2LC3	-.15		.39	.29	.42	.96	.15	-.15	-.19	-.07	.02	.00	-.04	-.02	-.01	-.03
T2LC4	.11		.94	.49	.38	.45	.56	-.46	-.41	-.36	-.16	-.22	-.20	-.16	-.05	-.05
T2LC5	.12		.58	.85	.48	.37	.64	.03	.04	-.14	-.11	-.14	-.11	-.13	.05	.04
T2LC6	-.19		.42	.45	.92	.41	.36	-.11	-.15	-.21	-.16	-.15	-.01	-.02	.12	.12
T2LC7	-.09		.53	.41	.46	.98	.36	-.21	-.25	-.14	-.01	-.04	-.09	-.06	-.03	-.05
T2LC8	.17		.54	.48	.28	.25	.79	-.24	-.21	-.25	-.13	-.16	-.20	-.18	-.06	-.05
T2L11X	.92		.30	.07	-.18	.01	.08	-.16	-.11	-.13	-.03	-.11	.04	.04	-.08	-.09
T2L12X	.97		.10	.15	-.18	-.07	.10	.04	.07	-.05	-.04	-.09	.10	.08	-.04	-.06
T2L13X	.93		.07	.09	-.02	-.07	.03	-.01	.02	-.08	-.06	-.11	.13	.11	.02	.00
T2L14X	.72		.31	.18	-.05	.41	.14	-.13	-.12	-.09	.01	-.05	.05	.05	-.04	-.08
T2L15X	--		.05	.02	-.25	-.12	.07	.00	.01	-.06	-.03	-.08	.10	.09	-.07	-.08
T2L11Y				.61	.46	.52	.64	-.45	-.38	-.32	-.12	-.18	-.18	-.15	-.02	-.03
T2L12Y				--	.57	.45	.74	.08	.13	-.06	-.05	-.06	-.06	-.08	.13	.09
T2L13Y					--	.47	.43	-.08	-.09	-.16	-.12	-.10	.02	.01	.18	.16
T2L14Y						--	.39	-.19	-.21	-.10	.02	-.01	-.07	-.05	-.01	-.03
T2L15Y							--	-.19	-.12	-.17	-.05	-.08	-.16	-.15	.03	.01
T3F1								--	.88	.32	-.11	-.02	.16	.12	.08	.07
T3F3									--	.31	-.06	-.04	.13	.06	.26	.20
T3F4										--	.82	.85	.00	-.01	.02	.04
T3F5											--	.95	-.08	-.08	-.04	-.02

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T2 L15X		T2 L11Y	T2 L12Y	T2 L13Y	T2 L14Y	T2 L15Y	T3 F1	T3 F3	T3 F4	T3 F5	T3 F6	T3 A1	T3 A2	T3 SC1	T3 SC2
T3F6												--	-.05	-.06	.05	.08
T3A1													--	.95	.18	.10
T3A2														--	.05	-.03
T3SC1															--	.90
T3SC2																--
T3SC3																
T3SC4																
T3LC1																
T3LC2																

	T3 SC3	T3 SC4	T3 LC1	T3 LC2	T3 LC3	T3 LC4	T3 LC5	T3 LC6	T3 LC7	T3 LC8	T3 L11X	T3 L12X	T3 L13X	T3 L14X	T3 L15X	T3 L11Y
T1F1	.16	.09	.50	-.16	-.19	-.35	.30	-.20	-.21	-.13	-.24	.04	-.19	-.20	-.14	-.38
T1F3	.16	.02	.43	-.17	-.19	-.30	.24	-.22	-.22	-.14	-.15	.09	-.14	-.15	-.06	-.32
T1F4	-.05	.06	.34	-.03	.00	-.11	.21	-.02	.01	-.05	-.08	.07	-.03	-.03	-.05	-.04
T1F5	-.08	-.01	.11	-.01	.04	.05	.10	.04	.08	.05	.01	.03	.02	.02	-.01	.12
T1F6	-.08	.04	.16	.02	.06	.02	.14	.07	.10	.04	-.01	.04	.03	.03	-.02	.09
T1A1	.11	.10	.15	.01	-.04	-.21	-.04	-.11	-.13	-.20	-.05	.01	-.04	-.06	-.04	-.26
T1A2	.09	.09	.13	.02	-.02	-.18	-.06	-.09	-.11	-.20	-.04	-.01	-.04	-.06	-.04	-.23
T1SC1	.13	.05	.06	-.03	-.07	-.12	.01	-.05	-.08	-.05	.05	.08	.05	.03	.08	-.10
T1SC2	.12	.02	.06	-.04	-.09	-.09	.01	-.07	-.09	-.05	.05	.07	.03	.01	.06	-.06
T1SC3	.12	.00	.06	-.05	-.09	-.07	.01	-.07	-.09	-.05	.06	.08	.03	.02	.07	-.05
T1SC4	.10	.22	.21	.04	-.02	-.20	.14	.02	-.02	-.03	-.18	-.04	-.08	-.08	-.12	-.24
T1LC1	.14	.16	.46	-.06	-.05	-.31	.25	-.15	-.11	-.16	-.19	.07	-.12	-.12	-.11	-.32
T1LC2	.07	.24	.07	.14	.12	-.08	.04	.11	.10	-.02	-.08	-.03	.02	.03	-.06	-.08
T1LC3	-.02	.19	.02	.15	.15	.00	.03	.14	.14	.01	-.11	-.09	-.02	.00	-.11	-.02
T1LC4	-.12	-.05	-.31	.08	.16	.30	-.06	.19	.23	.23	.16	.01	.14	.19	.11	.39
T1LC5	.06	.06	.27	-.11	-.06	-.13	.27	-.07	-.03	.06	-.12	.07	-.09	-.06	-.04	-.10
T1LC6	.00	.15	-.03	.09	.10	.05	.09	.15	.16	.14	-.03	-.02	.04	.07	-.01	.07
T1LC7	-.08	.09	-.08	.09	.13	.12	.07	.17	.19	.16	-.06	-.08	.00	.04	-.06	.13
T1LC8	-.11	-.12	-.20	-.08	-.01	.21	.06	.09	.11	.29	.07	.02	.05	.09	.09	.26
T1LC11X	-.03	-.12	.02	-.09	-.02	.04	.02	-.08	-.02	.02	.05	.05	-.01	.03	.04	.09
T1LC12X	.02	-.07	.15	-.11	-.06	-.04	.11	-.11	-.07	.00	-.02	.06	-.05	-.02	.00	-.01
T1LC13X	-.01	-.07	.08	-.10	-.03	-.01	.07	-.09	-.02	.03	.00	.05	-.03	.01	.01	.03
T1LC14X	-.04	-.09	.07	-.11	-.02	.01	.07	-.08	-.01	.04	-.01	.03	-.05	.00	-.01	.05
T1LC15X	-.02	-.12	.06	-.12	-.06	.02	.05	-.10	-.05	.03	.02	.05	-.04	.00	.02	.05
T1LC11Y	.04	.07	-.11	.11	.12	.15	-.02	.14	.15	.10	.05	-.02	.08	.10	.02	.21
T1LC12Y	.16	.14	.32	-.04	-.05	-.18	.22	-.06	-.05	-.03	-.16	.02	-.09	-.09	-.10	-.16
T1LC13Y	.09	.19	.07	.09	.07	-.04	.08	.10	.09	.03	-.09	-.04	.00	.01	-.07	-.03
T1LC14Y	.03	.14	.03	.10	.10	.03	.07	.12	.13	.06	-.10	-.08	-.02	.00	-.10	.04
T1LC15Y	.06	.02	.00	-.01	-.02	.05	.07	.05	.04	.11	-.04	-.02	-.01	.01	-.02	.08
T2F1	.16	.17	.60	-.08	-.16	-.40	.27	-.18	-.23	-.24	-.23	.06	-.16	-.20	-.13	-.43
T2F3	.11	.01	.55	-.14	-.18	-.33	.24	-.23	-.24	-.23	-.12	.13	-.11	-.14	-.05	-.34
T2F4	-.01	.12	.34	-.07	-.07	-.30	.12	-.15	-.12	-.17	-.20	-.02	-.14	-.15	-.13	-.25
T2F5	-.07	.03	.11	-.09	-.03	-.17	.03	-.12	-.05	-.07	-.14	-.07	-.12	-.12	-.11	-.13
T2F6	-.05	.13	.20	-.03	.00	-.23	.07	-.09	-.04	-.11	-.17	-.05	-.11	-.11	-.11	-.18
T2A1	-.03	.02	-.04	.07	.02	.06	-.05	.04	.00	-.03	.04	-.03	.02	.00	-.01	.05
T2A2	.01	.07	.01	.11	.06	.06	-.01	.08	.04	-.02	.02	-.03	.03	.01	-.03	.07
T2SC1	.10	.04	.28	-.01	-.06	-.17	.10	-.08	-.10	-.15	.12	.20	.11	.06	.13	-.16
T2SC2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	.16	.29	.34	.09	-.02	-.27	.14	.01	-.08	-.16	-.11	.06	-.02	-.06	-.04	-.29
T2LC1	.15	.09	.32	-.05	-.17	-.21	.14	-.12	-.21	-.14	-.20	-.04	-.17	-.23	-.16	-.24
T2LC2	.15	.21	.03	.14	.02	-.06	.00	.10	-.01	-.04	-.02	-.01	.04	-.01	-.02	-.07
T2LC3	-.05	.05	-.20	.05	.04	.11	-.08	.08	.06	.10	-.13	-.17	-.09	-.08	-.13	.09
T2LC4	-.09	-.12	-.48	.03	.04	.30	-.20	.13	.12	.23	.18	-.03	.14	.14	.13	.35
T2LC5	.04	.02	-.06	-.05	-.11	-.02	.00	-.02	-.07	.06	-.01	.03	.00	-.04	.04	-.02
T2LC6	.09	.15	-.17	.12	.02	.05	-.06	.14	.04	.08	.08	.03	.12	.08	.09	.05
T2LC7	-.07	.02	-.29	.03	.03	.15	-.11	.10	.08	.15	-.08	-.16	-.05	-.04	-.08	.14
T2LC8	-.07	-.05	-.32	-.01	.01	.16	-.09	.09	.09	.19	.16	.07	.15	.15	.18	.18

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T3 SC3	T3 SC4	T3 LC1	T3 LC2	T3 LC3	T3 LC4	T3 LC5	T3 LC6	T3 LC7	T3 LC8	T3 L11X	T3 L12X	T3 L13X	T3 L14X	T3 L15X	T3 L11Y
T2L11X	-.10	-.19	-.19	-.15	-.09	.09	-.09	-.11	-.07	.07	.12	.06	.02	.02	.13	.08
T2L12X	-.05	-.12	-.01	-.14	-.12	-.04	-.01	-.14	-.12	.00	.07	.11	.00	-.01	.11	-.08
T2L13X	-.01	-.08	-.05	-.10	-.10	-.01	-.03	-.10	-.10	.01	.10	.11	.04	.01	.14	-.06
T2L14X	-.09	-.11	-.21	-.09	-.07	.10	-.09	-.05	-.04	.09	-.05	-.09	-.08	-.09	-.04	.06
T2L15X	-.08	-.12	-.07	-.14	-.09	-.01	-.04	-.13	-.08	.02	.09	.11	.03	.03	.14	-.06
T2L11Y	-.07	-.12	-.45	.03	.02	.27	-.20	.11	.08	.19	.16	-.03	.13	.12	.11	.34
T2L12Y	.10	.03	.02	-.04	-.16	-.08	.00	-.05	-.14	-.02	-.05	.01	-.03	-.09	.00	-.05
T2L13Y	.13	.16	-.11	.13	-.02	.01	-.06	.12	-.01	.02	.05	.02	.10	.04	.06	.03
T2L14Y	-.05	.02	-.25	.03	.01	.12	-.10	.08	.05	.12	-.09	-.15	-.06	-.05	-.09	.13
T2L15Y	.01	-.03	-.25	.00	-.06	.09	-.11	.05	.00	.10	.13	.06	.13	.10	.15	.16
T3F1	.14	.23	.74	-.19	-.23	-.55	.43	-.25	-.27	-.18	-.33	.11	-.22	-.24	-.16	-.57
T3F3	.28	-.05	.66	-.38	-.54	-.49	.37	-.42	-.53	-.16	-.31	.07	-.31	-.39	-.15	-.50
T3F4	.08	.00	.31	-.10	-.10	-.15	.17	-.12	-.11	-.07	-.15	.02	-.13	-.14	-.09	-.11
T3F5	.00	-.17	-.03	-.08	-.04	.12	.00	-.05	-.01	.05	-.05	-.08	-.10	-.10	-.07	.16
T3F6	.10	.06	.05	.08	.07	.05	.03	.08	.07	.00	-.07	-.06	-.03	-.05	-.07	.10
T3A1	.12	.13	.12	.04	-.08	-.32	-.19	-.17	-.26	-.36	-.07	-.02	-.03	-.09	-.05	-.34
T3A2	-.02	.07	.08	.00	-.05	-.28	-.18	-.18	-.20	-.31	-.05	-.01	-.03	-.05	-.03	-.29
T3SC1	.88	.40	.12	.36	-.38	-.10	-.06	.22	-.41	-.19	-.09	-.06	.06	-.23	-.10	-.06
T3SC2	.98	.41	.13	.38	-.21	-.02	.00	.27	-.22	-.12	-.13	-.10	.02	-.20	-.16	.00
T3SC3	--	.38	.20	.33	-.24	-.06	.04	.21	-.25	-.13	-.14	-.08	.00	-.22	-.16	-.04
T3SC4		--	.23	.83	.50	-.20	.01	.62	.34	-.22	-.07	.01	.36	.24	-.07	-.20
T3LC1			--	.14	.04	-.22	.60	.01	-.06	-.23	-.17	.19	-.08	-.11	-.17	-.30
T3LC2				--	.68	.21	.01	.83	.53	-.13	.15	.05	.50	.40	.00	.18
T3LC3					--	.28	.05	.62	.88	.02	.16	.04	.39	.54	.03	.22
T3LC4						--	.44	.61	.60	.76	.41	.17	.32	.36	.24	.86
T3LC5							--	.35	.35	.63	.07	.32	.08	.10	.09	.24
T3LC6								--	.75	.42	.28	.16	.56	.51	.15	.50

	T3 L12Y	T3 L13Y	T3 L14Y	T3 L15Y	T4 F1	T4 F3	T4 F4	T4 F5	T4 F6	T4 A1	T4 A2	T4 SC1	T4 SC2	T4 SC3	T4 SC4	T4 LC1	T4 LC2
T1F1	.23	-.24	-.25	-.19	.70	.65	.34	.08	.11	.11	.04	.28	.35	.36	.17	.50	-.14
T1F3	.20	-.24	-.25	-.17	.59	.57	.28	.06	.09	.07	.02	.25	.35	.35	.13	.43	-.12
T1F4	.30	.03	.06	.05	.42	.30	.53	.32	.34	-.03	-.06	.06	.17	.19	.17	.22	-.06
T1F5	.20	.11	.14	.15	.08	.00	.29	.24	.24	-.08	-.09	-.08	.03	.05	.06	-.02	-.02
T1F6	.24	.13	.16	.14	.16	.06	.38	.29	.29	-.06	-.08	-.06	.07	.09	.09	.03	-.03
T1A1	-.12	-.16	-.19	-.25	.20	.20	.02	-.07	-.05	.27	.26	.19	.03	.01	.18	.20	.12
T1A2	-.13	-.14	-.16	-.24	.18	.16	.01	-.08	-.05	.29	.27	.17	.01	-.01	.19	.18	.14
T1SC1	.03	-.05	-.07	-.03	.16	.22	.04	-.03	-.02	.13	.12	.13	.16	.16	-.01	.13	-.05
T1SC2	.04	-.05	-.08	-.01	.16	.19	.02	-.05	-.04	.09	.09	.06	.15	.15	-.01	.13	-.04
T1SC3	.03	-.06	-.08	-.02	.15	.19	.01	-.06	-.05	.08	.08	.07	.16	.15	-.01	.14	-.04
T1SC4	.10	-.01	-.04	-.10	.32	.29	.13	.01	.00	.16	.12	.14	.10	.11	.06	.21	-.10
T1LC1	.24	-.15	-.12	-.16	.59	.52	.16	-.04	.00	.15	.11	.24	.22	.22	.19	.47	-.06
T1LC2	.05	.11	.10	-.03	.09	.08	-.05	-.08	-.08	.17	.16	.05	-.05	-.05	.02	.08	-.01
T1LC3	.03	.13	.13	-.01	-.02	-.06	-.08	-.06	-.07	.07	.08	-.05	-.16	-.15	.02	-.02	.00
T1LC4	.07	.27	.31	.35	-.47	-.47	-.21	.00	-.02	-.15	-.10	-.23	-.22	-.23	-.11	-.30	.14
T1LC5	.32	-.04	.00	.09	.33	.27	.16	.07	.09	-.03	-.04	.09	.14	.14	.10	.29	-.03
T1LC6	.13	.17	.18	.16	-.05	-.06	-.04	.00	-.02	.04	.05	-.04	-.08	-.08	-.04	-.03	.00
T1LC7	.10	.19	.21	.16	-.15	-.19	-.07	.01	-.01	-.05	-.03	-.13	-.17	-.16	-.04	-.12	.01
T1LC8	.15	.14	.16	.34	-.26	-.26	.03	.16	.12	-.24	-.22	-.16	-.06	-.05	-.10	-.20	.02
T1LC11X	.08	-.04	.01	.07	-.15	-.17	-.06	.02	-.01	-.09	-.03	-.17	-.11	-.11	-.09	-.09	-.02
T1LC12X	.15	-.08	-.04	.04	.01	-.03	-.01	.02	.00	-.05	.00	-.09	-.04	-.05	-.04	.05	-.04
T1LC13X	.12	-.05	.00	.07	-.06	-.09	-.04	.01	-.02	-.04	.01	-.11	-.08	-.09	-.08	-.03	-.05
T1LC14X	.12	-.05	.01	.07	-.09	-.12	-.04	.02	-.02	-.07	-.02	-.14	-.11	-.11	-.08	-.07	-.05
T1LC15X	.10	-.07	-.03	.06	-.09	-.12	-.03	.03	.00	-.09	-.04	-.15	-.08	-.08	-.07	-.04	-.03
T1LC11Y	.08	.20	.20	.19	-.26	-.30	-.20	-.10	-.10	-.16	-.11	-.20	-.22	-.22	-.05	-.15	.06
T1LC12Y	.26	-.04	-.04	.00	.34	.25	.08	-.04	-.01	-.04	-.04	.06	.06	.07	.12	.29	-.04
T1LC13Y	.11	.12	.11	.05	.03	-.01	-.06	-.07	-.08	.01	.03	-.03	-.09	-.08	.00	.03	-.02
T1LC14Y	.10	.14	.14	.07	-.05	-.12	-.09	-.07	-.08	-.06	-.04	-.11	-.16	-.15	.01	-.05	.00
T1LC15Y	.13	.09	.07	.15	-.07	-.13	-.02	-.01	-.01	-.21	-.19	-.13	-.09	-.08	-.01	-.04	-.01
T2F1	.22	-.21	-.26	-.28	.71	.61	.40	.12	.17	.10	.03	.28	.26	.28	.29	.50	-.04

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T3 L12Y	T3 L13Y	T3 L14Y	T3 L15Y	T4 F1	T4 F3	T4 F4	T4 F5	T4 F6	T4 A1	T4 A2	T4 SC1	T4 SC2	T4 SC3	T4 SC4	T4 LC1	T4 LC2
T2F3	.21	-.24	-.25	-.23	.58	.48	.28	.07	.10	.09	.04	.22	.19	.21	.27	.36	-.03
T2F4	.19	-.11	-.10	-.10	.44	.36	.51	.28	.31	.12	.08	.15	.11	.12	.19	.29	.00
T2F5	.09	-.09	-.03	-.01	.15	.12	.27	.18	.19	.09	.08	.05	-.01	-.01	.09	.09	.05
T2F6	.13	-.06	-.02	-.05	.28	.22	.34	.19	.22	.15	.14	.10	.03	.03	.16	.18	.05
T2A1	-.06	.02	-.01	-.03	-.08	-.12	-.14	-.15	-.12	.13	.18	-.09	-.08	-.08	.03	.02	.11
T2A2	.00	.07	.04	.00	-.04	-.11	-.12	-.15	-.10	.15	.21	-.03	-.07	-.07	.13	-.02	.17
T2SC1	.09	-.08	-.10	-.14	.23	.18	.05	-.04	-.02	.23	.20	.20	.12	.11	.22	.12	.10
T2SC2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	.09	-.02	-.11	-.20	.43	.38	.25	.08	.10	.20	.14	.27	.20	.20	.23	.35	.05
T2LC1	.09	-.15	-.24	-.18	.35	.28	.17	.06	.09	.03	.00	.04	.11	.12	.11	.33	-.04
T2LC2	-.03	.08	-.03	-.07	.06	.06	-.01	-.02	-.02	.16	.14	.10	.08	.07	.05	.11	.06
T2LC3	-.08	.07	.05	.08	-.11	-.07	-.02	.04	.01	-.07	-.05	-.12	-.07	-.08	-.17	-.03	-.09
T2LC4	-.13	.17	.16	.29	-.51	-.44	-.34	-.13	-.14	-.12	-.07	-.24	-.21	-.21	-.25	-.41	-.01
T2LC5	-.01	-.02	-.07	.04	-.02	-.02	-.11	-.07	-.05	-.06	-.04	-.06	-.03	-.02	-.03	-.02	-.04
T2LC6	-.07	.13	.04	.05	-.12	-.08	-.16	-.11	-.11	.09	.10	.02	.00	-.01	-.05	-.06	.03
T2LC7	-.11	.09	.08	.14	-.19	-.13	-.09	.01	-.01	-.10	-.07	-.15	-.10	-.10	-.20	-.11	-.10
T2LC8	-.07	.11	.11	.21	-.30	-.24	-.25	-.13	-.13	-.11	-.06	-.10	-.11	-.11	-.13	-.28	-.01
T2L11X	-.12	-.12	-.07	.06	-.18	-.14	-.16	-.10	-.09	-.13	-.06	-.11	-.07	-.07	-.16	-.18	-.14
T2L12X	-.09	-.18	-.15	-.07	.02	.04	-.06	-.08	-.07	-.09	-.05	.01	.02	.03	-.07	-.02	-.13
T2L13X	-.12	-.14	-.14	-.07	-.02	.02	-.09	-.10	-.09	-.07	-.03	.03	.04	.04	-.09	-.02	-.12
T2L14X	-.14	-.08	-.06	.04	-.11	-.06	-.10	-.06	-.06	-.16	-.13	-.09	-.04	-.04	-.16	-.07	-.14
T2L15X	-.11	-.16	-.11	-.04	-.03	.01	-.10	-.10	-.09	-.07	-.03	.00	.01	.01	-.09	-.07	-.14
T2L11Y	-.11	.17	.14	.28	-.49	-.41	-.33	-.13	-.15	-.14	-.09	-.26	-.22	-.22	-.25	-.36	-.02
T2L12Y	.04	-.03	-.12	.02	.06	.05	-.07	-.06	-.05	-.06	-.06	-.06	-.03	-.02	-.03	.07	-.05
T2L13Y	-.05	.12	.00	.03	-.07	-.03	-.14	-.10	-.11	.09	.09	.02	-.01	-.01	-.04	.01	.03
T2L14Y	-.08	.08	.06	.12	-.15	-.10	-.07	.01	-.01	-.10	-.07	-.15	-.10	-.10	-.19	-.07	-.10
T2L15Y	-.02	.11	.05	.19	-.23	-.17	-.23	-.13	-.14	-.12	-.09	-.12	-.14	-.13	-.13	-.19	-.03
T3F1	.38	-.28	-.30	-.23	.74	.63	.43	.12	.17	.17	.11	.26	.23	.24	.27	.54	-.07
T3F3	.34	-.42	-.55	-.19	.66	.64	.37	.09	.11	.13	.06	.31	.30	.30	.14	.47	-.14
T3F4	.24	-.09	-.08	-.01	.34	.29	.57	.38	.40	.00	-.05	.12	.08	.09	.13	.27	-.02
T3F5	.07	-.01	.03	.11	.04	.02	.32	.30	.30	-.12	-.15	-.02	-.05	-.04	.02	.03	.00
T3F6	.10	.12	.11	.07	.08	.03	.39	.35	.36	-.10	-.14	.01	-.04	-.03	.10	.08	.07
T3A1	-.25	-.19	-.29	-.38	.22	.26	-.02	-.08	-.07	.32	.29	.26	.09	.07	.05	.24	-.02
T3A2	-.23	-.20	-.22	-.32	.19	.23	-.05	-.09	-.08	.33	.31	.21	.03	.02	.01	.19	-.04
T3SC1	-.02	.24	-.38	-.12	.03	.04	.04	.01	.01	.08	.06	.17	.15	.13	.05	.13	.12
T3SC2	.04	.29	-.20	-.07	.03	.03	.04	.02	.03	.02	.00	.16	.17	.15	.07	.14	.15
T3SC3	.07	.23	-.24	-.09	.08	.08	.07	.02	.03	.02	.00	.17	.20	.17	.08	.19	.14
T3SC4	.02	.61	.34	-.20	.16	.01	.12	.05	.12	.08	.08	.04	.00	.01	.35	.20	.29
T3LC1	.50	-.06	-.12	-.32	.62	.51	.35	.07	.12	.21	.16	.23	.20	.21	.29	.56	.03
T3LC2	-.01	.79	.53	-.13	-.12	-.21	-.06	-.04	.00	.03	.05	-.03	-.04	-.04	.23	.00	.31
T3LC3	.01	.58	.84	-.03	-.14	-.23	-.08	-.03	.02	-.06	-.04	-.12	-.12	-.10	.19	-.07	.23
T3LC4	.32	.53	.54	.56	-.40	-.36	-.23	-.07	-.09	-.13	-.14	-.17	-.09	-.08	-.15	-.30	.06
T3LC5	.77	.21	.23	.32	.34	.26	.20	.07	.10	.06	-.01	.09	.11	.12	.16	.29	.03
T3LC6	.26	.92	.71	.28	-.20	-.27	-.09	-.01	.02	-.02	-.02	-.08	-.04	-.03	.15	-.10	.27

	T4 LC3	T4 LC4	T4 LC5	T4 LC6	T4 LC7	T4 LC8	T4 L11X	T4 L12X	T4 L13X	T4 L14X	T4 L15X	T4 L11Y
T1F1	-.23	-.31	.41	-.06	-.15	.08	-.34	-.04	-.25	-.30	-.20	-.27
T1F3	-.15	-.25	.36	-.04	-.07	.08	-.28	-.03	-.20	-.23	-.17	-.19
T1F4	-.06	-.21	.21	-.01	-.03	.07	-.26	-.11	-.20	-.16	-.17	-.12
T1F5	.05	-.07	.01	-.02	.05	.01	-.14	-.12	-.12	-.04	-.12	.07
T1F6	.03	-.10	.05	-.01	.04	.03	-.17	-.13	-.15	-.07	-.14	.03
T1A1	-.08	-.11	.12	.09	-.10	-.06	-.09	-.02	-.01	-.11	-.09	-.08
T1A2	-.08	-.11	.08	.09	-.12	-.09	-.04	.00	.04	-.07	-.06	-.10
T1SC1	-.04	-.07	.08	-.03	-.01	-.02	-.09	-.02	-.10	-.09	-.06	.01
T1SC2	.01	-.04	.11	.00	.05	.03	-.08	-.01	-.09	-.05	-.04	.03
T1SC3	.01	-.03	.12	.01	.05	.03	-.07	.00	-.08	-.05	-.03	.04
T1SC4	-.18	-.15	.17	-.07	-.14	.02	-.18	-.04	-.16	-.21	-.10	-.16
T1LC1	-.18	-.21	.41	.02	-.11	.11	-.28	-.02	-.19	-.25	-.17	-.17

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T4 LC3	T4 LC4	T4 LC5	T4 LC6	T4 LC7	T4 LC8	T4 L11X	T4 L12X	T4 L13X	T4 L14X	T4 L15X	T4 L11Y
T1LC2	-.09	.00	.07	.00	-.08	.02	-.07	-.03	-.07	-.10	-.05	-.03
T1LC3	-.05	.03	.00	.00	-.06	.03	-.02	-.03	-.02	-.04	-.02	-.04
T1LC4	.22	.37	-.18	.15	.19	.07	.26	.08	.21	.28	.16	.31
T1LC5	-.06	-.01	.31	.08	.01	.16	-.12	.06	-.07	-.07	-.05	-.02
T1LC6	-.01	.13	.03	.04	.01	.08	.03	.03	.01	.01	.04	.08
T1LC7	.03	.16	-.03	.05	.03	.09	.07	.03	.05	.07	.06	.06
T1LC8	.13	.23	-.08	.08	.15	.10	.17	.09	.13	.19	.14	.17
T1LC11X	.13	.09	-.08	-.03	.11	-.01	.02	-.02	-.02	.08	-.02	.07
T1LC12X	.06	.02	.03	-.03	.06	.01	-.05	-.02	-.05	.01	-.05	.03
T1LC13X	.05	.05	-.03	-.05	.05	.01	-.02	-.02	-.04	.02	-.03	.05
T1LC14X	.06	.06	-.05	-.04	.06	.02	-.01	-.02	-.04	.03	-.02	.03
T1LC15X	.11	.05	-.06	-.03	.09	-.01	-.01	-.03	-.03	.05	-.04	.04
T1LC11Y	.15	.29	-.01	.15	.17	.16	.17	.09	.14	.20	.14	.23
T1LC12Y	-.06	-.01	.32	.09	.02	.19	-.12	.06	-.07	-.07	-.04	-.01
T1LC13Y	-.03	.08	.09	.04	.01	.11	.00	.04	-.01	-.01	.04	.05
T1LC14Y	.01	.11	.03	.05	.02	.12	.03	.04	.03	.04	.05	.03
T1LC15Y	.08	.17	.08	.10	.13	.17	.09	.10	.08	.13	.11	.10
T2F1	-.19	-.39	.35	-.01	-.15	-.01	-.39	-.13	-.26	-.32	-.27	-.31
T2F3	-.17	-.32	.27	.00	-.13	.02	-.33	-.12	-.21	-.27	-.22	-.24
T2F4	-.11	-.22	.22	.01	-.09	.01	-.26	-.10	-.17	-.23	-.18	-.19
T2F5	-.02	-.08	.06	.03	-.03	-.02	-.11	-.06	-.06	-.10	-.08	-.06
T2F6	-.05	-.16	.11	.03	-.06	-.03	-.16	-.05	-.08	-.13	-.10	-.13
T2A1	.12	.07	-.04	.05	.07	-.08	.08	.03	.06	.08	.03	.06
T2A2	.11	-.04	-.10	.08	.04	-.13	.07	.02	.10	.10	.04	-.04
T2SC1	-.09	-.21	.03	.06	-.12	-.08	-.09	-.01	.00	-.10	-.04	-.12
T2SC2	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	-.13	-.31	.17	.01	-.14	-.11	-.20	-.01	-.09	-.18	-.11	-.25
T2LC1	-.02	-.15	.25	-.01	.01	.02	-.19	-.01	-.13	-.09	-.12	-.12
T2LC2	-.01	-.11	-.01	.00	-.04	-.13	.01	.08	.05	.01	.04	-.09
T2LC3	.02	.05	-.03	-.10	.02	-.02	.05	.04	-.02	.03	.03	-.03
T2LC4	.14	.32	-.26	.00	.13	.06	.25	.05	.13	.21	.16	.27
T2LC5	.03	.07	.06	.03	.07	.12	-.02	.00	-.05	.00	-.01	.06
T2LC6	.02	.02	-.08	.01	.01	-.05	.10	.09	.07	.06	.10	.03
T2LC7	.03	.11	-.06	-.10	.05	.02	.08	.04	-.01	.05	.05	.03
T2LC8	.05	.22	-.13	.05	.08	.13	.12	.00	.04	.07	.08	.19
T2L11X	-.04	.04	-.08	-.12	-.02	.07	.04	.02	-.03	.00	.06	.02
T2L12X	-.10	-.03	.06	-.07	-.04	.10	-.04	.02	-.06	-.07	.01	-.04
T2L13X	-.11	-.06	.01	-.10	-.07	.03	.00	.04	-.04	-.06	.04	-.04
T2L14X	-.05	.06	.02	-.11	.00	.09	.03	.03	-.04	-.01	.04	.02
T2L15X	-.11	-.03	.02	-.09	-.06	.09	-.02	.02	-.05	-.06	.03	-.03
T2L11Y	.14	.31	-.22	.00	.13	.07	.24	.06	.13	.23	.16	.26
T2L12Y	.01	.04	.12	.03	.06	.12	-.05	.00	-.06	.00	-.03	.04
T2L13Y	.01	.00	-.03	.01	.00	-.05	.07	.09	.06	.06	.07	.02
T2L14Y	.02	.10	-.03	-.09	.04	.02	.06	.03	-.02	.04	.03	.02
T2L15Y	.04	.19	-.06	.04	.07	.13	.10	.01	.03	.08	.06	.16
T3F1	-.22	-.38	.38	-.03	-.18	.00	-.36	-.05	-.23	-.30	-.22	-.31
T3F3	-.29	-.30	.35	-.08	-.22	.03	-.34	-.07	-.25	-.34	-.22	-.24
T3F4	-.10	-.19	.18	-.03	-.09	-.02	-.29	-.17	-.23	-.25	-.24	-.18
T3F5	-.01	-.07	.01	-.03	-.02	-.02	-.19	-.18	-.18	-.16	-.19	-.08
T3F6	.02	-.10	.01	.01	-.02	-.07	-.21	-.19	-.16	-.16	-.21	-.09
T3A1	-.20	-.15	.10	-.05	-.20	-.10	-.12	-.01	-.07	-.20	-.08	-.13
T3A2	-.21	-.14	.06	-.08	-.21	-.10	-.09	.01	-.07	-.19	-.05	-.12
T3SC1	-.01	.00	.01	.08	-.03	-.13	-.05	-.05	.00	-.08	-.10	-.01
T3SC2	.03	.04	.03	.10	.01	-.11	-.06	-.06	-.02	-.07	-.12	.03
T3SC3	.03	.01	.07	.10	.00	-.11	-.10	-.07	-.04	-.09	-.14	.04
T3SC4	.17	-.16	.04	.19	.10	-.16	-.07	-.01	.10	.07	-.07	-.16
T3LC1	-.17	-.27	.38	.04	-.15	-.02	-.26	.01	-.14	-.23	-.18	-.24
T3LC2	.20	.01	-.11	.20	.11	-.16	.10	.04	.19	.17	.03	-.04
T3LC3	.21	.01	-.11	.14	.13	-.08	.11	.04	.16	.19	.06	-.01
T3LC4	.13	.34	-.14	.08	.12	.12	.25	.06	.12	.17	.15	.24
T3LC5	-.06	-.02	.29	.09	-.02	.13	-.14	.00	-.11	-.13	-.09	-.06
T3LC6	.22	.11	-.12	.19	.15	-.07	.13	.02	.16	.17	.05	.02

	T4 L12 Y	T4 L13 Y	T4 L14 Y	T4 L15 Y	T5 F1	T5 F3	T5 F4	T5 F5	T5 F6	T5 A1	T5 A2	T5 SC1	T5 SC2	T5 SC3	T5 SC4	T5 LC1	T5 LC2
T1F1	.42	-.04	-.13	.11	.64	.54	.40	.10	.18	.14	.17	.18	.14	.15	.29	.41	.05
T1F3	.39	-.01	-.05	.14	.52	.44	.32	.06	.12	.14	.19	.12	.11	.12	.20	.36	.02
T1F4	.28	.05	.02	.18	.34	.26	.46	.28	.37	.11	.04	.15	.16	.18	.26	.22	.14
T1F5	.12	.08	.11	.18	.05	.01	.22	.19	.24	.05	-.01	.04	.05	.07	.09	-.01	.07
T1F6	.16	.08	.09	.19	.11	.06	.27	.21	.29	.07	-.01	.09	.11	.13	.16	.06	.12
T1A1	.13	.10	-.09	-.02	.06	.16	.10	.12	.11	.15	.20	.13	-.04	-.03	.05	.09	.05
T1A2	.08	.08	-.12	-.07	.08	.14	.10	.11	.09	.13	.18	.10	-.07	-.06	.05	.07	.04
T1SC1	.15	.01	.02	.07	.04	.10	.05	-.01	.02	.13	.17	.10	.06	.06	.04	.11	.07
T1SC2	.17	.03	.07	.10	.03	.06	.04	-.01	.01	.14	.18	.05	.05	.05	.01	.10	.05
T1SC3	.17	.04	.07	.10	.01	.05	.03	-.01	.01	.14	.18	.04	.05	.05	.01	.10	.04
T1SC4	.16	-.07	-.14	.01	.27	.27	.15	.05	.10	.06	.05	.21	.09	.10	.21	.19	.11
T1LC1	.42	.05	-.09	.14	.44	.41	.25	.03	.08	.14	.17	.19	.12	.14	.21	.42	.12
T1LC2	.06	-.01	-.08	.00	.00	.06	-.02	-.02	.01	.01	.01	.16	.04	.05	.11	.09	.14
T1LC3	-.05	-.03	-.08	-.05	-.03	.00	-.07	-.01	-.01	-.07	-.11	.10	-.01	.00	.07	.01	.09
T1LC4	-.19	.13	.18	.02	-.43	-.40	-.25	-.04	-.12	-.14	-.14	-.19	-.07	-.09	-.22	-.24	-.04
T1LC5	.32	.09	.02	.17	.29	.22	.18	.06	.06	.02	.05	.01	.06	.07	.09	.25	.01
T1LC6	.02	.02	-.01	.04	-.07	-.05	-.05	.02	.00	-.06	-.06	.04	.01	.00	.02	.00	.06
T1LC7	-.09	.00	.00	-.02	-.09	-.11	-.08	.02	-.01	-.13	-.16	-.02	-.03	-.04	-.01	-.07	.01
T1LC8	-.10	.05	.13	.04	-.13	-.19	-.05	.05	-.02	-.14	-.12	-.21	-.04	-.08	-.14	-.16	-.14
T1LC11X	-.06	-.03	.12	.00	-.16	-.12	-.11	.05	-.02	.00	.03	-.11	-.01	.00	-.13	-.03	-.01
T1LC12X	.07	-.01	.08	.04	-.01	.01	-.04	.04	.00	.02	.05	-.06	.00	.01	-.06	.06	-.01
T1LC13X	.00	-.04	.06	.02	-.08	-.04	-.08	.04	-.01	-.02	.02	-.06	-.02	-.01	-.08	.02	.00
T1LC14X	-.04	-.05	.06	.00	-.09	-.06	-.08	.07	.00	-.03	.00	-.07	-.02	-.01	-.09	-.01	-.01
T1LC15X	-.03	-.03	.10	-.01	-.09	-.06	-.08	.04	-.01	-.01	.03	-.10	-.01	-.01	-.10	.00	-.03
T1LC11Y	-.02	.12	.15	.10	-.23	-.21	-.20	-.12	-.16	-.12	-.13	-.10	-.05	-.05	-.13	-.01	.04
T1LC12Y	.34	.10	.03	.19	.31	.25	.15	-.02	.00	.01	.03	.05	.05	.06	.12	.32	.06
T1LC13Y	.09	.03	.00	.08	.03	.05	-.04	-.05	-.04	-.05	-.06	.07	.02	.02	.06	.11	.09
T1LC14Y	.00	.02	.00	.03	.00	.00	-.07	-.04	-.05	-.12	-.16	.03	-.02	-.01	.03	.06	.06
T1LC15Y	.06	.07	.11	.10	.03	-.01	-.03	-.05	-.07	-.13	-.13	-.10	-.03	-.03	-.05	.08	-.03
T2F1	.41	.04	-.12	.09	.71	.60	.45	.15	.24	.16	.14	.27	.19	.21	.40	.43	.14
T2F3	.33	.06	-.10	.11	.57	.52	.35	.10	.16	.10	.11	.25	.16	.18	.30	.36	.10
T2F4	.23	.03	-.08	.04	.47	.40	.50	.29	.36	.19	.17	.21	.15	.18	.29	.24	.09
T2F5	.06	.04	-.03	.00	.18	.17	.32	.27	.29	.11	.10	.12	.09	.09	.14	.04	.03
T2F6	.13	.05	-.05	.01	.28	.24	.36	.27	.32	.18	.16	.16	.10	.11	.22	.11	.07
T2A1	-.05	.05	.06	-.08	-.13	-.09	-.09	-.06	-.02	.01	.03	.06	-.02	-.02	.03	.00	.11
T2A2	-.12	.06	.04	-.13	-.07	-.07	-.08	-.05	-.01	.03	.04	.06	-.01	-.01	.09	.03	.16
T2SC1	.10	.11	-.09	.02	.20	.22	.15	.09	.13	.11	.14	.18	.01	.02	.19	.08	.08
T2SC2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	.22	.04	-.12	-.03	.39	.29	.25	.13	.18	.18	.14	.14	.03	.04	.31	.18	.12
T2LC1	.27	.02	.03	.06	.24	.15	.18	.10	.13	.04	.12	.04	.08	.09	.13	.26	.10
T2LC2	.01	.01	-.04	-.10	-.03	-.09	-.01	.06	.07	.09	.10	-.06	-.08	-.08	.07	-.06	.06
T2LC3	-.10	-.15	.00	-.11	-.13	-.21	-.09	-.03	-.06	.00	.00	-.23	-.07	-.08	-.13	-.12	-.10
T2LC4	-.29	-.03	.12	.00	-.50	-.50	-.30	-.08	-.14	-.22	-.18	-.29	-.11	-.12	-.27	-.26	-.06
T2LC5	.05	.04	.08	.11	-.02	-.15	-.03	.00	-.01	-.19	-.13	-.17	-.01	.01	.00	.08	.04
T2LC6	-.07	.01	.01	-.03	-.15	-.26	-.11	-.01	-.01	-.07	-.06	-.19	-.12	-.11	.01	-.14	.03
T2LC7	-.12	-.14	.03	-.06	-.18	-.28	-.13	-.05	-.09	-.06	-.07	-.28	-.08	-.09	-.15	-.13	-.10
T2LC8	-.14	.03	.07	.08	-.20	-.29	-.17	-.08	-.12	-.26	-.25	-.23	-.08	-.07	-.10	-.12	-.04
T2L11X	-.11	-.13	-.02	.02	-.26	-.19	-.15	-.03	-.05	-.11	-.08	-.10	-.07	-.08	-.18	-.16	-.09
T2L12X	.04	-.08	-.05	.07	-.05	-.03	-.04	-.01	-.01	-.08	-.05	-.04	-.04	-.04	-.06	-.02	-.05
T2L13X	.02	-.09	-.06	.04	-.08	-.05	-.06	.00	.00	-.04	-.02	-.03	-.06	-.07	-.05	-.07	-.05
T2L14X	-.03	-.13	-.02	.02	-.14	-.12	-.10	-.03	-.06	-.05	-.08	-.10	-.06	-.07	-.13	-.11	-.10
T2L15X	.00	-.09	-.06	.07	-.08	-.05	-.06	-.02	-.03	-.08	-.06	-.05	-.05	-.05	-.09	-.07	-.07
T2L11Y	-.26	-.02	.12	.00	-.48	-.48	-.31	-.09	-.14	-.22	-.16	-.29	-.10	-.12	-.29	-.23	-.07
T2L12Y	.13	.04	.07	.11	.04	-.08	-.02	-.01	-.01	-.16	-.09	-.14	.02	.03	.01	.14	.04
T2L13Y	-.02	.02	.01	-.04	-.12	-.21	-.11	-.01	-.01	-.05	-.03	-.17	-.10	-.10	.01	-.09	.03
T2L14Y	-.09	-.13	.01	-.07	-.16	-.25	-.12	-.05	-.09	-.04	-.04	-.27	-.07	-.08	-.15	-.11	-.10
T2L15Y	-.07	.03	.07	.08	-.16	-.23	-.18	-.10	-.13	-.24	-.22	-.22	-.04	-.05	-.09	-.05	-.04
T3F1	.43	.02	-.15	.08	.63	.55	.39	.15	.22	.13	.11	.24	.16	.19	.32	.49	.15
T3F3	.40	-.04	-.20	.09	.57	.51	.34	.14	.19	.13	.15	.20	.12	.13	.27	.41	.10

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T4 L12 Y	T4 L13 Y	T4 L14 Y	T4 L15 Y	T5 F1	T5 F3	T5 F4	T5 F5	T5 F6	T5 A1	T5 A2	T5 SC1	T5 SC2	T5 SC3	T5 SC4	T5 LC1	T5 LC2
T3F4	.19	-.01	-.09	-.01	.33	.24	.60	.44	.52	.07	.04	.18	.13	.15	.33	.16	.15
T3F5	.00	-.04	-.03	-.03	.05	.00	.41	.41	.44	.01	-.01	.06	.04	.05	.16	-.07	.06
T3F6	.03	.02	-.01	-.05	.12	.05	.46	.41	.48	.03	.00	.11	.08	.09	.25	.01	.12
T3A1	.11	-.04	-.20	-.07	.15	.10	.12	.06	.06	.25	.31	.04	-.02	-.02	.11	.14	.08
T3A2	.09	-.06	-.20	-.06	.10	.05	.08	.05	.04	.24	.31	.02	-.04	-.04	.07	.10	.07
T3SC1	.03	.07	-.05	-.12	.08	.04	.01	.01	.05	.09	.14	.02	.01	.01	.15	.12	.13
T3SC2	.05	.10	.00	-.09	.07	.01	.03	.03	.09	.04	.10	.02	-.01	.00	.18	.10	.16
T3SC3	.12	.12	.01	-.04	.13	.06	.06	.04	.11	.08	.13	.05	.01	.02	.21	.11	.16
T3SC4	.06	.19	.10	-.12	.18	.14	.09	.02	.08	.05	.02	.19	.18	.18	.26	.24	.24
T3LC1	.42	.08	-.14	.05	.52	.46	.30	.09	.17	.10	.12	.21	.14	.16	.29	.49	.17
T3LC2	-.11	.17	.09	-.16	-.05	-.06	-.08	-.07	-.02	.01	.00	.10	.13	.12	.14	.07	.20
T3LC3	-.11	.13	.13	-.07	-.09	-.07	-.07	-.08	-.05	-.06	-.10	.08	.11	.09	.04	.00	.10
T3LC4	-.21	.02	.09	.01	-.36	-.34	-.25	-.11	-.14	-.13	-.13	-.19	-.13	-.15	-.21	-.22	-.06
T3LC5	.26	.07	-.03	.08	.24	.21	.12	.05	.07	-.03	-.03	.05	.01	.02	.08	.30	.07
T3LC6	-.16	.13	.12	-.14	-.13	-.15	-.12	-.06	-.04	-.06	-.09	.00	.04	.02	.04	.01	.13

	T5 LC3	T5 LC4	T5 LC5	T5 LC6	T5 LC7	T5 LC8	T5 L11 X	T5 L12 X	T5 L13 X	T5 L14 X	T5 L15 X	T5 L11 Y	T5 L12 Y	T5 L13 Y	T5 L14 Y	T5 L15 Y
T1F1	-.15	-.23	.25	.05	-.14	.00	-.19	.05	-.03	-.14	-.07	-.25	.22	.01	-.14	-.03
T1F3	-.12	-.16	.22	.04	-.10	.01	-.12	.07	.00	-.08	-.03	-.17	.21	.02	-.10	.01
T1F4	-.06	-.12	.12	.11	-.05	-.02	-.08	.03	.05	-.03	-.04	-.16	.06	.06	-.07	-.08
T1F5	.01	-.01	.00	.06	.01	.01	.01	.01	.06	.04	.01	-.04	-.03	.03	-.01	-.03
T1F6	.01	-.03	.05	.11	.02	.02	-.03	.00	.05	.01	-.02	-.07	-.01	.06	-.01	-.06
T1A1	-.15	-.15	-.11	-.09	-.21	-.21	-.10	-.06	-.08	-.19	-.09	-.17	-.13	-.10	-.20	-.23
T1A2	-.13	-.18	-.13	-.11	-.19	-.22	-.10	-.06	-.08	-.17	-.08	-.18	-.13	-.11	-.18	-.21
T1SC1	-.08	.05	.08	.07	-.06	.01	-.01	.01	.01	-.06	-.02	.01	.04	.04	-.08	-.04
T1SC2	-.04	.07	.10	.06	-.02	.04	-.01	-.01	-.02	-.06	-.03	.04	.06	.04	-.04	.00
T1SC3	-.03	.07	.09	.06	-.01	.03	-.02	-.02	-.03	-.07	-.04	.04	.06	.03	-.02	-.01
T1SC4	-.13	-.08	.13	.10	-.10	.03	-.14	-.04	-.05	-.17	-.10	-.16	.05	.03	-.15	-.09
T1LC1	-.14	-.08	.26	.12	-.12	.02	-.17	-.02	-.06	-.20	-.12	-.14	.19	.06	-.15	-.06
T1LC2	-.08	.06	.08	.13	-.05	.03	-.08	-.10	-.06	-.16	-.12	-.03	-.03	.05	-.10	-.10
T1LC3	-.03	.03	.02	.08	-.02	.03	-.07	-.10	-.06	-.12	-.09	-.03	-.06	.02	-.06	-.08
T1LC4	.18	.32	.00	.08	.21	.19	.17	-.02	.02	.14	.07	.30	-.03	.08	.18	.15
T1LC5	-.02	.05	.27	.12	.03	.16	-.05	.05	-.01	-.03	-.01	.00	.22	.08	.00	.10
T1LC6	.01	.15	.11	.13	.06	.14	.00	-.04	-.02	-.04	-.03	.08	.02	.07	.00	.02
T1LC7	.06	.13	.06	.09	.09	.14	.01	-.05	-.03	-.01	-.02	.08	.00	.04	.05	.05
T1LC8	.15	.18	.08	.02	.21	.22	.14	.09	.05	.20	.14	.18	.10	.04	.19	.23
T1LC1 1X	.08	.22	.11	.08	.12	.16	.17	.11	.11	.14	.14	.16	.05	.05	.09	.10
T1LC1 2X	.03	.16	.16	.10	.07	.16	.13	.13	.11	.09	.13	.09	.10	.05	.04	.08
T1LC1 3X	.04	.18	.12	.09	.08	.15	.14	.11	.10	.09	.13	.12	.06	.04	.04	.07
T1LC1 4X	.05	.17	.11	.08	.09	.15	.15	.11	.10	.10	.13	.12	.05	.04	.06	.08
T1LC1 5X	.06	.19	.14	.08	.11	.18	.18	.16	.14	.15	.18	.13	.08	.05	.07	.11
T1LC1 1Y	.06	.26	.11	.13	.08	.14	.09	-.01	.00	.04	.01	.26	.08	.12	.05	.12
T1LC1 2Y	-.08	.03	.29	.14	-.04	.12	-.09	.02	-.03	-.09	-.05	.00	.24	.10	-.07	.07
T1LC1 3Y	-.05	.10	.15	.14	-.02	.10	-.05	-.05	-.04	-.09	-.07	.06	.08	.09	-.06	.02
T1LC1 4Y	-.02	.09	.11	.11	.00	.10	-.03	-.05	-.04	-.07	-.05	.06	.06	.07	-.03	.04
T1LC1 5Y	.02	.12	.17	.09	.05	.16	.06	.08	.03	.08	.06	.14	.18	.10	.04	.17
T2F1	-.17	-.27	.27	.14	-.14	-.01	-.22	.06	-.01	-.15	-.07	-.32	.21	.07	-.17	-.07
T2F3	-.19	-.20	.23	.10	-.16	-.01	-.23	-.03	-.08	-.24	-.14	-.21	.21	.07	-.17	-.03
T2F4	-.10	-.22	.13	.07	-.07	-.02	-.14	.06	.04	-.09	-.01	-.26	.08	.05	-.09	-.09
T2F5	-.04	-.15	.01	.01	-.02	-.02	-.07	.03	.05	-.04	.03	-.17	-.02	.01	-.03	-.06
T2F6	-.07	-.18	.04	.04	-.05	-.03	-.09	.05	.07	-.05	.02	-.23	-.01	.02	-.07	-.10
T2A1	-.04	.08	-.04	.06	-.05	-.05	.02	-.05	.00	-.10	-.04	.02	-.11	.01	-.08	-.13
T2A2	.03	.01	-.05	.07	-.01	-.09	-.09	-.13	-.07	-.09	-.13	-.06	-.14	.00	-.04	-.19
T2SC1	-.16	-.20	-.06	-.01	-.18	-.16	-.15	-.07	-.01	-.21	-.10	-.21	-.08	-.02	-.19	-.17

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T5 LC3	T5 LC4	T5 LC5	T5 LC6	T5 LC7	T5 LC8	T5 L11 X	T5 L12 X	T5 L13 X	T5 L14 X	T5 L15 X	T5 L11 Y	T5 L12 Y	T5 L13 Y	T5 L14 Y	T5 L15 Y
T2SC2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T2SC4	-.08	-.29	.03	.05	-.11	-.11	-.10	.08	.09	-.01	.03	-.36	-.05	-.02	-.13	-.19
T2LC1	-.01	-.02	.18	.12	.01	.04	-.02	.07	.03	.02	.01	-.02	.20	.12	.01	.05
T2LC2	.05	-.11	-.11	-.01	.00	-.09	.06	.08	.13	.13	.11	-.14	-.13	-.02	.00	-.11
T2LC3	.17	.05	-.03	-.05	.16	.07	.12	.10	.06	.26	.14	.07	.01	-.03	.19	.12
T2LC4	.21	.32	-.06	.00	.22	.13	.24	.04	.06	.23	.12	.40	.03	.09	.25	.23
T2LC5	.15	.16	.19	.13	.19	.16	.09	.09	.07	.17	.07	.22	.27	.21	.21	.26
T2LC6	.16	.00	-.08	.02	.12	.00	.11	.08	.13	.21	.12	.03	-.03	.07	.14	.06
T2LC7	.21	.11	-.01	-.03	.21	.11	.15	.10	.06	.29	.15	.15	.06	.02	.24	.18
T2LC8	.19	.18	.06	.06	.21	.15	.12	.06	.07	.19	.08	.27	.16	.16	.24	.27
T2L11 X	-.05	.16	-.06	-.07	-.05	.04	.17	.06	.03	.06	.11	.15	-.08	-.09	-.06	.02
T2L12 X	-.08	.07	.02	-.03	-.08	.03	.11	.10	.05	.03	.10	.05	-.01	-.05	-.09	.01
T2L13 X	-.10	.00	-.07	-.07	-.11	-.04	.11	.08	.06	.03	.11	-.02	-.10	-.10	-.13	-.07
T2L14 X	-.04	.05	-.06	-.09	-.05	.02	.14	.10	.04	.07	.13	.06	-.05	-.09	-.04	.02
T2L15 X	-.08	.05	-.04	-.07	-.09	.00	.09	.06	.02	.01	.08	.03	-.06	-.09	-.09	-.02
T2L11 Y	.21	.31	-.05	-.01	.22	.12	.21	.02	.03	.20	.09	.40	.07	.09	.26	.25
T2L12 Y	.13	.11	.19	.11	.16	.13	.01	.04	.01	.10	.00	.20	.31	.21	.19	.26
T2L13 Y	.15	-.03	-.07	.01	.11	-.02	.06	.05	.09	.17	.08	.02	.00	.08	.14	.07
T2L14 Y	.20	.09	.00	-.04	.19	.10	.12	.09	.04	.26	.12	.14	.07	.01	.23	.18
T2L15 Y	.19	.15	.08	.04	.20	.12	.05	.01	.00	.12	.01	.27	.23	.18	.25	.30
T3F1	-.17	-.14	.34	.19	-.12	.05	-.15	.10	.03	-.15	-.05	-.21	.27	.11	-.15	-.04
T3F3	-.16	-.13	.30	.14	-.12	.06	-.15	.06	-.01	-.16	-.06	-.19	.24	.08	-.14	-.01
T3F4	-.07	-.16	.08	.12	-.06	-.03	-.13	.00	.05	-.06	-.05	-.22	.02	.09	-.09	-.10
T3F5	-.01	-.10	-.07	.03	-.01	-.04	-.08	-.06	.01	.00	-.04	-.14	-.11	.01	-.04	-.09
T3F6	.00	-.11	-.01	.09	.00	-.03	-.10	-.04	.05	.00	-.04	-.17	-.07	.06	-.04	-.10
T3A1	-.03	-.16	-.06	-.04	-.10	-.20	-.07	.01	.01	-.02	-.04	-.18	-.07	-.04	-.09	-.20
T3A2	-.02	-.13	-.08	-.05	-.09	-.19	-.07	-.02	-.01	-.02	-.05	-.15	-.09	-.05	-.08	-.19
T3SC1	.08	-.13	-.04	.02	.00	-.15	-.15	-.09	-.08	-.04	-.14	-.15	-.06	.00	.00	-.18
T3SC2	.09	-.11	-.04	.05	.00	-.14	-.12	-.07	-.04	-.02	-.12	-.09	-.03	.07	.01	-.12
T3SC3	.05	-.13	-.03	.05	-.02	-.14	-.14	-.08	-.04	-.05	-.12	-.12	-.02	.07	-.02	-.12
T3SC4	-.02	-.12	.07	.15	-.04	-.11	-.17	-.05	-.01	-.08	-.14	-.18	.00	.08	-.07	-.19
T3LC1	-.11	-.14	.27	.15	-.09	-.04	-.24	-.04	-.08	-.21	-.18	-.20	.20	.07	-.12	-.10
T3LC2	.07	-.07	-.08	.07	.01	-.16	-.18	-.16	-.10	-.08	-.20	-.12	-.13	.02	-.01	-.21
T3LC3	.00	.02	-.04	.06	.01	-.04	-.05	-.07	-.02	-.03	-.07	.02	-.05	.05	.00	-.04
T3LC4	.16	.21	-.09	-.04	.15	.06	.11	-.06	-.03	.10	.00	.21	-.09	-.02	.14	.07
T3LC5	-.03	.08	.28	.14	.03	.12	-.04	.05	.00	-.05	-.04	.01	.20	.07	-.01	.03
T3LC6	.11	.04	-.03	.07	.09	-.05	-.09	-.12	-.08	-.01	-.14	.00	-.08	.02	.06	-.10

	T3 LC8	T3 L11X	T3 L12X	T3 L13X	T3 L14X	T3 L15X	T3 L11Y	T3 L12Y	T3 L13Y	T3 L14Y	T3 L15Y	T4 F1	T4 F3	T4 F4
T3LC7	.47	.26	.14	.43	.59	.16	.48	.24	.67	.93	.31	-.18	-.25	-.09
T3LC8	--	.25	.20	.16	.23	.27	.61	.48	.32	.40	.73	-.19	-.18	-.08
T3L11X		--	.86	.87	.86	.94	.45	.13	.32	.31	.29	-.29	-.27	-.13
T3L12X			--	.78	.76	.92	.19	.36	.19	.17	.21	.03	-.02	.06
T3L13X				--	.94	.82	.36	.15	.60	.49	.22	-.21	-.25	-.08
T3L14X					--	.81	.39	.15	.54	.63	.25	-.22	-.26	-.09
T3L15X						--	.29	.16	.21	.21	.32	-.19	-.19	-.05
T3L11Y							--	.46	.63	.60	.79	-.48	-.45	-.17
T3L12Y								--	.38	.35	.65	.22	.11	.26
T3L13Y									--	.77	.47	-.26	-.35	-.04
T3L14Y										--	.48	-.25	-.34	-.05
T3L15Y											--	-.30	-.32	.00
T4F1												--	.85	.39
T4F3													--	.31
T4F4														--

	T4 F5	T4 F6	T4 A1	T4 A2	T4 SC1	T4 SC2	T4 SC3	T4 SC4	T4 LC1	T4 LC2	T4 LC3	T4 LC4	T4 LC5	T4 LC6	T4 LC7	T4 LC8	T4 L11X
T3LC7	-.01	.04	-.10	-.10	-.15	-.10	-.08	.13	-.13	.19	.21	.10	-.10	.13	.15	.00	.12
T3LC8	.03	.02	-.14	-.17	-.13	-.06	-.06	-.10	-.19	.00	.09	.25	-.02	.06	.12	.17	.11
T3L11X	-.06	-.05	-.08	-.07	-.07	-.03	-.03	-.04	-.19	.08	.08	.14	-.12	.08	.06	.02	.19
T3L12X	.02	.04	-.01	-.02	.04	.06	.07	.10	.06	.06	.00	.02	.10	.11	.01	.07	.02
T3L13X	-.03	.00	-.06	-.04	-.05	-.02	-.01	.09	-.13	.17	.13	.07	-.11	.15	.10	-.03	.14
T3L14X	-.03	.00	-.09	-.07	-.09	-.06	-.04	.06	-.17	.12	.12	.08	-.11	.12	.10	.01	.15
T3L15X	.00	.02	-.07	-.07	-.03	-.02	.00	.01	-.13	.07	.06	.09	-.06	.09	.05	.04	.12
T3L11Y	.01	.00	-.17	-.18	-.21	-.10	-.09	-.13	-.37	.11	.18	.36	-.20	.12	.15	.10	.27
T3L12Y	.17	.21	.01	-.06	.02	.07	.08	.18	.19	.10	.02	.04	.22	.17	.03	.13	-.10
T3L13Y	.06	.10	-.05	-.06	-.12	-.05	-.04	.16	-.16	.31	.26	.14	-.16	.23	.17	-.08	.15
T3L14Y	.05	.10	-.12	-.12	-.19	-.11	-.09	.14	-.18	.23	.26	.12	-.14	.17	.19	-.01	.14
T3L15Y	.14	.14	-.18	-.22	-.18	-.09	-.08	-.06	-.29	.08	.16	.28	-.10	.13	.16	.15	.14
T4F1	-.01	.06	.22	.18	.26	.28	.30	.31	.67	-.15	-.26	-.50	.53	-.08	-.19	.07	-.47
T4F3	-.02	-.03	.31	.21	.50	.43	.44	-.05	.58	-.39	-.59	-.43	.42	-.34	-.49	.01	-.38
T4F4	.85	.86	-.03	-.12	.21	.21	.22	.24	.29	.06	-.10	-.22	.19	.05	-.10	-.03	-.25
T4F5	--	.96	-.11	-.21	.12	.10	.10	.12	.01	.11	-.02	-.07	-.05	.05	-.05	-.09	-.12
T4F6		--	-.10	-.20	.15	.11	.11	.32	.09	.28	.07	-.12	-.02	.19	.00	-.12	-.13
T4A1			--	.91	.36	.09	.09	.07	.19	.02	-.30	-.29	-.07	-.14	-.37	-.29	-.03
T4A2				--	.14	-.06	-.06	.03	.12	-.02	-.16	-.29	-.10	-.17	-.23	-.27	-.03
T4SC1					--	.60	.58	.23	.21	.14	-.65	-.31	-.05	.00	-.69	-.29	-.10
T4SC2						--	.99	.11	.28	.04	-.23	-.10	.18	.03	-.20	-.04	-.18
T4SC3							--	.11	.30	.04	-.23	-.11	.19	.03	-.19	-.04	-.18
T4SC4								--	.28	.82	.35	-.28	.10	.70	.21	-.16	-.17
T4LC1									--	.20	.00	-.08	.71	.21	-.01	-.03	-.26
T4LC2										--	.54	.10	-.01	.83	.33	-.23	.09
T4LC3											--	.35	.09	.56	.93	.13	.08
T4LC4												--	.40	.45	.51	.65	.31
T4LC5													--	.39	.31	.68	-.18
T4LC6														--	.56	.33	.09
T4LC7															--	.45	.05
T4LC8																--	.00
T4L11X																	--
T4L12X																	
T4L13X																	

	T4 L12X	T4 L13X	T4 L14X	T4 L15X	T4 L11Y	T4 L12Y	T4 L13Y	T4 L14Y	T4 L15Y	T5 F1	T5 F3	T5 F4	T5 F5	T5 F6	T5 A1
T3LC7	.02	.12	.17	.07	.05	-.12	.11	.15	-.04	-.14	-.14	-.10	-.07	-.07	-.10
T3LC8	.01	.02	.09	.08	.17	-.09	.01	.10	.07	-.20	-.19	-.13	-.02	-.07	-.14
T3L11X	.09	.16	.14	.14	.15	-.13	.08	.06	.00	-.28	-.26	-.18	-.08	-.15	-.04
T3L12X	.07	.06	.02	.03	.03	.10	.12	.01	.07	-.01	-.05	-.02	-.03	-.07	-.01
T3L13X	.07	.17	.15	.10	.06	-.11	.14	.09	-.04	-.18	-.19	-.14	-.07	-.10	-.01
T3L14X	.07	.17	.16	.12	.07	-.12	.11	.09	-.01	-.21	-.20	-.15	-.09	-.13	-.05
T3L15X	.07	.13	.11	.10	.11	-.05	.10	.06	.04	-.18	-.20	-.12	-.05	-.11	-.04
T3L11Y	.03	.15	.22	.15	.32	-.21	.10	.15	.06	-.40	-.39	-.24	-.07	-.11	-.12
T3L12Y	-.02	-.06	-.05	-.08	.06	.25	.19	.05	.16	.18	.13	.13	.08	.09	-.03
T3L13Y	.00	.18	.21	.06	.09	-.16	.20	.16	-.10	-.16	-.19	-.11	-.04	-.02	-.06
T3L14Y	.01	.14	.22	.08	.11	-.14	.17	.19	-.01	-.18	-.19	-.10	-.05	-.05	-.10
T3L15Y	-.02	.08	.16	.09	.28	-.09	.14	.17	.14	-.25	-.26	-.11	.02	-.03	-.11
T4F1	-.07	-.32	-.36	-.28	-.48	.52	-.07	-.18	.08	.69	.60	.49	.18	.28	.25
T4F3	-.06	-.40	-.54	-.23	-.45	.37	-.36	-.49	-.03	.56	.49	.45	.15	.26	.28
T4F4	-.10	-.14	-.19	-.19	-.21	.20	.07	-.09	-.02	.39	.27	.52	.33	.41	.12
T4F5	-.12	-.05	-.09	-.12	-.09	-.06	.05	-.05	-.11	.10	.04	.32	.29	.28	.02
T4F6	-.10	.03	-.04	-.12	-.12	-.01	.21	.01	-.11	.17	.11	.35	.30	.30	.04
T4A1	.04	-.02	-.21	-.01	-.28	-.06	-.15	-.38	-.26	.10	.11	.16	.14	.13	.27
T4A2	.03	-.03	-.13	.00	-.23	-.05	-.14	-.22	-.18	.03	.05	.09	.12	.11	.21
T4SC1	-.03	.00	-.46	-.07	-.33	-.10	-.03	-.70	-.32	.26	.25	.25	.13	.17	.28
T4SC2	-.10	-.13	-.27	-.18	-.15	.08	-.02	-.22	-.12	.28	.23	.26	.08	.14	.25
T4SC3	-.10	-.13	-.26	-.18	-.16	.10	-.02	-.21	-.12	.30	.24	.26	.08	.14	.23
T4SC4	-.04	.30	.16	-.12	-.26	.10	.69	.20	-.13	.35	.32	.21	.12	.15	.07
T4LC1	.07	-.13	-.17	-.25	-.08	.68	.21	-.01	-.02	.52	.44	.34	.12	.18	.17
T4LC2	.03	.47	.35	-.03	.10	-.01	.81	.32	-.22	.06	.05	.01	.05	.03	-.03
T4LC3	-.01	.26	.61	-.03	.35	.11	.56	.92	.14	-.14	-.11	-.16	-.09	-.11	-.19
T4LC4	.11	.18	.32	.13	.77	.19	.30	.43	.35	-.37	-.33	-.30	-.18	-.26	-.13
T4LC5	.10	-.12	-.04	-.13	.20	.77	.25	.24	.39	.35	.29	.20	.01	.03	.11
T4LC6	.08	.44	.40	.01	.31	.24	.87	.48	.12	.05	.06	-.03	-.03	-.05	-.03
T4LC7	.00	.17	.57	-.02	.44	.25	.50	.96	.34	-.14	-.10	-.17	-.12	-.15	-.16
T4LC8	.05	-.06	.10	.06	.38	.40	.14	.35	.59	-.05	-.03	-.08	-.12	-.16	-.03

(Appendix 11.6 continues)

(Appendix 11.6 continued)

	T4 L12X	T4 L13X	T4 L14X	T4 L15X	T4 L11Y	T4 L12Y	T4 L13Y	T4 L14Y	T4 L15Y	T5 F1	T5 F3	T5 F4	T5 F5	T5 F6	T5 A1
T4L11X	.87	.86	.76	.94	.32	-.16	.09	.06	.01	-.29	-.31	-.30	-.22	-.24	-.17
T4L12X	--	.81	.70	.93	.12	.11	.08	.00	.08	-.03	-.09	-.13	-.14	-.14	-.13
T4L13X		--	.84	.83	.18	-.12	.41	.16	-.06	-.13	-.16	-.19	-.14	-.15	-.15
T4L14X			--	.72	.33	-.02	.39	.57	.12	-.20	-.21	-.26	-.18	-.20	-.24
T4L15X				--	.14	-.12	.01	-.01	.07	-.19	-.23	-.23	-.18	-.19	-.15
T4L11Y					--	.44	.49	.53	.68	-.30	-.30	-.26	-.15	-.17	-.07
T4L12Y						--	.43	.34	.71	.39	.30	.22	.04	.12	.15
T4L13Y							--	.56	.37	.09	.08	.00	.00	.01	.01
T4L14Y								--	.47	-.11	-.09	-.15	-.11	-.11	-.14
T4L15Y									--	.04	.00	-.02	-.06	-.02	.04
T5F1										--	.80	.42	-.04	.10	.13
T5F3											--	.29	-.07	.01	.22
T5F4												--	.80	.86	.15
T5F5													--	.92	.12
T5F6														--	.15
T5A1															--
T5A2															
T5SC1															
T5SC2															
T5SC3															
T5SC4															

	T5 A2	T5 SC1	T5 SC2	T5 SC3	T5 SC4	T5 LC1	T5 LC2	T5 LC3	T5 LC4	T5 LC5	T5 LC6	T5 LC7	T5 LC8
T3LC7	-.15	-.01	.02	.00	-.04	-.04	.04	.04	.11	.00	.06	.06	.04
T3LC8	-.16	-.15	-.13	-.14	-.18	-.12	-.08	.07	.23	.07	.03	.12	.18
T3L11X	-.02	-.17	-.10	-.12	-.23	-.15	-.10	.05	.16	-.07	-.08	.04	.01
T3L12X	.00	-.11	-.06	-.07	-.09	.09	-.03	.01	.11	.12	.02	.03	.06
T3L13X	-.02	-.07	-.02	-.04	-.09	-.04	.02	.04	.09	-.04	.00	.02	-.04
T3L14X	-.07	-.08	-.02	-.04	-.13	-.06	-.01	.03	.14	-.02	.00	.03	.01
T3L15X	-.03	-.16	-.09	-.10	-.18	-.08	-.08	.04	.14	.00	-.04	.04	.05
T3L11Y	-.12	-.24	-.17	-.18	-.24	-.26	-.08	.18	.23	-.10	-.03	.17	.07
T3L12Y	-.03	-.03	-.04	-.02	.04	.24	.04	.00	.12	.27	.15	.07	.15
T3L13Y	-.08	-.06	-.01	-.02	.00	-.02	.10	.13	.07	-.03	.07	.11	-.03
T3L14Y	-.16	-.06	-.02	-.03	-.07	-.08	.02	.06	.13	.00	.06	.08	.06
T3L15Y	-.13	-.22	-.17	-.17	-.20	-.17	-.11	.10	.24	.04	.03	.14	.18
T4F1	.20	.31	.22	.24	.43	.55	.24	-.18	-.25	.29	.19	-.17	-.06
T4F3	.26	.29	.20	.21	.38	.44	.22	-.15	-.23	.19	.15	-.17	-.10
T4F4	.04	.17	.21	.24	.32	.29	.19	-.01	-.11	.19	.20	.00	.01
T4F5	-.02	.00	.06	.08	.09	.08	.04	.07	.00	.09	.08	.09	.05
T4F6	-.02	.05	.09	.11	.14	.13	.08	.01	-.04	.10	.10	.02	.02
T4A1	.26	.09	-.04	-.03	.10	.12	.09	-.08	-.20	-.15	-.10	-.17	-.28
T4A2	.22	.07	-.07	-.05	.06	.05	.07	-.07	-.17	-.18	-.10	-.16	-.26
T4SC1	.27	.24	.13	.13	.28	.14	.17	-.17	-.27	-.07	.04	-.23	-.21
T4SC2	.21	.13	.11	.12	.21	.16	.09	-.08	-.17	.04	.04	-.10	-.08
T4SC3	.20	.13	.11	.12	.21	.18	.09	-.08	-.16	.06	.05	-.09	-.07
T4SC4	.00	.22	.13	.16	.28	.25	.17	-.15	-.21	.11	.12	-.15	-.07
T4LC1	.16	.14	.09	.12	.27	.50	.14	-.06	-.18	.26	.12	-.06	-.06
T4LC2	-.06	.05	.00	.02	.08	.07	.07	-.03	-.09	.00	.04	-.04	-.07
T4LC3	-.22	-.11	-.07	-.05	-.15	-.03	-.07	.09	.14	.07	.00	.13	.10
T4LC4	-.07	-.26	-.18	-.17	-.35	-.16	-.19	.19	.33	.04	-.06	.25	.17
T4LC5	.11	.05	.01	.05	.09	.37	.02	-.03	.01	.29	.09	.02	.08
T4LC6	-.05	.03	.00	.03	.00	.10	.01	-.03	.03	.11	.06	.01	.04
T4LC7	-.18	-.10	-.06	-.04	-.18	-.01	-.10	.08	.21	.12	.01	.14	.16
T4LC8	-.02	-.07	-.05	-.04	-.16	.03	-.12	.03	.23	.17	.02	.11	.20
T4L11X	-.18	-.19	-.12	-.13	-.17	-.19	-.06	.14	.06	-.16	-.08	.10	-.06
T4L12X	-.14	-.11	-.06	-.06	-.04	.02	-.01	.09	-.03	-.04	-.04	.05	-.07
T4L13X	-.18	-.07	-.05	-.05	-.04	-.08	.01	.06	-.05	-.12	-.04	.02	-.10
T4L14X	-.28	-.15	-.08	-.09	-.13	-.10	-.05	.11	.07	-.06	-.04	.10	.00
T4L15X	-.18	-.13	-.08	-.09	-.10	-.13	-.04	.10	.01	-.13	-.07	.06	-.06
T4L11Y	-.03	-.19	-.10	-.11	-.22	-.15	-.09	.19	.25	.01	.00	.24	.12
T4L12Y	.13	.11	.10	.12	.20	.38	.12	-.01	-.04	.29	.17	.03	.05
T4L13Y	-.02	.07	.05	.07	.08	.12	.08	-.01	-.01	.11	.10	.02	.02
T4L14Y	-.17	-.09	-.03	-.02	-.14	-.01	-.07	.09	.18	.12	.04	.15	.14
T4L15Y	.02	.02	.07	.05	.02	.05	.02	.05	.12	.13	.11	.10	.12
T5F1	.08	.22	.14	.17	.46	.50	.03	-.17	-.44	.29	.03	-.14	-.03
T5F3	.13	.54	.26	.30	.23	.39	-.09	-.58	-.33	.22	-.07	-.51	-.01
T5F4	.12	.24	.22	.23	.42	.28	.25	-.05	-.28	.04	.13	-.09	-.17
T5F5	.13	.12	.08	.09	.19	.00	.20	-.01	-.16	-.17	.03	-.08	-.23
T5F6	.11	.28	.26	.25	.48	.09	.43	-.01	-.22	-.12	.24	-.09	-.24
T5A1	.87	.37	.25	.25	.23	.07	.19	-.24	-.18	-.12	.04	-.28	-.20
T5A2	--	.22	.11	.10	.11	.05	.10	-.17	-.14	-.13	-.04	-.21	-.19
T5SC1		--	.69	.68	.56	.20	.52	-.64	-.15	.00	.36	-.61	-.15
T5SC2			--	.98	.50	.30	.57	-.12	.00	.13	.44	-.11	-.06
T5SC3				--	.49	.33	.55	-.13	.01	.16	.44	-.12	-.04
T5SC4					--	.31	.79	-.03	-.39	-.01	.54	-.11	-.26
T5LC1						--	.48	.12	.15	.62	.45	.13	.01
T5LC2							--	.13	.03	.14	.76	.03	-.20
T5LC3								--	.17	.12	.13	.92	.03
T5LC4									--	.69	.52	.42	.77
T5LC5										--	.65	.40	.79
T5LC6											--	.28	.47
T5LC7												--	.38
T5LC8													--
T5L11X													
T5L12X													
T5L13X													
T5L14X													
T5L15X													
T5L11Y													
T5L12Y													
T5L13Y													
T5L14Y													

	T5 L11X	T5 L12X	T5 L13X	T5 L14X	T5 L15X	T5 L11Y	T5 L12Y	T5 L13Y	T5 L14Y	T5 L15Y
T3LC7	.02	-.04	.00	.02	-.03	.10	-.01	.05	.04	.03
T3LC8	.18	.09	.07	.14	.12	.20	.05	.02	.10	.15
T3L11X	.14	.01	-.01	.06	.05	.18	-.05	-.06	.05	.06
T3L12X	.07	.06	.02	.02	.03	.11	.12	.02	.02	.07
T3L13X	.01	-.05	-.03	-.01	-.05	.08	-.05	-.01	.01	-.03
T3L14X	.06	-.01	.00	.02	-.01	.15	-.01	.01	.03	.03
T3L15X	.14	.07	.04	.08	.08	.16	.03	-.01	.05	.09
T3L11Y	.10	-.09	-.03	.11	-.01	.25	-.07	.00	.16	.10
T3L12Y	-.06	.00	-.02	-.05	-.07	.07	.22	.10	.02	.09
T3L13Y	-.09	-.13	-.07	.01	-.14	.04	-.06	.04	.08	-.06
T3L14Y	.02	-.06	.00	.03	-.04	.14	.00	.06	.06	.06
T3L15Y	.14	.03	.05	.13	.07	.24	.05	.04	.13	.18
T4F1	-.28	-.02	-.05	-.23	-.18	-.30	.22	.13	-.19	-.12
T4F3	-.22	-.01	-.03	-.16	-.14	-.25	.16	.12	-.17	-.11
T4F4	-.03	.12	.14	.04	.03	-.16	.13	.14	-.03	-.06
T4F5	.12	.18	.19	.19	.15	-.04	.05	.05	.05	.00
T4F6	.08	.16	.18	.12	.12	-.08	.06	.06	-.01	-.03
T4A1	-.10	-.05	-.04	-.08	-.09	-.19	-.14	-.08	-.15	-.26
T4A2	-.11	-.10	-.07	-.13	-.12	-.14	-.14	-.07	-.13	-.21
T4SC1	-.11	.03	.06	-.07	-.02	-.30	-.10	.01	-.22	-.21
T4SC2	-.04	.09	.08	-.03	.04	-.18	.03	.03	-.10	-.08
T4SC3	-.04	.10	.08	-.03	.03	-.17	.05	.03	-.10	-.07
T4SC4	-.18	-.02	.00	-.18	-.10	-.28	.02	.05	-.18	-.17
T4LC1	-.16	.06	.01	-.12	-.10	-.20	.24	.10	-.07	-.08
T4LC2	-.04	.02	.05	-.06	-.02	-.12	-.03	.01	-.05	-.11
T4LC3	.11	.05	.03	.03	.05	.14	.06	.01	.12	.08
T4LC4	.27	.11	.08	.21	.13	.37	.08	-.01	.24	.20
T4LC5	-.04	.09	.02	-.05	-.04	.00	.27	.08	.00	.05
T4LC6	.01	.05	.05	-.03	.00	-.01	.06	.02	-.02	-.03
T4LC7	.15	.08	.04	.05	.07	.20	.12	.01	.12	.13
T4LC8	.12	.06	.01	.06	.04	.22	.16	.02	.09	.16
T4L11X	.11	.02	.07	.17	.06	.12	-.09	-.01	.13	.03
T4L12X	.01	.02	.04	.10	.00	.00	-.01	.01	.07	-.02
T4L13X	.03	.02	.07	.08	.02	-.04	-.10	-.01	.03	-.07
T4L14X	.10	.04	.06	.10	.05	.10	-.02	.00	.11	.03
T4L15X	.06	.01	.05	.14	.04	.05	-.08	-.01	.09	.00
T4L11Y	.16	.03	.08	.14	.07	.29	.06	.05	.23	.15
T4L12Y	-.13	.01	.01	-.10	-.11	-.05	.25	.14	.01	.01
T4L13Y	-.05	.00	.04	-.07	-.05	-.05	.06	.07	.00	-.04
T4L14Y	.11	.05	.03	.03	.05	.18	.12	.04	.13	.12
T4L15Y	-.02	-.04	.00	-.02	-.05	.11	.11	.09	.08	.08
T5F1	-.34	.04	-.08	-.19	-.12	-.45	.28	.00	-.14	-.03
T5F3	-.22	.08	-.10	-.43	-.04	-.41	.13	-.17	-.52	-.12
T5F4	-.22	-.05	.01	-.10	-.13	-.27	.05	.14	-.08	-.14
T5F5	-.12	-.12	-.01	-.06	-.12	-.16	-.16	.04	-.06	-.21
T5F6	-.17	-.12	.08	-.07	-.14	-.22	-.13	.24	-.08	-.22
T5A1	-.08	-.04	.01	-.18	-.03	-.25	-.21	-.04	-.30	-.29
T5A2	-.07	-.04	-.02	-.15	-.03	-.20	-.18	-.10	-.22	-.25
T5SC1	-.13	-.06	.09	-.49	-.10	-.30	-.19	.20	-.65	-.35
T5SC2	-.16	-.12	-.01	-.20	-.18	-.12	-.02	.29	-.15	-.21
T5SC3	-.15	-.10	.00	-.20	-.17	-.11	.01	.29	-.16	-.20
T5SC4	-.33	-.13	.17	-.14	-.22	-.44	-.08	.49	-.14	-.31
T5LC1	-.13	.07	.02	-.07	-.23	.09	.55	.39	.08	-.05
T5LC2	-.15	-.11	.22	-.06	-.25	-.04	.03	.69	-.01	-.29
T5LC3	-.04	-.09	-.03	.55	-.12	.27	.23	.24	.93	.18
T5LC4	.48	.26	.21	.25	.23	.83	.51	.39	.30	.51
T5LC5	.18	.28	.16	.13	.10	.51	.80	.50	.27	.51
T5LC6	.09	.11	.35	.09	-.02	.33	.42	.84	.15	.19
T5LC7	.06	.01	.02	.55	-.03	.44	.42	.31	.96	.39
T5LC8	.35	.31	.19	.21	.32	.59	.58	.31	.26	.69
T5L11X	--	.90	.82	.68	.92	.48	.18	.11	.04	.33
T5L12X		--	.88	.67	.94	.27	.29	.13	.00	.31
T5L13X			--	.66	.84	.21	.15	.37	.00	.18
T5L14X				--	.67	.34	.22	.18	.56	.32
T5L15X					--	.25	.12	.01	-.02	.33
T5L11Y						--	.69	.52	.47	.77
T5L12Y							--	.62	.46	.79
T5L13Y								--	.33	.46
T5L14Y									--	.47

Appendix 12.1

*Regression analysis summary for productive vocabulary knowledge predicting each component score (*n = *225)*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-1.36	0.23		-5.90	0.00				
PF1Size	0.00	0.00	0.15	1.28	0.20	0.45	0.09	0.07	0.28
Deri	0.01	0.03	0.04	0.45	0.65	0.39	0.03	0.03	0.07
Anto	0.10	0.03	0.29	2.88	0.00	0.48	0.19	0.17	0.58
Collo	0.02	0.03	0.05	0.56	0.58	0.36	0.04	0.03	0.07
(Constant)	-0.29	0.26		-1.12	0.26				
PF2Size	0.00	0.00	-0.02	-0.17	0.86	0.07	-0.01	-0.01	-0.18
Deri	0.02	0.03	0.07	0.61	0.54	0.09	0.04	0.04	0.63
Anto	0.00	0.04	0.01	0.09	0.93	0.07	0.01	0.01	0.08
Collo	-0.29	0.26	0.05	-1.12	0.26	0.09	0.04	0.04	0.47
(Constant)	-0.58	0.26		-2.26	0.02	0.23	0.06	0.06	0.45
PF3Size	0.00	0.00	0.13	0.95	0.34	0.16	-0.03	-0.03	-0.12
Deri	-0.01	0.03	-0.05	-0.43	0.67	0.24	0.11	0.11	0.74
Anto	0.07	0.04	0.19	1.70	0.09	0.14	-0.02	-0.02	-0.07
Collo	-0.01	0.03	-0.03	-0.34	0.73	0.23	0.06	0.06	0.45
(Constant)	-0.27	0.26		-1.04	0.30				
PF4Size	0.00	0.00	0.06	0.42	0.67	0.13	0.03	0.03	0.34
Deri	0.03	0.03	0.11	0.95	0.34	0.14	0.06	0.06	0.67
Anto	0.01	0.04	0.03	0.25	0.80	0.12	0.02	0.02	0.16
Collo	-0.02	0.03	-0.05	-0.52	0.61	0.07	-0.03	-0.03	-0.17
(Constant)	-0.05	0.26		-0.17	0.86	0.02	0.06	0.06	0.51
PF5Size	0.00	0.00	0.12	0.89	0.37	0.00	-0.01	-0.01	-0.01
Deri	0.00	0.03	-0.01	-0.13	0.90	-0.02	-0.06	-0.06	0.49
Anto	-0.04	0.04	-0.11	-0.91	0.36	0.00	0.00	0.00	0.00
Collo	0.00	0.03	-0.01	-0.06	0.95	0.02	0.06	0.06	0.51
(Constant)	-0.57	0.26		-2.18	0.03				
PF6Size	0.00	0.00	0.22	1.67	0.10	0.20	0.11	0.11	1.12
Deri	-0.01	0.03	-0.02	-0.21	0.83	0.14	-0.01	-0.01	-0.08
Anto	0.00	0.04	-0.01	-0.11	0.91	0.15	-0.01	-0.01	-0.05
Collo	0.00	0.03	0.00	0.05	0.96	0.13	0.00	0.00	0.01
(Constant)	-1.05	0.25		-4.29	0.00				
PA1Size	0.00	0.00	0.51	4.01	0.00	0.36	0.26	0.25	1.25
Deri	-0.06	0.03	-0.20	-1.99	0.05	0.20	-0.13	-0.12	-0.28
Anto	0.00	0.04	0.00	-0.02	0.99	0.28	0.00	0.00	0.00
Collo	0.01	0.03	0.02	0.23	0.82	0.23	0.02	0.01	0.03
(Constant)	-1.05	0.25		-4.21	0.00				
PA2Size	0.00	0.00	0.21	1.63	0.10	0.33	0.11	0.10	0.59
Deri	-0.01	0.03	-0.04	-0.38	0.70	0.26	-0.03	-0.02	-0.09
Anto	0.05	0.04	0.13	1.20	0.23	0.32	0.08	0.08	0.35
Collo	0.02	0.03	0.07	0.74	0.46	0.26	0.05	0.05	0.15
(Constant)	-0.85	0.25		-3.35	0.00				
PA3Size	0.00	0.00	0.19	1.50	0.14	0.29	0.10	0.10	0.62
Deri	-0.01	0.03	-0.02	-0.20	0.84	0.23	-0.01	-0.01	-0.05
Anto	0.05	0.04	0.13	1.18	0.24	0.28	0.08	0.08	0.41
Collo	0.00	0.03	0.01	0.13	0.90	0.21	0.01	0.01	0.03
(Constant)	-0.89	0.25		-3.59	0.00				
PA4Size	0.00	0.00	0.41	3.17	0.00	0.33	0.21	0.20	1.13
Deri	-0.04	0.03	-0.16	-1.52	0.13	0.19	-0.10	-0.10	-0.26
Anto	0.02	0.04	0.07	0.61	0.54	0.28	0.04	0.04	0.16
Collo	-0.01	0.03	-0.02	-0.19	0.85	0.20	-0.01	-0.01	-0.03
(Constant)	-0.43	0.26		-1.65	0.10	0.14	0.08	0.08	1.13
PA5Size	0.00	0.00	0.17	1.24	0.22	0.09	-0.03	-0.03	-0.23
Deri	-0.01	0.03	-0.05	-0.48	0.63	0.11	0.00	0.00	0.00
Anto	0.00	0.04	0.00	0.01	0.99	0.10	0.01	0.01	0.09
Collo	0.01	0.03	0.02	0.21	0.84	0.14	0.08	0.08	1.13
(Constant)	-0.58	0.26		-2.23	0.03				
PSC1Size	0.00	0.00	0.25	1.86	0.06	0.19	0.12	0.12	1.24
Deri	-0.01	0.03	-0.02	-0.20	0.84	0.13	-0.01	-0.01	-0.08
Anto	-0.03	0.04	-0.09	-0.74	0.46	0.12	-0.05	-0.05	-0.28
Collo	0.01	0.03	0.03	0.35	0.73	0.13	0.02	0.02	0.12

(Appendix 12.1 continues)

(Appendix 12.1 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-0.26	0.26		-1.01	0.31				
PSC2Size	0.00	0.00	0.07	0.55	0.59	0.13	0.04	0.04	0.41
Deri	0.03	0.03	0.12	1.07	0.29	0.15	0.07	0.07	0.71
Anto	0.00	0.04	0.01	0.11	0.91	0.12	0.01	0.01	0.06
Collo	-0.02	0.03	-0.06	-0.64	0.52	0.07	-0.04	-0.04	-0.18
(Constant)	-0.45	0.26		-1.72	0.09				
PSC3Size	0.00	0.00	0.05	0.39	0.70	0.16	0.03	0.03	0.27
Deri	0.03	0.03	0.09	0.86	0.39	0.17	0.06	0.06	0.51
Anto	0.01	0.04	0.04	0.36	0.72	0.15	0.02	0.02	0.21
Collo	0.00	0.03	0.00	0.02	0.99	0.12	0.00	0.00	0.01
(Constant)	-0.55	0.26		-2.11	0.04				
PSC4Size	0.00	0.00	0.20	1.51	0.13	0.21	0.10	0.10	0.83
Deri	0.03	0.03	0.12	1.07	0.29	0.20	0.07	0.07	0.46
Anto	-0.03	0.04	-0.08	-0.69	0.49	0.15	-0.05	-0.05	-0.23
Collo	-0.01	0.03	-0.02	-0.26	0.80	0.14	-0.02	-0.02	-0.06
(Constant)	-1.21	0.23		-5.25	0.00				
PSC5Size	0.00	0.00	0.35	2.91	0.00	0.48	0.19	0.17	0.69
Deri	0.03	0.03	0.09	0.97	0.33	0.41	0.07	0.06	0.16
Anto	0.05	0.03	0.14	1.34	0.18	0.43	0.09	0.08	0.25
Collo	-0.03	0.03	-0.07	-0.88	0.38	0.30	-0.06	-0.05	-0.09
(Constant)	-1.14	0.24		-4.83	0.00				
PLC1Size	0.00	0.00	0.10	0.81	0.42	0.41	0.05	0.05	0.20
Deri	0.03	0.03	0.10	1.00	0.32	0.38	0.07	0.06	0.18
Anto	0.10	0.04	0.29	2.79	0.01	0.44	0.18	0.17	0.62
Collo	0.00	0.03	0.00	-0.04	0.97	0.31	0.00	0.00	0.00
(Constant)	0.13	0.26		0.49	0.63				
PLC2Size	0.00	0.00	0.10	0.73	0.47	0.05	0.05	0.05	0.33
Deri	0.02	0.03	0.09	0.80	0.43	0.06	0.05	0.05	0.33
Anto	-0.01	0.04	-0.02	-0.17	0.87	0.03	-0.01	-0.01	-0.04
Collo	-0.05	0.03	-0.15	-1.58	0.11	-0.04	-0.11	-0.11	0.38
(Constant)	0.04	0.26		0.16	0.87				
PLC3Size	0.00	0.00	0.24	1.80	0.07	0.02	0.12	0.12	0.26
Deri	-0.01	0.03	-0.05	-0.45	0.65	-0.02	-0.03	-0.03	0.06
Anto	-0.06	0.04	-0.18	-1.59	0.11	-0.06	-0.11	-0.11	0.56
Collo	-0.02	0.03	-0.05	-0.58	0.56	-0.04	-0.04	-0.04	0.11
(Constant)	0.69	0.26		2.68	0.01				
PLC4Size	0.00	0.00	0.06	0.42	0.67	-0.18	0.03	0.03	-0.19
Deri	-0.02	0.03	-0.09	-0.81	0.42	-0.19	-0.05	-0.05	0.33
Anto	-0.04	0.04	-0.12	-1.06	0.29	-0.20	-0.07	-0.07	0.46
Collo	-0.04	0.03	-0.10	-1.13	0.26	-0.20	-0.08	-0.07	0.40
(Constant)	0.43	0.26		1.67	0.10				
PLC5Size	0.00	0.00	-0.28	-2.12	0.03	-0.19	-0.14	-0.14	1.24
Deri	0.03	0.03	0.11	1.00	0.32	-0.10	0.07	0.07	-0.25
Anto	-0.01	0.04	-0.03	-0.28	0.78	-0.15	-0.02	-0.02	0.12
Collo	0.02	0.03	0.05	0.59	0.56	-0.08	0.04	0.04	-0.11
(Constant)	0.88	0.25		3.49	0.00				
PLC6Size	0.00	0.00	-0.20	-1.54	0.12	-0.25	-0.10	-0.10	0.56
Deri	0.05	0.03	0.20	1.89	0.06	-0.14	0.13	0.12	-0.31
Anto	-0.06	0.04	-0.19	-1.70	0.09	-0.27	-0.11	-0.11	0.55
Collo	-0.03	0.03	-0.08	-0.94	0.35	-0.21	-0.06	-0.06	0.19
(Constant)	0.13	0.26		0.49	0.62				
PLC7Size	0.00	0.00	0.09	0.71	0.48	-0.04	0.05	0.05	-0.14
Deri	0.03	0.03	0.11	1.01	0.31	-0.01	0.07	0.07	-0.03
Anto	-0.09	0.04	-0.28	-2.41	0.02	-0.12	-0.16	-0.16	1.16
Collo	0.00	0.03	0.00	0.01	0.99	-0.04	0.00	0.00	0.00
(Constant)	0.03	0.26		0.10	0.92				
PLC8Size	0.00	0.00	0.02	0.11	0.91	-0.06	0.01	0.01	-0.03
Deri	0.03	0.03	0.12	1.08	0.28	-0.01	0.07	0.07	-0.03
Anto	-0.10	0.04	-0.28	-2.46	0.01	-0.13	-0.16	-0.16	1.06
Collo	0.03	0.03	0.09	0.94	0.35	0.00	0.06	0.06	-0.01
(Constant)	0.07	0.26		0.29	0.78				
PLC9Size	0.00	0.00	-0.31	-2.28	0.02	-0.14	-0.15	-0.15	1.06
Deri	0.00	0.03	-0.01	-0.09	0.93	-0.09	-0.01	-0.01	0.02
Anto	0.02	0.04	0.06	0.53	0.60	-0.08	0.04	0.03	-0.12
Collo	0.06	0.03	0.18	1.94	0.05	0.01	0.13	0.13	0.03

(Appendix 12.1 continues)

(Appendix 12.1 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.37	0.26		1.42	0.16				
PLC10Size	0.00	0.00	-0.15	-1.09	0.28	-0.08	-0.07	-0.07	0.70
Deri	0.04	0.03	0.16	1.48	0.14	-0.01	0.10	0.10	-0.09
Anto	-0.01	0.04	-0.02	-0.14	0.89	-0.07	-0.01	-0.01	0.06
Collo	-0.03	0.03	-0.07	-0.77	0.44	-0.08	-0.05	-0.05	0.32
(Constant)	0.60	0.26		2.35	0.02				
PLC11Size	0.00	0.00	-0.18	-1.34	0.18	-0.24	-0.09	-0.09	0.62
Deri	0.01	0.03	0.05	0.46	0.65	-0.17	0.03	0.03	-0.12
Anto	-0.06	0.04	-0.17	-1.49	0.14	-0.25	-0.10	-0.10	0.60
Collo	0.02	0.03	0.05	0.52	0.60	-0.14	0.03	0.03	-0.10
(Constant)	0.73	0.26		2.83	0.01				
PLC12Size	0.00	0.00	-0.04	-0.30	0.76	-0.21	-0.02	-0.02	0.15
Deri	-0.02	0.03	-0.08	-0.73	0.47	-0.21	-0.05	-0.05	0.30
Anto	-0.03	0.04	-0.08	-0.68	0.50	-0.21	-0.05	-0.04	0.29
Collo	-0.03	0.03	-0.07	-0.80	0.43	-0.20	-0.05	-0.05	0.26

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index.

Variable	R	R^2	Adjusted R^2	SEE	F	p	DW
PF1	0.49	0.24	0.23	0.88	17.62	0.00	1.38
PF2	0.10	0.01	-0.01	1.00	0.54	0.71	1.80
PF3	0.25	0.06	0.05	0.98	3.72	0.01	2.20
PF4	0.15	0.02	0.00	1.00	1.23	0.30	2.00
PF5	0.07	0.00	-0.01	1.01	0.26	0.90	1.91
PF6	0.20	0.04	0.02	0.99	2.26	0.06	2.06
PA1	0.38	0.14	0.13	0.93	9.28	0.00	2.01
PA2	0.34	0.12	0.10	0.95	7.35	0.00	2.06
PA3	0.30	0.09	0.08	0.96	5.57	0.00	2.05
PA4	0.34	0.12	0.10	0.95	7.30	0.00	2.03
PA5	0.14	0.02	0.00	1.00	1.16	0.33	2.06
PSC1	0.19	0.04	0.02	0.99	2.12	0.08	1.97
PSC2	0.16	0.02	0.01	1.00	1.36	0.25	2.22
PSC3	0.17	0.03	0.01	0.99	1.73	0.14	1.84
PSC4	0.23	0.05	0.03	0.98	2.96	0.02	1.65
PSC5	0.49	0.24	0.23	0.88	17.50	0.00	2.03
PLC1	0.45	0.20	0.19	0.90	14.14	0.00	1.62
PLC2	0.12	0.02	0.00	1.00	0.86	0.49	1.76
PLC3	0.14	0.02	0.00	1.00	1.03	0.39	1.53
PLC4	0.23	0.05	0.03	0.98	3.03	0.02	1.95
PLC5	0.21	0.04	0.03	0.99	2.46	0.05	1.99
PLC6	0.30	0.09	0.07	0.96	5.53	0.00	1.95
PLC7	0.17	0.03	0.01	0.99	1.70	0.15	2.04
PLC8	0.19	0.04	0.02	0.99	2.04	0.09	1.88
PLC9	0.20	0.04	0.02	0.99	2.36	0.05	1.78
PLC10	0.13	0.02	0.00	1.00	0.98	0.42	1.91
PLC11	0.26	0.07	0.05	0.97	4.10	0.00	1.82
PLC12	0.24	0.06	0.04	0.98	3.26	0.01	1.96

Note. n = 225. SEE = Standard Error of the Estimate; F (4, 220); DW = Durbin-Watson.

Appendix 12.3

Regression analysis summary for productive vocabulary knowledge predicting each component score (outliers excluded)

Variable	R	R^2	Adjusted R^2	SEE	F	p	DW
PF1	0.52	0.27	0.26	0.86	20.61	0.00	1.42
PF2	0.25	0.06	0.04	0.89	3.49	0.01	1.67
PF3	0.24	0.06	0.04	0.94	3.45	0.01	2.12
PF4	0.15	0.02	0.01	0.88	1.29	0.28	1.90
PF5	0.09	0.01	-0.01	0.98	0.50	0.74	1.93
PF6	0.19	0.04	0.02	0.96	2.08	0.08	2.01
PA1	0.39	0.15	0.13	0.91	9.65	0.00	1.93
PSC2	0.21	0.04	0.03	0.84	2.50	0.04	2.16
PSC3	0.30	0.09	0.07	0.60	5.39	0.00	1.84

(Appendix 12.3 continues)

(Appendix 12.3 continued)

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
PSC4	0.27	0.07	0.06	0.93	4.27	0.00	1.56
PLC1	0.48	0.23	0.22	0.88	16.80	0.00	1.65
PLC2	0.09	0.01	-0.01	0.95	0.46	0.77	1.80
PLC3	0.14	0.02	0.00	0.97	1.06	0.38	1.53
PLC4	0.21	0.04	0.03	0.80	2.46	0.05	1.92
PLC8	0.19	0.04	0.02	0.95	2.14	0.08	1.95
PLC9	0.23	0.05	0.04	0.92	3.13	0.02	1.77
PLC10	0.14	0.02	0.00	0.94	1.06	0.38	1.77
PLC12	0.26	0.07	0.05	0.96	4.03	0.00	1.72

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson.

Appendix 12.4

*Regression analysis summary for productive vocabulary knowledge predicting each token and type (*n = *225)*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	27.15	3.45		7.86	0.00				
T1TokenSize	0.00	0.00	0.06	0.48	0.63	0.27	0.03	0.03	0.18
Deri	0.09	0.39	0.02	0.22	0.83	0.23	0.01	0.06	
Anto	1.13	0.52	0.24	2.18	0.03	0.30	0.15	0.14	0.79
Collo	-0.06	0.43	-0.01	-0.13	0.90	0.20	-0.01	-0.01	-0.02
(Constant)	18.56	2.02		9.18	0.00				
T1TypeSize	0.00	0.00	0.17	1.31	0.19	0.32	0.09	0.08	0.46
Deri	0.04	0.23	0.02	0.19	0.85	0.27	0.01	0.01	0.05
Anto	0.53	0.30	0.19	1.76	0.08	0.33	0.12	0.11	0.54
Collo	-0.08	0.25	-0.03	-0.31	0.76	0.22	-0.02	-0.02	-0.05
(Constant)	14.15	3.11		4.55	0.00				
T2TokenSize	0.00	0.00	0.13	1.02	0.31	0.40	0.07	0.06	0.27
Deri	0.18	0.35	0.05	0.52	0.60	0.35	0.04	0.03	0.10
Anto	1.06	0.47	0.24	2.26	0.02	0.41	0.15	0.14	0.54
Collo	0.24	0.39	0.05	0.61	0.54	0.32	0.04	0.04	0.09
(Constant)	11.31	1.15		9.87	0.00				
T2TypeSize	0.00	0.00	0.06	0.50	0.62	0.35	0.03	0.03	0.14
Deri	0.04	0.13	0.03	0.28	0.78	0.30	0.02	0.02	0.06
Anto	0.50	0.17	0.31	2.89	0.00	0.39	0.19	0.18	0.78
Collo	0.03	0.14	0.02	0.19	0.85	0.27	0.01	0.01	0.03
(Constant)	10.22	2.58		3.97	0.00				
T3TokenSize	0.00	0.00	0.14	1.21	0.23	0.44	0.08	0.07	0.28
Deri	0.33	0.29	0.11	1.11	0.27	0.41	0.07	0.07	0.20
Anto	0.76	0.39	0.20	1.97	0.05	0.44	0.13	0.12	0.40
Collo	0.29	0.32	0.08	0.91	0.36	0.37	0.06	0.05	0.12
(Constant)	9.34	1.46		6.39	0.00				
T3TypeSize	0.00	0.00	0.15	1.23	0.22	0.43	0.08	0.07	0.30
Deri	0.20	0.17	0.12	1.19	0.24	0.40	0.08	0.07	0.22
Anto	0.40	0.22	0.19	1.81	0.07	0.43	0.12	0.11	0.38
Collo	0.12	0.18	0.06	0.67	0.50	0.35	0.05	0.04	0.09
(Constant)	12.70	2.62		4.84	0.00				
T4TokenSize	0.01	0.00	0.18	1.53	0.13	0.47	0.10	0.09	0.31
Deri	0.03	0.30	0.01	0.12	0.91	0.40	0.01	0.01	0.02
Anto	1.40	0.40	0.35	3.53	0.00	0.51	0.23	0.20	0.67
Collo	0.01	0.33	0.00	0.04	0.97	0.35	0.00	0.00	0.00
(Constant)	11.12	1.38		8.05	0.00				
T4TypeSize	0.00	0.00	0.24	2.09	0.04	0.50	0.14	0.12	0.43
Deri	0.05	0.16	0.03	0.33	0.74	0.41	0.02	0.02	0.05
Anto	0.66	0.21	0.31	3.15	0.00	0.51	0.21	0.18	0.57
Collo	-0.08	0.17	-0.04	-0.47	0.64	0.34	-0.03	-0.03	-0.05
(Constant)	18.08	2.88		6.29	0.00	0.38	0.06	0.05	0.23
T5TokenSize	0.00	0.00	0.11	0.88	0.38	0.31	0.01	0.00	0.01
Deri	0.03	0.33	0.01	0.08	0.94	0.42	0.21	0.20	0.80
Anto	1.40	0.43	0.34	3.23	0.00	0.27	-0.02	-0.02	-0.04
Collo	-0.10	0.36	-0.02	-0.29	0.77	0.38	0.06	0.05	0.23
(Constant)	12.38	1.35		9.14	0.00				
T5TypeSize	0.00	0.00	0.07	0.59	0.56	0.40	0.04	0.04	0.14
Deri	0.16	0.15	0.11	1.06	0.29	0.37	0.07	0.06	0.19
Anto	0.63	0.20	0.32	3.10	0.00	0.44	0.20	0.19	0.70
Collo	-0.04	0.17	-0.02	-0.23	0.82	0.30	-0.02	-0.01	-0.03

Note. *n* = 225. *r* = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; *d* = relative Pratt index.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1Tokens	0.31	0.09	0.08	13.15	5.69	0.00	1.60
T1Types	0.34	0.12	0.10	7.70	7.22	0.00	1.58
T2Tokens	0.43	0.18	0.17	11.84	12.34	0.00	1.43
T2Types	0.40	0.16	0.14	4.36	10.26	0.00	1.57
T3Tokens	0.48	0.23	0.21	9.81	16.12	0.00	1.60
T3Types	0.46	0.21	0.20	5.57	14.71	0.00	1.61
T4Tokens	0.52	0.27	0.26	9.98	20.31	0.00	1.50
T4Types	0.53	0.28	0.27	5.26	21.35	0.00	1.47
T5Tokens	0.43	0.18	0.17	10.95	12.17	0.00	1.66
T5Types	0.45	0.21	0.19	5.16	14.21	0.00	1.64

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (4, 220); DW = Durbin-Watson.

Appendix 12.6

Regression analysis summary for productive vocabulary knowledge predicting each token and type (outliers excluded)

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T2Types	0.39	0.16	0.14	4.23	10.12	0.00	1.62
T3Tokens	0.51	0.26	0.25	9.52	19.37	0.00	1.63
T3Types	0.52	0.27	0.26	5.16	20.43	0.00	1.63
T4Types	0.56	0.32	0.31	5.08	25.50	0.00	1.53

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson.

Appendix 12.7

*Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (*n *= 225) in Task 1*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	36.20	4.60		7.86	0.00	0.27	0.03	0.03	0.18
T1F1Size	0.00	0.01	0.06	0.48	0.63	0.23	0.01	0.01	0.06
Deri	0.11	0.52	0.02	0.22	0.83	0.30	0.15	0.14	0.79
Anto	1.51	0.69	0.24	2.18	0.03	0.20	-0.01	-0.01	-0.02
Collo	-0.07	0.57	-0.01	-0.13	0.90	0.27	0.03	0.03	0.18
(Constant)	7.21	0.91		7.90	0.00				
T1F3Size	0.00	0.00	0.12	0.90	0.37	0.22	0.06	0.06	0.47
Deri	-0.02	0.10	-0.02	-0.15	0.88	0.17	-0.01	-0.01	-0.05
Anto	0.17	0.14	0.14	1.23	0.22	0.23	0.08	0.08	0.57
Collo	0.00	0.11	0.00	0.03	0.98	0.16	0.00	0.00	0.01
(Constant)	1.90	1.29		1.47	0.14				
T1F4Size	0.00	0.00	-0.01	-0.09	0.93	0.15	-0.01	-0.01	-0.05
Deri	0.12	0.15	0.09	0.79	0.43	0.17	0.05	0.05	0.40
Anto	0.05	0.19	0.03	0.26	0.79	0.15	0.02	0.02	0.12
Collo	0.19	0.16	0.11	1.18	0.24	0.18	0.08	0.08	0.53
(Constant)	0.12	0.03		3.71	0.00				
T1F5Size	0.00	0.00	-0.07	-0.53	0.60	-0.04	-0.04	-0.04	1.12
Deri	0.00	0.00	0.04	0.39	0.70	-0.02	0.03	0.03	-0.25
Anto	0.00	0.00	-0.02	-0.14	0.89	-0.04	-0.01	-0.01	0.21
Collo	0.00	0.00	0.01	0.13	0.90	-0.02	0.01	0.01	-0.08
(Constant)	0.56	0.18		3.10	0.00				
T1F6Size	0.00	0.00	-0.03	-0.22	0.83	0.02	-0.01	-0.01	-0.33
Deri	0.01	0.02	0.05	0.46	0.64	0.04	0.03	0.03	1.10
Anto	0.00	0.03	0.00	-0.01	0.99	0.02	0.00	0.00	-0.01
Collo	0.00	0.02	0.02	0.17	0.87	0.03	0.01	0.01	0.25
(Constant)	0.54	0.05		11.06	0.00				
T1A1Size	0.00	0.00	0.50	3.97	0.00	0.36	0.26	0.25	1.26
Deri	-0.01	0.01	-0.19	-1.85	0.07	0.20	-0.12	-0.12	-0.27
Anto	0.00	0.01	-0.02	-0.16	0.87	0.27	-0.01	-0.01	-0.03
Collo	0.00	0.01	0.03	0.30	0.76	0.23	0.02	0.02	0.04
(Constant)	0.52	0.05		10.47	0.00	0.35	0.26	0.25	1.24
T1A2Size	0.00	0.00	0.50	3.97	0.00	0.19	-0.14	-0.13	-0.29
Deri	-0.01	0.01	-0.22	-2.09	0.04	0.28	0.01	0.01	0.03
Anto	0.00	0.01	0.01	0.13	0.90	0.22	0.01	0.01	0.02
Collo	0.00	0.01	0.01	0.16	0.88	0.35	0.26	0.25	1.24

(Appendix 12.7 continues)

(Appendix 12.7 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	1.00	0.04		23.98	0.00				
T1SC1Size	0.00	0.00	0.27	1.98	0.05	0.19	0.13	0.13	1.27
Deri	0.00	0.00	-0.04	-0.37	0.71	0.13	-0.03	-0.02	-0.13
Anto	0.00	0.01	-0.08	-0.67	0.50	0.13	-0.05	-0.04	-0.25
Collo	0.00	0.01	0.03	0.37	0.71	0.14	0.02	0.02	0.12
(Constant)	0.03	0.04		0.74	0.46				
T1SC2Size	0.00	0.00	0.27	2.01	0.05	0.18	0.13	0.13	1.35
Deri	0.00	0.00	-0.03	-0.31	0.76	0.12	-0.02	-0.02	-0.11
Anto	-0.01	0.01	-0.10	-0.86	0.39	0.11	-0.06	-0.06	-0.30
Collo	0.00	0.00	0.02	0.23	0.82	0.12	0.02	0.02	0.07
(Constant)	0.02	0.03		0.80	0.42				
T1SC3Size	0.00	0.00	0.27	2.03	0.04	0.18	0.14	0.13	1.29
Deri	0.00	0.00	-0.02	-0.17	0.87	0.13	-0.01	-0.01	-0.06
Anto	0.00	0.00	-0.12	-1.00	0.32	0.11	-0.07	-0.07	-0.33
Collo	0.00	0.00	0.03	0.32	0.75	0.13	0.02	0.02	0.10
(Constant)	5.23	0.29		17.99	0.00				
T1SC4Size	0.00	0.00	0.10	0.80	0.43	0.25	0.05	0.05	0.35
Deri	0.03	0.03	0.10	0.96	0.34	0.23	0.06	0.06	0.33
Anto	0.05	0.04	0.14	1.24	0.22	0.25	0.08	0.08	0.48
Collo	-0.03	0.04	-0.08	-0.86	0.39	0.14	-0.06	-0.06	-0.16
(Constant)	3.57	0.16		22.02	0.00				
T1LC1Size	0.00	0.00	0.26	2.06	0.04	0.34	0.14	0.13	0.73
Deri	0.00	0.02	0.00	-0.01	0.99	0.27	0.00	0.00	0.00
Anto	0.03	0.02	0.13	1.22	0.22	0.32	0.08	0.08	0.35
Collo	-0.01	0.02	-0.04	-0.45	0.66	0.22	-0.03	-0.03	-0.07
(Constant)	3.75	0.23		16.11	0.00				
T1LC2Size	0.00	0.00	0.22	1.65	0.10	0.23	0.11	0.11	0.80
Deri	0.02	0.03	0.08	0.73	0.47	0.19	0.05	0.05	0.25
Anto	0.01	0.04	0.05	0.41	0.69	0.19	0.03	0.03	0.14
Collo	-0.04	0.03	-0.14	-1.47	0.14	0.09	-0.10	-0.10	-0.20
(Constant)	3.80	0.23		16.80	0.00				
T1LC3Size	0.00	0.00	0.04	0.32	0.75	0.09	0.02	0.02	0.15
Deri	0.02	0.03	0.11	0.96	0.34	0.10	0.06	0.06	0.41
Anto	0.03	0.03	0.10	0.85	0.39	0.10	0.06	0.06	0.39
Collo	-0.05	0.03	-0.16	-1.75	0.08	-0.01	-0.12	-0.12	0.04
(Constant)	0.51	0.02		29.09	0.00				
T1LC4Size	0.00	0.00	0.09	0.72	0.47	-0.25	0.05	0.05	-0.22
Deri	0.00	0.00	-0.06	-0.60	0.55	-0.25	-0.04	-0.04	0.15
Anto	-0.01	0.00	-0.27	-2.41	0.02	-0.31	-0.16	-0.15	0.77
Collo	0.00	0.00	-0.12	-1.37	0.17	-0.27	-0.09	-0.09	0.30
(Constant)	2.60	0.12		21.00	0.00				
T1LC5Size	0.00	0.00	0.10	0.75	0.45	0.07	0.05	0.05	0.78
Deri	0.00	0.01	0.00	-0.02	0.98	0.04	0.00	0.00	-0.01
Anto	0.01	0.02	0.03	0.29	0.77	0.06	0.02	0.02	0.23
Collo	-0.01	0.02	-0.09	-0.93	0.35	0.00	-0.06	-0.06	0.00
(Constant)	2.70	0.16		16.97	0.00				
T1LC6Size	0.00	0.00	0.14	1.02	0.31	0.03	0.07	0.07	0.28
Deri	0.01	0.02	0.05	0.45	0.66	0.02	0.03	0.03	0.07
Anto	-0.01	0.02	-0.06	-0.49	0.62	0.00	-0.03	-0.03	0.02
Collo	-0.03	0.02	-0.15	-1.59	0.11	-0.06	-0.11	-0.11	0.63
(Constant)	2.71	0.15		17.53	0.00				
T1LC7Size	0.00	0.00	-0.01	-0.07	0.94	-0.09	0.00	0.00	0.04
Deri	0.01	0.02	0.06	0.53	0.60	-0.07	0.04	0.04	-0.16
Anto	0.00	0.02	-0.01	-0.07	0.94	-0.08	-0.01	0.00	0.03
Collo	-0.04	0.02	-0.17	-1.84	0.07	-0.15	-0.12	-0.12	1.10
(Constant)	0.72	0.02		34.62	0.00				
T1LC8Size	0.00	0.00	-0.16	-1.24	0.22	-0.32	-0.08	-0.08	0.45
Deri	0.00	0.00	0.00	0.01	0.99	-0.26	0.00	0.00	0.00
Anto	0.00	0.00	-0.16	-1.42	0.16	-0.32	-0.10	-0.09	0.44
Collo	0.00	0.00	-0.05	-0.59	0.55	-0.26	-0.04	-0.04	0.12
(Constant)	0.05	0.01		4.63	0.00				
T1LC11XSize	0.00	0.00	0.25	1.82	0.07	0.02	0.12	0.12	0.22
Deri	0.00	0.00	-0.05	-0.44	0.66	-0.02	-0.03	-0.03	0.06
Anto	0.00	0.00	-0.21	-1.78	0.08	-0.06	-0.12	-0.12	0.67
Collo	0.00	0.00	-0.03	-0.37	0.71	-0.03	-0.03	-0.02	0.05

(Appendix 12.7 continues)

(Appendix 12.7 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.25	0.06		4.36	0.00				
T1LC12XSize	0.00	0.00	0.26	1.93	0.05	0.08	0.13	0.13	1.18
Deri	0.00	0.01	-0.04	-0.36	0.72	0.03	-0.02	-0.02	-0.07
Anto	-0.01	0.01	-0.12	-1.05	0.29	0.02	-0.07	-0.07	-0.11
Collo	-0.01	0.01	-0.07	-0.76	0.45	0.00	-0.05	-0.05	0.00
(Constant)	0.24	0.06		3.93	0.00				
T1LC13XSize	0.00	0.00	0.27	2.01	0.05	0.09	0.13	0.13	1.21
Deri	0.00	0.01	-0.04	-0.37	0.71	0.04	-0.02	-0.02	-0.08
Anto	-0.01	0.01	-0.14	-1.21	0.23	0.02	-0.08	-0.08	-0.11
Collo	0.00	0.01	-0.06	-0.61	0.54	0.01	-0.04	-0.04	-0.03
(Constant)	0.24	0.06		4.32	0.00				
T1LC14XSize	0.00	0.00	0.25	1.86	0.06	0.06	0.12	0.12	0.85
Deri	0.00	0.01	-0.04	-0.34	0.74	0.01	-0.02	-0.02	-0.03
Anto	-0.01	0.01	-0.14	-1.19	0.24	-0.01	-0.08	-0.08	0.08
Collo	-0.01	0.01	-0.08	-0.85	0.40	-0.02	-0.06	-0.06	0.10
(Constant)	0.07	0.01		4.86	0.00				
T1LC15XSize	0.00	0.00	0.20	1.48	0.14	0.01	0.10	0.10	0.17
Deri	0.00	0.00	-0.04	-0.34	0.73	-0.02	-0.02	-0.02	0.06
Anto	0.00	0.00	-0.17	-1.42	0.16	-0.05	-0.10	-0.09	0.67
Collo	0.00	0.00	-0.04	-0.42	0.67	-0.03	-0.03	-0.03	0.10
(Constant)	0.34	0.02		19.89	0.00				
T1LC11YSize	0.00	0.00	0.06	0.45	0.65	-0.06	0.03	0.03	-0.19
Deri	0.00	0.00	0.05	0.48	0.63	-0.04	0.03	0.03	-0.10
Anto	0.00	0.00	-0.21	-1.83	0.07	-0.12	-0.12	-0.12	1.33
Collo	0.00	0.00	0.02	0.16	0.87	-0.04	0.01	0.01	-0.03
(Constant)	1.71	0.12		14.61	0.00				
T1LC12YSize	0.00	0.00	0.11	0.84	0.40	0.18	0.06	0.06	0.60
Deri	0.01	0.01	0.05	0.47	0.64	0.16	0.03	0.03	0.25
Anto	0.00	0.02	0.01	0.13	0.90	0.15	0.01	0.01	0.07
Collo	0.00	0.01	0.02	0.20	0.84	0.14	0.01	0.01	0.08
(Constant)	1.79	0.15		11.65	0.00				
T1LC13YSize	0.00	0.00	0.11	0.81	0.42	0.12	0.05	0.05	0.67
Deri	0.02	0.02	0.10	0.93	0.35	0.13	0.06	0.06	0.65
Anto	-0.01	0.02	-0.03	-0.29	0.77	0.09	-0.02	-0.02	-0.15
Collo	-0.01	0.02	-0.06	-0.61	0.54	0.06	-0.04	-0.04	-0.17
(Constant)	1.83	0.14		12.67	0.00				
T1LC14YSize	0.00	0.00	-0.02	-0.12	0.91	0.02	-0.01	-0.01	-0.05
Deri	0.02	0.02	0.12	1.04	0.30	0.05	0.07	0.07	0.83
Anto	0.00	0.02	0.01	0.07	0.94	0.02	0.01	0.01	0.03
Collo	-0.02	0.02	-0.09	-0.91	0.36	-0.02	-0.06	-0.06	0.19
(Constant)	0.48	0.02		22.10	0.00				
T1LC15YSize	0.00	0.00	-0.10	-0.71	0.48	-0.08	-0.05	-0.05	0.44
Deri	0.00	0.00	0.08	0.76	0.45	-0.03	0.05	0.05	-0.13
Anto	0.00	0.00	-0.13	-1.11	0.27	-0.10	-0.07	-0.07	0.73
Collo	0.00	0.00	0.08	0.88	0.38	-0.01	0.06	0.06	-0.04

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.31	0.09	0.08	17.53	5.69	0.00	1.60
T1F3	0.24	0.06	0.04	3.47	3.22	0.01	1.67
T1F4	0.19	0.04	0.02	4.92	2.09	0.08	1.78
T1F5	0.05	0.00	-0.02	0.12	0.15	0.96	1.90
T1F6	0.04	0.00	-0.02	0.69	0.10	0.98	1.83
T1A1	0.38	0.14	0.13	0.19	9.05	0.00	2.04
T1A2	0.38	0.14	0.13	0.19	9.16	0.00	1.99
T1SC1	0.20	0.04	0.02	0.16	2.33	0.06	1.92
T1SC2	0.19	0.04	0.02	0.15	2.03	0.09	1.97
T1SC3	0.20	0.04	0.02	0.11	2.21	0.07	1.98
T1SC4	0.27	0.07	0.06	1.11	4.31	0.00	1.75
T1LC1	0.35	0.12	0.11	0.62	7.79	0.00	1.65
T1LC2	0.25	0.06	0.04	0.89	3.59	0.01	1.87
T1LC3	0.16	0.03	0.01	0.86	1.48	0.21	1.84
T1LC4	0.33	0.11	0.09	0.07	6.66	0.00	1.68
T1LC5	0.09	0.01	-0.01	0.47	0.49	0.74	1.85
T1LC6	0.12	0.01	0.00	0.61	0.82	0.51	1.76
T1LC7	0.15	0.02	0.01	0.59	1.30	0.27	1.67
T1LC8	0.34	0.11	0.10	0.08	7.09	0.00	1.81
T1LC11X	0.14	0.02	0.00	0.04	1.10	0.36	1.52
T1LC12X	0.14	0.02	0.00	0.22	1.03	0.39	1.59
T1LC13X	0.14	0.02	0.00	0.23	1.11	0.35	1.58
T1LC14X	0.13	0.02	0.00	0.22	0.97	0.43	1.54
T1LC15X	0.12	0.01	0.00	0.05	0.74	0.57	1.53
T1LC11Y	0.14	0.02	0.00	0.06	1.06	0.38	1.68
T1LC12Y	0.18	0.03	0.02	0.45	1.86	0.12	1.85
T1LC13Y	0.14	0.02	0.00	0.59	1.13	0.34	1.87
T1LC14Y	0.09	0.01	-0.01	0.55	0.42	0.79	1.87
T1LC15Y	0.13	0.02	0.00	0.08	0.96	0.43	1.84

Note. n = 225. SEE = Standard Error of the Estimate; F (4, 220); DW = Durbin-Watson.

Appendix 12.8

*Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (*n = *225) in Task 2*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	18.86	4.15		4.55	0.00				
T1F1Size	0.01	0.01	0.13	1.02	0.31	0.40	0.07	0.06	0.27
Deri	0.25	0.47	0.05	0.52	0.60	0.35	0.04	0.03	0.10
Anto	1.41	0.62	0.24	2.26	0.02	0.41	0.15	0.14	0.54
Collo	0.31	0.52	0.05	0.61	0.54	0.32	0.04	0.04	0.09
(Constant)	3.17	0.64		4.98	0.00				
T1F3Size	0.00	0.00	0.21	1.62	0.11	0.32	0.11	0.10	0.62
Deri	0.03	0.07	0.04	0.42	0.68	0.28	0.03	0.03	0.11
Anto	0.03	0.10	0.04	0.32	0.75	0.28	0.02	0.02	0.09
Collo	0.07	0.08	0.07	0.83	0.41	0.26	0.06	0.05	0.18
(Constant)	0.88	1.44		0.61	0.54				
T1F4Size	0.00	0.00	0.15	1.19	0.23	0.31	0.08	0.08	0.43
Deri	-0.02	0.16	-0.01	-0.13	0.90	0.25	-0.01	-0.01	-0.03
Anto	0.38	0.22	0.19	1.74	0.08	0.32	0.12	0.11	0.56
Collo	0.04	0.18	0.02	0.23	0.82	0.23	0.02	0.01	0.04
(Constant)	0.13	0.04		3.65	0.00				
T1F5Size	0.00	0.00	0.08	0.60	0.55	0.09	0.04	0.04	0.51
Deri	0.00	0.00	-0.05	-0.46	0.65	0.05	-0.03	-0.03	-0.18
Anto	0.01	0.01	0.11	0.96	0.34	0.11	0.06	0.06	0.80
Collo	0.00	0.00	-0.06	-0.63	0.53	0.03	-0.04	-0.04	-0.13
(Constant)	0.44	0.23		1.88	0.06				
T1F6Size	0.00	0.00	0.16	1.18	0.24	0.23	0.08	0.08	0.57
Deri	-0.01	0.03	-0.06	-0.56	0.58	0.16	-0.04	-0.04	-0.15
Anto	0.05	0.04	0.17	1.52	0.13	0.24	0.10	0.10	0.66
Collo	-0.01	0.03	-0.03	-0.34	0.74	0.14	-0.02	-0.02	-0.07

(Appendix 12.8 continues)

(Appendix 12.8 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.17	0.06		2.59	0.01				
T1A1Size	0.00	0.00	0.19	1.38	0.17	0.11	0.09	0.09	1.27
Deri	-0.01	0.01	-0.09	-0.84	0.40	0.05	-0.06	-0.06	-0.30
Anto	0.00	0.01	-0.02	-0.21	0.84	0.08	-0.01	-0.01	-0.11
Collo	0.00	0.01	0.03	0.32	0.75	0.08	0.02	0.02	0.14
(Constant)	0.17	0.06		2.96	0.00				
T1A2Size	0.00	0.00	0.14	1.01	0.31	0.16	0.07	0.07	0.85
Deri	0.00	0.01	-0.01	-0.06	0.95	0.12	0.00	0.00	-0.03
Anto	0.00	0.01	0.03	0.23	0.82	0.14	0.02	0.02	0.15
Collo	0.00	0.01	0.01	0.09	0.93	0.11	0.01	0.01	0.04
(Constant)	0.65	0.05		12.71	0.00				
T1SC1Size	0.00	0.00	0.32	2.53	0.01	0.35	0.17	0.16	0.83
Deri	0.01	0.01	0.09	0.92	0.36	0.31	0.06	0.06	0.22
Anto	-0.01	0.01	-0.13	-1.15	0.25	0.25	-0.08	-0.07	-0.23
Collo	0.01	0.01	0.09	0.98	0.33	0.28	0.07	0.06	0.18
(Constant)	4.25	0.42		10.08	0.00				
T1SC4Size	0.00	0.00	0.09	0.75	0.45	0.37	0.05	0.05	0.21
Deri	0.03	0.05	0.06	0.61	0.54	0.33	0.04	0.04	0.13
Anto	0.15	0.06	0.25	2.34	0.02	0.39	0.16	0.14	0.60
Collo	0.02	0.05	0.03	0.40	0.69	0.29	0.03	0.02	0.06
(Constant)	2.88	0.11		26.80	0.00				
T1LC1Size	0.00	0.00	-0.11	-0.79	0.43	0.11	-0.05	-0.05	-0.30
Deri	0.00	0.01	0.01	0.08	0.94	0.10	0.01	0.01	0.02
Anto	0.04	0.02	0.28	2.40	0.02	0.18	0.16	0.16	1.33
Collo	0.00	0.01	-0.03	-0.29	0.77	0.08	-0.02	-0.02	-0.06
(Constant)	3.23	0.24		13.42	0.00				
T1LC2Size	0.00	0.00	-0.11	-0.84	0.40	0.10	-0.06	-0.06	-0.40
Deri	0.02	0.03	0.07	0.60	0.55	0.12	0.04	0.04	0.27
Anto	0.07	0.04	0.21	1.82	0.07	0.16	0.12	0.12	1.16
Collo	0.00	0.03	-0.01	-0.11	0.91	0.09	-0.01	-0.01	-0.03
(Constant)	5.70	0.61		9.34	0.00				
T1LC3Size	0.00	0.00	-0.33	-2.54	0.01	-0.21	-0.17	-0.17	1.10
Deri	-0.06	0.07	-0.09	-0.85	0.40	-0.18	-0.06	-0.06	0.26
Anto	0.19	0.09	0.23	2.05	0.04	-0.09	0.14	0.13	-0.34
Collo	0.01	0.08	0.01	0.10	0.92	-0.13	0.01	0.01	-0.02
(Constant)	0.48	0.03		18.16	0.00				
T1LC4Size	0.00	0.00	-0.24	-1.93	0.06	-0.39	-0.13	-0.12	0.53
Deri	0.00	0.00	0.11	1.07	0.29	-0.29	0.07	0.07	-0.17
Anto	-0.01	0.00	-0.23	-2.12	0.04	-0.40	-0.14	-0.13	0.50
Collo	0.00	0.00	-0.08	-0.97	0.33	-0.31	-0.07	-0.06	0.15
(Constant)	1.94	0.09		21.21	0.00				
T1LC5Size	0.00	0.00	-0.19	-1.43	0.15	-0.16	-0.10	-0.09	0.75
Deri	0.02	0.01	0.19	1.70	0.09	-0.06	0.11	0.11	-0.29
Anto	-0.01	0.01	-0.06	-0.48	0.63	-0.14	-0.03	-0.03	0.20
Collo	-0.01	0.01	-0.10	-1.05	0.29	-0.14	-0.07	-0.07	0.35
(Constant)	2.19	0.16		13.78	0.00				
T1LC6Size	0.00	0.00	-0.19	-1.37	0.17	-0.08	-0.09	-0.09	0.81
Deri	0.03	0.02	0.18	1.61	0.11	0.00	0.11	0.11	0.04
Anto	0.00	0.02	0.00	0.00	1.00	-0.06	0.00	0.00	0.00
Collo	-0.01	0.02	-0.05	-0.50	0.62	-0.06	-0.03	-0.03	0.15
(Constant)	3.81	0.38		10.00	0.00				
T1LC7Size	0.00	0.00	-0.36	-2.73	0.01	-0.28	-0.18	-0.18	1.16
Deri	-0.01	0.04	-0.02	-0.22	0.83	-0.22	-0.01	-0.01	0.06
Anto	0.07	0.06	0.15	1.29	0.20	-0.18	0.09	0.08	-0.30
Collo	-0.02	0.05	-0.04	-0.39	0.69	-0.20	-0.03	-0.03	0.08
(Constant)	0.68	0.03		26.94	0.00				
T1LC8Size	0.00	0.00	-0.14	-1.11	0.27	-0.28	-0.07	-0.07	0.33
Deri	0.01	0.00	0.21	1.97	0.05	-0.16	0.13	0.12	-0.28
Anto	-0.01	0.00	-0.29	-2.61	0.01	-0.32	-0.17	-0.16	0.76
Collo	0.00	0.00	-0.10	-1.10	0.27	-0.24	-0.07	-0.07	0.19
(Constant)	0.05	0.01		5.33	0.00				
T1LC11XSize	0.00	0.00	0.11	0.79	0.43	-0.08	0.05	0.05	-0.18
Deri	0.00	0.00	0.09	0.86	0.39	-0.05	0.06	0.06	-0.09
Anto	0.00	0.00	-0.35	-3.06	0.00	-0.18	-0.20	-0.20	1.31
Collo	0.00	0.00	0.03	0.34	0.73	-0.06	0.02	0.02	-0.04

(Appendix 12.8 continues)

(Appendix 12.8 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.21	0.04		5.13	0.00				
T1LC12XSize	0.00	0.00	0.14	1.05	0.30	0.01	0.07	0.07	0.02
Deri	0.01	0.00	0.15	1.33	0.19	0.04	0.09	0.09	0.17
Anto	-0.02	0.01	-0.32	-2.79	0.01	-0.10	-0.19	-0.18	0.80
Collo	0.00	0.01	0.02	0.22	0.83	0.01	0.01	0.01	0.00
(Constant)	0.21	0.05		4.68	0.00				
T1LC13XSize	0.00	0.00	0.12	0.86	0.39	0.05	0.06	0.06	0.20
Deri	0.01	0.01	0.15	1.39	0.16	0.08	0.09	0.09	0.47
Anto	-0.01	0.01	-0.24	-2.10	0.04	-0.04	-0.14	-0.14	0.31
Collo	0.00	0.01	0.01	0.16	0.87	0.04	0.01	0.01	0.02
(Constant)	0.41	0.07		5.55	0.00				
T1LC14XSize	0.00	0.00	-0.07	-0.49	0.62	-0.12	-0.03	-0.03	0.47
Deri	0.00	0.01	0.03	0.26	0.79	-0.09	0.02	0.02	-0.15
Anto	-0.01	0.01	-0.07	-0.58	0.56	-0.12	-0.04	-0.04	0.48
Collo	0.00	0.01	-0.03	-0.35	0.73	-0.10	-0.02	-0.02	0.20
(Constant)	0.07	0.01		5.02	0.00				
T1LC15XSize	0.00	0.00	0.19	1.40	0.16	0.02	0.09	0.09	0.06
Deri	0.00	0.00	0.14	1.29	0.20	0.05	0.09	0.09	0.14
Anto	-0.01	0.00	-0.36	-3.17	0.00	-0.10	-0.21	-0.21	0.79
Collo	0.00	0.00	0.02	0.22	0.82	0.01	0.02	0.01	0.00
(Constant)	0.49	0.03		18.17	0.00				
T1LC11YSize	0.00	0.00	-0.23	-1.83	0.07	-0.37	-0.12	-0.11	0.55
Deri	0.00	0.00	0.08	0.81	0.42	-0.28	0.05	0.05	-0.15
Anto	-0.01	0.00	-0.20	-1.88	0.06	-0.37	-0.13	-0.12	0.48
Collo	0.00	0.00	-0.07	-0.75	0.46	-0.29	-0.05	-0.05	0.12
(Constant)	1.98	0.10		20.10	0.00				
T1LC12YSize	0.00	0.00	-0.14	-1.06	0.29	-0.10	-0.07	-0.07	0.81
Deri	0.01	0.01	0.13	1.21	0.23	-0.03	0.08	0.08	-0.25
Anto	0.00	0.01	-0.01	-0.11	0.91	-0.08	-0.01	-0.01	0.06
Collo	-0.01	0.01	-0.07	-0.76	0.45	-0.09	-0.05	-0.05	0.37
(Constant)	2.23	0.16		13.50	0.00				
T1LC13YSize	0.00	0.00	-0.16	-1.18	0.24	-0.03	-0.08	-0.08	0.46
Deri	0.03	0.02	0.15	1.37	0.17	0.03	0.09	0.09	0.46
Anto	0.01	0.02	0.05	0.42	0.68	0.00	0.03	0.03	-0.01
Collo	-0.01	0.02	-0.04	-0.45	0.65	-0.02	-0.03	-0.03	0.09
(Constant)	3.91	0.43		9.20	0.00				
T1LC14YSize	0.00	0.00	-0.33	-2.53	0.01	-0.25	-0.17	-0.16	1.18
Deri	-0.02	0.05	-0.04	-0.33	0.74	-0.20	-0.02	-0.02	0.10
Anto	0.09	0.06	0.15	1.34	0.18	-0.15	0.09	0.09	-0.33
Collo	-0.01	0.05	-0.02	-0.20	0.84	-0.17	-0.01	-0.01	0.04
(Constant)	0.69	0.02		28.35	0.00				
T1LC15YSize	0.00	0.00	-0.11	-0.81	0.42	-0.24	-0.05	-0.05	0.28
Deri	0.00	0.00	0.16	1.53	0.13	-0.14	0.10	0.10	-0.27
Anto	-0.01	0.00	-0.26	-2.33	0.02	-0.28	-0.16	-0.15	0.82
Collo	0.00	0.00	-0.07	-0.79	0.43	-0.20	-0.05	-0.05	0.16

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.43	0.18	0.17	15.79	12.34	0.00	1.43
T1F3	0.33	0.11	0.09	2.42	6.71	0.00	1.83
T1F4	0.33	0.11	0.09	5.48	6.87	0.00	2.22
T1F5	0.12	0.01	0.00	0.14	0.82	0.51	2.14
T1F6	0.25	0.06	0.05	0.89	3.66	0.01	2.30
T1A1	0.13	0.02	0.00	0.25	0.94	0.44	2.08
T1A2	0.16	0.03	0.01	0.21	1.43	0.23	2.02
T1SC1	0.37	0.13	0.12	0.20	8.57	0.00	2.12
T1SC4	0.40	0.16	0.15	1.61	10.75	0.00	2.05
T1LC1	0.19	0.04	0.02	0.41	2.15	0.08	1.79
T1LC2	0.17	0.03	0.01	0.92	1.63	0.17	2.03
T1LC3	0.25	0.06	0.05	2.32	3.79	0.01	2.10
T1LC4	0.42	0.18	0.16	0.10	11.89	0.00	1.71
T1LC5	0.20	0.04	0.02	0.35	2.29	0.06	2.07
T1LC6	0.13	0.02	0.00	0.60	1.01	0.40	2.03
T1LC7	0.29	0.09	0.07	1.45	5.18	0.00	2.06
T1LC8	0.35	0.12	0.10	0.10	7.55	0.00	2.00
T1LC11X	0.22	0.05	0.03	0.04	2.80	0.03	2.01
T1LC12X	0.20	0.04	0.02	0.15	2.19	0.07	2.04
T1LC13X	0.17	0.03	0.01	0.17	1.56	0.19	2.02
T1LC14X	0.13	0.02	0.00	0.28	0.95	0.44	2.17
T1LC15X	0.22	0.05	0.03	0.05	2.72	0.03	2.03
T1LC11Y	0.40	0.16	0.14	0.10	10.19	0.00	1.61
T1LC12Y	0.13	0.02	0.00	0.38	0.97	0.43	1.94
T1LC13Y	0.11	0.01	-0.01	0.63	0.62	0.65	1.98
T1LC14Y	0.27	0.07	0.05	1.62	4.15	0.00	2.04
T1LC15Y	0.30	0.09	0.07	0.09	5.31	0.00	1.93

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (4, 220); DW = Durbin-Watson.

Appendix 12.9

*Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (*n *= 225) in Task 3*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	13.63	3.44		3.97	0.00				
T1F1Size	0.01	0.00	0.14	1.21	0.23	0.44	0.08	0.07	0.28
Deri	0.43	0.39	0.11	1.11	0.27	0.41	0.07	0.07	0.20
Anto	1.02	0.52	0.20	1.97	0.05	0.44	0.13	0.12	0.40
Collo	0.39	0.43	0.08	0.91	0.36	0.37	0.06	0.05	0.12
(Constant)	2.84	0.62		4.56	0.00				
T1F3Size	0.00	0.00	0.06	0.46	0.65	0.37	0.03	0.03	0.13
Deri	0.09	0.07	0.13	1.31	0.19	0.36	0.09	0.08	0.29
Anto	0.15	0.09	0.18	1.65	0.10	0.37	0.11	0.10	0.40
Collo	0.08	0.08	0.09	1.08	0.28	0.33	0.07	0.07	0.18
(Constant)	2.26	1.44		1.57	0.12				
T1F4Size	0.00	0.00	0.17	1.24	0.22	0.17	0.08	0.08	0.87
Deri	0.05	0.16	0.03	0.31	0.76	0.15	0.02	0.02	0.15
Anto	-0.13	0.22	-0.07	-0.62	0.54	0.12	-0.04	-0.26	
Collo	0.10	0.18	0.05	0.58	0.56	0.14	0.04	0.04	0.24
(Constant)	0.19	0.04		4.21	0.00				
T1F5Size	0.00	0.00	0.06	0.43	0.67	-0.05	0.03	0.03	-0.33
Deri	0.00	0.01	-0.01	-0.07	0.94	-0.05	-0.01	-0.01	0.05
Anto	-0.01	0.01	-0.14	-1.17	0.24	-0.09	-0.08	-0.08	1.34
Collo	0.00	0.01	0.01	0.13	0.89	-0.04	0.01	0.01	-0.05
(Constant)	1.11	0.26		4.30	0.00				
T1F6Size	0.00	0.00	0.13	0.95	0.35	-0.01	0.06	0.06	-0.09
Deri	0.00	0.03	0.00	-0.03	0.98	-0.02	0.00	0.00	0.01
Anto	-0.04	0.04	-0.13	-1.07	0.28	-0.05	-0.07	-0.07	0.85
Collo	-0.02	0.03	-0.04	-0.47	0.64	-0.04	-0.03	-0.03	0.23
(Constant)	0.30	0.07		4.41	0.00				
T1A1Size	0.00	0.00	0.22	1.75	0.08	0.34	0.12	0.11	0.62
Deri	0.00	0.01	-0.03	-0.32	0.75	0.26	-0.02	-0.02	-0.07
Anto	0.01	0.01	0.13	1.22	0.22	0.32	0.08	0.08	0.35
Collo	0.00	0.01	0.05	0.53	0.60	0.26	0.04	0.03	0.10

(Appendix 12.9 continues)

(Appendix 12.9 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.29	0.07		4.22	0.00				
T1A2Size	0.00	0.00	0.19	1.48	0.14	0.31	0.10	0.09	0.55
Deri	0.00	0.01	-0.05	-0.44	0.66	0.24	-0.03	-0.03	-0.10
Anto	0.01	0.01	0.13	1.17	0.24	0.30	0.08	0.07	0.36
Collo	0.01	0.01	0.08	0.92	0.36	0.26	0.06	0.06	0.20
(Constant)	1.00	0.06		17.72	0.00				
T1SC1Size	0.00	0.00	0.03	0.21	0.84	0.13	0.01	0.01	0.15
Deri	0.01	0.01	0.13	1.15	0.25	0.15	0.08	0.08	0.81
Anto	0.00	0.01	0.01	0.09	0.92	0.12	0.01	0.01	0.05
Collo	0.00	0.01	0.00	-0.05	0.96	0.10	0.00	0.00	-0.02
(Constant)	0.06	0.05		1.22	0.22				
T1SC2Size	0.00	0.00	0.05	0.40	0.69	0.10	0.03	0.03	0.37
Deri	0.01	0.01	0.11	1.00	0.32	0.11	0.07	0.07	0.88
Anto	0.00	0.01	-0.01	-0.13	0.90	0.08	-0.01	-0.01	-0.08
Collo	0.00	0.01	-0.05	-0.50	0.62	0.05	-0.03	-0.03	-0.17
(Constant)	0.04	0.03		1.44	0.15				
T1SC3Size	0.00	0.00	0.06	0.47	0.64	0.11	0.03	0.03	0.38
Deri	0.00	0.00	0.12	1.09	0.28	0.12	0.07	0.07	0.83
Anto	0.00	0.00	-0.01	-0.05	0.96	0.09	0.00	0.00	-0.03
Collo	0.00	0.00	-0.06	-0.68	0.49	0.05	-0.05	-0.05	-0.18
(Constant)	5.79	0.37		15.59	0.00				
T1SC4Size	0.00	0.00	0.18	1.39	0.17	0.20	0.09	0.09	0.74
Deri	0.01	0.04	0.03	0.31	0.75	0.16	0.02	0.02	0.11
Anto	0.04	0.06	0.09	0.79	0.43	0.18	0.05	0.05	0.34
Collo	-0.06	0.05	-0.13	-1.38	0.17	0.07	-0.09	-0.09	-0.19
(Constant)	2.84	0.14		20.49	0.00				
T1LC1Size	0.00	0.00	0.12	0.96	0.34	0.35	0.06	0.06	0.30
Deri	0.01	0.02	0.10	0.93	0.35	0.33	0.06	0.06	0.22
Anto	0.03	0.02	0.15	1.41	0.16	0.35	0.09	0.09	0.37
Collo	0.01	0.02	0.05	0.60	0.55	0.29	0.04	0.04	0.11
(Constant)	4.76	0.34		14.13	0.00				
T1LC2Size	0.00	0.00	0.09	0.64	0.52	-0.01	0.04	0.04	-0.15
Deri	-0.01	0.04	-0.01	-0.13	0.90	-0.03	-0.01	-0.01	0.06
Anto	-0.01	0.05	-0.02	-0.21	0.84	-0.03	-0.01	-0.01	0.09
Collo	-0.05	0.04	-0.10	-1.09	0.28	-0.07	-0.07	-0.07	1.00
(Constant)	4.99	0.37		13.63	0.00				
T1LC3Size	0.00	0.00	0.11	0.84	0.40	-0.10	0.06	0.06	-0.35
Deri	-0.05	0.04	-0.13	-1.22	0.22	-0.15	-0.08	-0.08	0.62
Anto	-0.01	0.06	-0.02	-0.19	0.85	-0.10	-0.01	-0.01	0.07
Collo	-0.06	0.05	-0.13	-1.42	0.16	-0.16	-0.10	-0.09	0.66
(Constant)	0.50	0.02		20.42	0.00				
T1LC4Size	0.00	0.00	-0.19	-1.48	0.14	-0.36	-0.10	-0.09	0.47
Deri	0.00	0.00	-0.03	-0.25	0.80	-0.30	-0.02	-0.02	0.05
Anto	-0.01	0.00	-0.19	-1.78	0.08	-0.36	-0.12	-0.11	0.48
Collo	0.00	0.00	0.00	0.01	0.99	-0.26	0.00	0.00	0.00
(Constant)	1.81	0.11		16.85	0.00				
T1LC5Size	0.00	0.00	-0.10	-0.77	0.44	0.07	-0.05	-0.05	-0.44
Deri	0.02	0.01	0.14	1.29	0.20	0.12	0.09	0.09	1.02
Anto	0.01	0.02	0.05	0.45	0.66	0.08	0.03	0.03	0.27
Collo	0.00	0.01	0.03	0.31	0.76	0.08	0.02	0.02	0.15
(Constant)	2.93	0.21		13.72	0.00				
T1LC6Size	0.00	0.00	-0.01	-0.05	0.96	-0.14	0.00	0.00	0.04
Deri	0.00	0.02	-0.01	-0.12	0.91	-0.13	-0.01	-0.01	0.06
Anto	-0.02	0.03	-0.08	-0.70	0.49	-0.15	-0.05	-0.05	0.43
Collo	-0.02	0.03	-0.09	-0.93	0.35	-0.15	-0.06	-0.06	0.47
(Constant)	3.10	0.24		12.83	0.00				
T1LC7Size	0.00	0.00	0.02	0.16	0.87	-0.19	0.01	0.01	-0.07
Deri	-0.03	0.03	-0.11	-0.97	0.33	-0.21	-0.07	-0.06	0.39
Anto	-0.02	0.04	-0.06	-0.50	0.62	-0.19	-0.03	-0.03	0.19
Collo	-0.04	0.03	-0.13	-1.37	0.17	-0.21	-0.09	-0.09	0.49
(Constant)	0.63	0.02		25.23	0.00				
T1LC8Size	0.00	0.00	-0.24	-1.79	0.07	-0.26	-0.12	-0.12	0.85
Deri	0.00	0.00	0.08	0.75	0.46	-0.18	0.05	0.05	-0.20
Anto	0.00	0.00	-0.10	-0.88	0.38	-0.24	-0.06	-0.06	0.33
Collo	0.00	0.00	-0.01	-0.09	0.93	-0.18	-0.01	-0.01	0.02

(Appendix 12.9 continues)

(Appendix 12.9 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.11	0.01		7.85	0.00				
T1LC11XSize	0.00	0.00	-0.04	-0.31	0.76	-0.14	-0.02	-0.02	0.10
Deri	0.00	0.00	0.11	1.00	0.32	-0.07	0.07	0.07	-0.14
Anto	-0.01	0.00	-0.31	-2.67	0.01	-0.20	-0.18	-0.18	1.13
Collo	0.00	0.00	0.10	1.08	0.28	-0.05	0.07	0.07	-0.09
(Constant)	0.39	0.06		6.97	0.00				
T1LC12XSize	0.00	0.00	0.00	-0.02	0.99	0.03	0.00	0.00	0.00
Deri	0.01	0.01	0.17	1.57	0.12	0.10	0.11	0.10	0.52
Anto	-0.02	0.01	-0.21	-1.83	0.07	-0.03	-0.12	-0.12	0.17
Collo	0.01	0.01	0.11	1.19	0.23	0.09	0.08	0.08	0.31
(Constant)	0.66	0.09		6.92	0.00				
T1LC13XSize	0.00	0.00	0.01	0.06	0.95	-0.09	0.00	0.00	-0.03
Deri	0.01	0.01	0.07	0.66	0.51	-0.05	0.04	0.04	-0.14
Anto	-0.03	0.01	-0.22	-1.88	0.06	-0.14	-0.13	-0.12	1.24
Collo	0.00	0.01	0.04	0.39	0.69	-0.05	0.03	0.03	-0.07
(Constant)	0.66	0.09		7.02	0.00				
T1LC14XSize	0.00	0.00	0.05	0.37	0.71	-0.12	0.03	0.02	-0.18
Deri	0.00	0.01	0.00	0.03	0.97	-0.10	0.00	0.00	-0.01
Anto	-0.03	0.01	-0.24	-2.08	0.04	-0.18	-0.14	-0.14	1.27
Collo	0.00	0.01	0.04	0.39	0.70	-0.08	0.03	0.03	-0.08
(Constant)	0.13	0.02		7.98	0.00				
T1LC15XSize	0.00	0.00	0.00	-0.02	0.99	-0.06	0.00	0.00	0.00
Deri	0.00	0.00	0.15	1.38	0.17	0.00	0.09	0.09	0.01
Anto	-0.01	0.00	-0.29	-2.56	0.01	-0.14	-0.17	-0.17	0.99
Collo	0.00	0.00	0.09	0.95	0.34	0.00	0.06	0.06	0.00
(Constant)	0.47	0.02		19.11	0.00				
T1LC11YSize	0.00	0.00	-0.11	-0.87	0.39	-0.35	-0.06	-0.05	0.25
Deri	0.00	0.00	-0.05	-0.52	0.61	-0.31	-0.03	-0.03	0.10
Anto	-0.01	0.00	-0.29	-2.68	0.01	-0.39	-0.18	-0.17	0.72
Collo	0.00	0.00	0.04	0.50	0.62	-0.24	0.03	0.03	-0.07
(Constant)	1.67	0.10		16.73	0.00	0.05	-0.01	-0.01	-0.10
T1LC12YSize	0.00	0.00	-0.02	-0.18	0.86	0.08	0.06	0.06	0.64
Deri	0.01	0.01	0.10	0.86	0.39	0.02	-0.04	-0.04	-0.16
Anto	-0.01	0.02	-0.07	-0.64	0.52	0.08	0.06	0.06	0.61
Collo	0.01	0.01	0.09	0.91	0.36	0.05	-0.01	-0.01	-0.10
(Constant)	2.72	0.21		13.24	0.00				
T1LC13YSize	0.00	0.00	0.04	0.30	0.76	-0.15	0.02	0.02	-0.17
Deri	-0.01	0.02	-0.04	-0.41	0.68	-0.15	-0.03	-0.03	0.19
Anto	-0.04	0.03	-0.15	-1.30	0.19	-0.18	-0.09	-0.09	0.77
Collo	-0.01	0.03	-0.05	-0.55	0.59	-0.15	-0.04	-0.04	0.21
(Constant)	2.86	0.22		12.75	0.00				
T1LC14YSize	0.00	0.00	0.06	0.48	0.63	-0.20	0.03	0.03	-0.20
Deri	-0.03	0.03	-0.12	-1.13	0.26	-0.22	-0.08	-0.07	0.43
Anto	-0.04	0.03	-0.13	-1.19	0.24	-0.22	-0.08	-0.08	0.47
Collo	-0.03	0.03	-0.09	-0.99	0.32	-0.21	-0.07	-0.06	0.30
(Constant)	0.58	0.03		21.90	0.00				
T1LC15YSize	0.00	0.00	-0.09	-0.72	0.48	-0.24	-0.05	-0.05	0.29
Deri	0.00	0.00	0.01	0.09	0.93	-0.19	0.01	0.01	-0.02
Anto	-0.01	0.00	-0.23	-2.08	0.04	-0.27	-0.14	-0.13	0.83
Collo	0.00	0.00	0.05	0.54	0.59	-0.15	0.04	0.03	-0.10

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.48	0.23	0.21	13.08	16.12	0.00	1.60
T1F3	0.41	0.17	0.15	2.37	10.89	0.00	1.60
T1F4	0.18	0.03	0.02	5.49	1.86	0.12	1.85
T1F5	0.09	0.01	-0.01	0.17	0.50	0.74	1.98
T1F6	0.09	0.01	-0.01	0.98	0.42	0.79	1.94
T1A1	0.35	0.12	0.11	0.26	7.66	0.00	2.03
T1A2	0.33	0.11	0.09	0.26	6.71	0.00	2.09
T1SC1	0.15	0.02	0.01	0.21	1.35	0.25	2.26
T1SC2	0.12	0.01	0.00	0.18	0.79	0.53	2.27
T1SC3	0.13	0.02	0.00	0.11	1.02	0.40	2.27
T1SC4	0.22	0.05	0.03	1.41	2.85	0.02	1.92
T1LC1	0.38	0.14	0.13	0.53	9.22	0.00	1.72
T1LC2	0.08	0.01	-0.01	1.28	0.40	0.81	2.01
T1LC3	0.18	0.03	0.01	1.39	1.78	0.13	2.04
T1LC4	0.38	0.15	0.13	0.09	9.34	0.00	1.83
T1LC5	0.13	0.02	0.00	0.41	0.90	0.46	1.96
T1LC6	0.17	0.03	0.01	0.81	1.57	0.18	2.02
T1LC7	0.24	0.06	0.04	0.92	3.24	0.01	2.03
T1LC8	0.27	0.07	0.05	0.10	4.25	0.00	1.90
T1LC11X	0.23	0.05	0.04	0.05	3.18	0.01	1.76
T1LC12X	0.18	0.03	0.01	0.21	1.79	0.13	2.10
T1LC13X	0.16	0.02	0.01	0.36	1.36	0.25	0.16
T1LC14X	0.18	0.03	0.02	0.36	1.88	0.11	1.82
T1LC15X	0.20	0.04	0.02	0.06	2.35	0.06	1.94
T1LC11Y	0.39	0.16	0.14	0.09	10.14	0.00	1.69
T1LC12Y	0.11	0.01	-0.01	0.38	0.66	0.62	1.87
T1LC13Y	0.19	0.04	0.02	0.78	2.00	0.10	1.93
T1LC14Y	0.25	0.06	0.05	0.85	3.77	0.01	1.95
T1LC15Y	0.28	0.08	0.06	0.10	4.62	0.00	1.76

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (4, 220); DW = Durbin-Watson.

Appendix 12.10

*Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (*n = *225) in Task 4*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	16.93	3.50		4.84	0.00				
T1F1Size	0.01	0.00	0.18	1.53	0.13	0.47	0.10	0.09	0.31
Deri	0.05	0.40	0.01	0.12	0.91	0.40	0.01	0.01	0.02
Anto	1.86	0.53	0.35	3.53	0.00	0.51	0.23	0.20	0.67
Collo	0.02	0.44	0.00	0.03	0.97	0.35	0.00	0.00	0.00
(Constant)	2.51	0.70		3.61	0.00				
T1F3Size	0.00	0.00	0.10	0.82	0.42	0.43	0.05	0.05	0.19
Deri	0.02	0.08	0.03	0.31	0.76	0.37	0.02	0.02	0.05
Anto	0.32	0.10	0.32	3.10	0.00	0.46	0.20	0.18	0.66
Collo	0.07	0.09	0.07	0.83	0.40	0.35	0.06	0.05	0.11
(Constant)	1.12	1.24		0.91	0.37				
T1F4Size	0.00	0.00	0.06	0.44	0.66	0.25	0.03	0.03	0.19
Deri	0.17	0.14	0.13	1.18	0.24	0.25	0.08	0.08	0.42
Anto	0.17	0.19	0.10	0.90	0.37	0.25	0.06	0.06	0.33
Collo	0.03	0.15	0.02	0.21	0.83	0.20	0.01	0.01	0.05
(Constant)	0.12	0.03		3.47	0.00				
T1F5Size	0.00	0.00	0.03	0.23	0.82	0.05	0.02	0.02	0.30
Deri	0.00	0.00	0.08	0.69	0.49	0.07	0.05	0.05	0.90
Anto	0.00	0.01	-0.01	-0.06	0.95	0.04	0.00	0.00	-0.06
Collo	0.00	0.00	-0.05	-0.51	0.61	0.02	-0.03	-0.03	-0.15
(Constant)	0.69	0.21		3.22	0.00				
T1F6Size	0.00	0.00	0.08	0.57	0.57	0.10	0.04	0.04	0.42
Deri	0.02	0.02	0.10	0.89	0.37	0.11	0.06	0.06	0.57
Anto	0.01	0.03	0.03	0.22	0.83	0.09	0.01	0.01	0.12
Collo	-0.03	0.03	-0.11	-1.18	0.24	0.02	-0.08	-0.08	-0.12

(Appendix 12.10 continues)

(Appendix 12.10 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.29	0.07		4.47	0.00				
T1A1Size	0.00	0.00	0.17	1.35	0.18	0.31	0.09	0.09	0.51
Deri	0.00	0.01	0.00	0.00	1.00	0.26	0.00	0.00	0.00
Anto	0.01	0.01	0.15	1.32	0.19	0.31	0.09	0.08	0.42
Collo	0.00	0.01	0.03	0.34	0.74	0.24	0.02	0.02	0.07
(Constant)	0.37	0.06		6.05	0.00				
T1A2Size	0.00	0.00	0.21	1.59	0.11	0.26	0.11	0.10	0.75
Deri	0.00	0.01	-0.04	-0.40	0.69	0.19	-0.03	-0.03	-0.11
Anto	0.01	0.01	0.11	0.99	0.33	0.25	0.07	0.06	0.38
Collo	0.00	0.01	-0.01	-0.08	0.94	0.17	-0.01	-0.01	-0.02
(Constant)	0.84	0.05		18.14	0.00				
T1SC1Size	0.00	0.00	0.11	0.81	0.42	0.26	0.05	0.05	0.35
Deri	0.01	0.01	0.15	1.43	0.15	0.27	0.10	0.09	0.51
Anto	0.00	0.01	0.04	0.31	0.76	0.24	0.02	0.02	0.10
Collo	0.00	0.01	0.01	0.15	0.88	0.20	0.01	0.01	0.03
(Constant)	0.00	0.03		0.02	0.99				
T1SC2Size	0.00	0.00	0.22	1.62	0.11	0.19	0.11	0.11	0.98
Deri	0.00	0.00	0.10	0.91	0.36	0.17	0.06	0.06	0.42
Anto	0.00	0.00	-0.10	-0.83	0.41	0.12	-0.06	-0.05	-0.28
Collo	0.00	0.00	-0.05	-0.51	0.61	0.10	-0.03	-0.03	-0.12
(Constant)	0.00	0.02		0.14	0.89				
T1SC3Size	0.00	0.00	0.24	1.83	0.07	0.19	0.12	0.12	1.06
Deri	0.00	0.00	0.09	0.83	0.41	0.17	0.06	0.05	0.35
Anto	0.00	0.00	-0.10	-0.87	0.39	0.12	-0.06	-0.06	-0.27
Collo	0.00	0.00	-0.07	-0.73	0.47	0.09	-0.05	-0.05	-0.14
(Constant)	5.37	0.35		15.28	0.00				
T1SC4Size	0.00	0.00	0.28	2.15	0.03	0.24	0.14	0.14	0.89
Deri	0.02	0.04	0.06	0.60	0.55	0.19	0.04	0.04	0.16
Anto	0.02	0.05	0.04	0.40	0.69	0.20	0.03	0.03	0.12
Collo	-0.09	0.04	-0.19	-2.05	0.04	0.07	-0.14	-0.13	-0.17
(Constant)	3.08	0.13		24.64	0.00				
T1LC1Size	0.00	0.00	0.29	2.34	0.02	0.42	0.16	0.14	0.61
Deri	0.00	0.01	0.02	0.15	0.88	0.34	0.01	0.01	0.03
Anto	0.04	0.02	0.23	2.16	0.03	0.41	0.14	0.13	0.48
Collo	-0.02	0.02	-0.09	-1.04	0.30	0.25	-0.07	-0.06	-0.11
(Constant)	4.45	0.29		15.29	0.00	0.05	0.13	0.13	0.34
T1LC2Size	0.00	0.00	0.25	1.91	0.06	0.03	0.04	0.04	0.05
Deri	0.02	0.03	0.06	0.57	0.57	-0.01	-0.07	-0.07	0.05
Anto	-0.05	0.04	-0.12	-1.05	0.29	-0.10	-0.16	-0.16	0.56
Collo	-0.09	0.04	-0.23	-2.48	0.01	0.05	0.13	0.13	0.34
(Constant)	5.80	0.47		12.42	0.00				
T1LC3Size	0.00	0.00	0.13	1.00	0.32	-0.15	0.07	0.07	-0.35
Deri	-0.03	0.05	-0.06	-0.58	0.56	-0.17	-0.04	-0.04	0.19
Anto	-0.08	0.07	-0.13	-1.18	0.24	-0.18	-0.08	-0.08	0.45
Collo	-0.11	0.06	-0.18	-1.96	0.05	-0.22	-0.13	-0.13	0.71
(Constant)	0.47	0.02		22.53	0.00				
T1LC4Size	0.00	0.00	-0.06	-0.47	0.64	-0.27	-0.03	-0.03	0.18
Deri	0.00	0.00	0.00	-0.03	0.98	-0.23	0.00	0.00	0.01
Anto	-0.01	0.00	-0.22	-1.97	0.05	-0.30	-0.13	-0.13	0.71
Collo	0.00	0.00	-0.05	-0.50	0.62	-0.22	-0.03	-0.03	0.11
(Constant)	1.78	0.11		16.88	0.00				
T1LC5Size	0.00	0.00	0.16	1.23	0.22	0.26	0.08	0.08	0.54
Deri	0.00	0.01	0.03	0.26	0.80	0.21	0.02	0.02	0.08
Anto	0.02	0.02	0.15	1.37	0.17	0.26	0.09	0.09	0.52
Collo	-0.01	0.01	-0.07	-0.78	0.44	0.15	-0.05	-0.05	-0.14
(Constant)	2.51	0.17		15.00	0.00				
T1LC6Size	0.00	0.00	0.20	1.47	0.14	0.02	0.10	0.10	0.12
Deri	0.01	0.02	0.07	0.66	0.51	0.01	0.04	0.04	0.03
Anto	-0.02	0.03	-0.11	-0.97	0.33	-0.04	-0.07	-0.06	0.13
Collo	-0.05	0.02	-0.21	-2.30	0.02	-0.11	-0.15	-0.15	0.72
(Constant)	2.71	0.15		17.53	0.00				
T1LC7Size	0.00	0.00	-0.01	-0.07	0.94	-0.16	0.04	0.04	-0.24
Deri	0.01	0.02	0.06	0.53	0.60	-0.17	-0.04	-0.03	0.20
Anto	0.00	0.02	-0.01	-0.07	0.94	-0.19	-0.07	-0.07	0.43
Collo	-0.04	0.02	-0.17	-1.84	0.07	-0.21	-0.11	-0.10	0.60

(Appendix 12.10 continues)

(Appendix 12.10 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.57	0.02		25.84	0.00				
T1LC8Size	0.00	0.00	-0.07	-0.49	0.63	-0.07	-0.03	-0.03	0.96
Deri	0.00	0.00	0.01	0.12	0.91	-0.05	0.01	0.01	-0.13
Anto	0.00	0.00	-0.02	-0.14	0.89	-0.06	-0.01	-0.01	0.21
Collo	0.00	0.00	0.00	0.04	0.97	-0.04	0.00	0.00	-0.04
(Constant)	0.10	0.01		8.73	0.00				
T1LC11XSize	0.00	0.00	-0.34	-2.55	0.01	-0.21	-0.17	-0.17	1.06
Deri	0.00	0.00	-0.02	-0.20	0.84	-0.15	-0.01	-0.01	0.05
Anto	0.00	0.00	0.01	0.05	0.96	-0.15	0.00	0.00	-0.01
Collo	0.00	0.00	0.20	2.21	0.03	-0.03	0.15	0.14	-0.10
(Constant)	0.40	0.05		8.07	0.00				
T1LC12XSize	0.00	0.00	-0.28	-2.08	0.04	-0.02	-0.14	-0.14	0.17
Deri	0.00	0.01	-0.02	-0.14	0.89	0.01	-0.01	-0.01	0.00
Anto	0.01	0.01	0.15	1.30	0.19	0.05	0.09	0.09	0.19
Collo	0.01	0.01	0.22	2.34	0.02	0.12	0.16	0.15	0.65
(Constant)	0.58	0.07		8.24	0.00				
T1LC13XSize	0.00	0.00	-0.14	-1.05	0.29	-0.10	-0.07	-0.07	1.18
Deri	0.00	0.01	0.01	0.10	0.92	-0.07	0.01	0.01	-0.06
Anto	0.00	0.01	-0.01	-0.06	0.95	-0.08	0.00	0.00	0.05
Collo	0.01	0.01	0.06	0.62	0.54	-0.03	0.04	0.04	-0.16
(Constant)	0.73	0.08		8.69	0.00				
T1LC14XSize	0.00	0.00	-0.21	-1.57	0.12	-0.20	-0.11	-0.10	0.98
Deri	-0.01	0.01	-0.06	-0.53	0.60	-0.17	-0.04	-0.03	0.23
Anto	0.00	0.01	0.03	0.26	0.80	-0.15	0.02	0.02	-0.10
Collo	0.00	0.01	0.04	0.44	0.66	-0.12	0.03	0.03	-0.11
(Constant)	0.13	0.01		8.53	0.00				
T1LC15XSize	0.00	0.00	-0.36	-2.72	0.01	-0.13	-0.18	-0.18	0.84
Deri	0.00	0.00	0.00	0.04	0.97	-0.07	0.00	0.00	-0.01
Anto	0.00	0.00	0.07	0.62	0.53	-0.06	0.04	0.04	-0.08
Collo	0.00	0.00	0.25	2.71	0.01	0.06	0.18	0.18	0.24
(Constant)	0.43	0.02		22.35	0.00				
T1LC11YSize	0.00	0.00	-0.12	-0.89	0.38	-0.28	-0.06	-0.06	0.37
Deri	0.00	0.00	-0.06	-0.60	0.55	-0.26	-0.04	-0.04	0.19
Anto	0.00	0.00	-0.07	-0.67	0.51	-0.26	-0.04	-0.04	0.22
Collo	0.00	0.00	-0.08	-0.90	0.37	-0.25	-0.06	-0.06	0.23
(Constant)	1.63	0.10		16.91	0.00				
T1LC12YSize	0.00	0.00	0.11	0.82	0.42	0.23	0.05	0.05	0.30
Deri	-0.01	0.01	-0.06	-0.52	0.60	0.16	-0.04	-0.03	-0.11
Anto	0.04	0.01	0.28	2.52	0.01	0.27	0.17	0.16	0.95
Collo	-0.01	0.01	-0.09	-1.02	0.31	0.12	-0.07	-0.07	-0.13
(Constant)	2.30	0.15		15.18	0.00				
T1LC13YSize	0.00	0.00	0.16	1.22	0.22	0.01	0.08	0.08	0.05
Deri	0.00	0.02	0.01	0.10	0.92	-0.02	0.01	0.01	-0.01
Anto	0.00	0.02	0.00	0.01	0.99	-0.01	0.00	0.00	0.00
Collo	-0.05	0.02	-0.24	-2.62	0.01	-0.13	-0.17	-0.17	0.96
(Constant)	3.12	0.30		10.41	0.00				
T1LC14YSize	0.00	0.00	0.05	0.41	0.68	-0.16	0.03	0.03	-0.18
Deri	-0.03	0.03	-0.09	-0.81	0.42	-0.18	-0.05	-0.05	0.32
Anto	-0.02	0.05	-0.05	-0.48	0.63	-0.17	-0.03	-0.03	0.19
Collo	-0.06	0.04	-0.16	-1.72	0.09	-0.21	-0.11	-0.11	0.67
(Constant)	0.53	0.02		24.61	0.00				
T1LC15YSize	0.00	0.00	-0.11	-0.78	0.44	-0.08	-0.05	-0.05	0.43
Deri	0.00	0.00	-0.09	-0.83	0.40	-0.09	-0.06	-0.06	0.45
Anto	0.00	0.00	0.17	1.47	0.14	-0.01	0.10	0.10	-0.12
Collo	0.00	0.00	-0.06	-0.59	0.55	-0.08	-0.04	-0.04	0.23

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.52	0.27	0.26	13.31	20.31	0.00	1.50
T1F3	0.47	0.23	0.21	2.65	15.99	0.00	1.53
T1F4	0.27	0.08	0.06	4.71	4.49	0.00	2.01
T1F5	0.07	0.01	-0.01	0.13	0.31	0.87	2.00
T1F6	0.13	0.02	0.00	0.81	1.01	0.40	1.95
T1A1	0.33	0.11	0.09	0.25	6.58	0.00	1.97
T1A2	0.27	0.07	0.06	0.23	4.30	0.00	2.11
T1SC1	0.28	0.08	0.06	0.18	4.79	0.00	1.85
T1SC2	0.20	0.04	0.02	0.10	2.36	0.05	1.67
T1SC3	0.21	0.04	0.03	0.08	2.49	0.04	1.67
T1SC4	0.28	0.08	0.06	1.34	4.58	0.00	1.69
T1LC1	0.44	0.20	0.18	0.48	13.52	0.00	1.68
T1LC2	0.20	0.04	0.02	1.11	2.22	0.07	1.67
T1LC3	0.24	0.06	0.04	1.78	3.22	0.01	1.98
T1LC4	0.30	0.09	0.08	0.08	5.64	0.00	1.96
T1LC5	0.28	0.08	0.06	0.40	4.57	0.00	2.06
T1LC6	0.18	0.03	0.01	0.64	1.79	0.13	1.92
T1LC7	0.22	0.05	0.03	1.26	2.91	0.02	2.02
T1LC8	0.07	0.00	-0.01	0.08	0.25	0.91	1.91
T1LC11X	0.26	0.07	0.05	0.05	3.96	0.00	1.83
T1LC12X	0.20	0.04	0.02	0.19	2.23	0.07	1.83
T1LC13X	0.11	0.01	-0.01	0.27	0.68	0.61	1.71
T1LC14X	0.21	0.04	0.03	0.32	2.50	0.04	1.84
T1LC15X	0.24	0.06	0.04	0.06	3.33	0.01	1.76
T1LC11Y	0.30	0.09	0.07	0.07	5.33	0.00	2.13
T1LC12Y	0.29	0.08	0.07	0.37	4.91	0.00	2.26
T1LC13Y	0.18	0.03	0.01	0.58	1.84	0.12	2.06
T1LC14Y	0.22	0.05	0.03	1.14	2.91	0.02	2.15
T1LC15Y	0.14	0.02	0.00	0.08	1.05	0.38	2.03

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (4, 220); DW = Durbin-Watson.

Appendix 12.11

*Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (*n = *225) in Task 5*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	24.10	3.83		6.29	0.00				
T1F1Size	0.00	0.00	0.11	0.88	0.38	0.38	0.06	0.05	0.23
Deri	0.04	0.44	0.01	0.08	0.94	0.31	0.01	0.00	0.01
Anto	1.87	0.58	0.34	3.23	0.00	0.42	0.21	0.20	0.80
Collo	-0.14	0.48	-0.02	-0.29	0.77	0.27	-0.02	-0.02	-0.04
(Constant)	3.24	0.58		5.61	0.00				
T1F3Size	0.00	0.00	0.10	0.84	0.40	0.34	0.06	0.05	0.22
Deri	-0.04	0.07	-0.06	-0.62	0.53	0.26	-0.04	-0.04	-0.10
Anto	0.30	0.09	0.37	3.39	0.00	0.39	0.22	0.21	0.91
Collo	-0.02	0.07	-0.02	-0.23	0.82	0.24	-0.02	-0.01	-0.03
(Constant)	1.68	1.43		1.17	0.24				
T1F4Size	0.00	0.00	0.22	1.69	0.09	0.29	0.11	0.11	0.70
Deri	-0.10	0.16	-0.06	-0.61	0.54	0.21	-0.04	-0.04	-0.15
Anto	0.26	0.22	0.14	1.23	0.22	0.28	0.08	0.08	0.42
Collo	0.03	0.18	0.01	0.14	0.89	0.20	0.01	0.01	0.03
(Constant)	0.11	0.04		3.06	0.00	0.13	0.11	0.11	1.43
T1F5Size	0.00	0.00	0.22	1.64	0.10	0.08	-0.04	-0.04	-0.22
Deri	0.00	0.00	-0.06	-0.54	0.59	0.09	-0.02	-0.02	-0.15
Anto	0.00	0.01	-0.03	-0.30	0.77	0.07	-0.01	-0.01	-0.06
Collo	0.00	0.00	-0.02	-0.19	0.85	0.13	0.11	0.11	1.43
(Constant)	0.70	0.29		2.42	0.02				
T1F6Size	0.00	0.00	0.24	1.81	0.07	0.17	0.12	0.12	1.34
Deri	0.00	0.03	-0.01	-0.13	0.90	0.12	-0.01	-0.01	-0.06
Anto	-0.02	0.04	-0.05	-0.43	0.67	0.12	-0.03	-0.03	-0.19
Collo	-0.01	0.04	-0.03	-0.35	0.73	0.09	-0.02	-0.02	-0.09

(Appendix 12.11 continues)

(Appendix 12.11 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.11	0.07		1.58	0.12				
T1A1Size	0.00	0.00	0.39	3.02	0.00	0.31	0.20	0.19	1.12
Deri	-0.01	0.01	-0.15	-1.40	0.16	0.19	-0.09	-0.09	-0.25
Anto	0.01	0.01	0.08	0.69	0.49	0.27	0.05	0.04	0.19
Collo	0.00	0.01	-0.03	-0.37	0.71	0.18	-0.02	-0.02	-0.05
(Constant)	0.12	0.06		1.96	0.05				
T1A2Size	0.00	0.00	0.40	3.11	0.00	0.32	0.21	0.20	1.15
Deri	-0.01	0.01	-0.16	-1.54	0.12	0.18	-0.10	-0.10	-0.27
Anto	0.00	0.01	0.05	0.47	0.64	0.26	0.03	0.03	0.12
Collo	0.00	0.01	0.00	0.00	1.00	0.20	0.00	0.00	0.00
(Constant)	0.76	0.07		10.36	0.00				
T1SC1Size	0.00	0.00	0.16	1.22	0.22	0.21	0.08	0.08	0.71
Deri	0.00	0.01	-0.05	-0.48	0.64	0.15	-0.03	-0.03	-0.16
Anto	0.01	0.01	0.11	0.93	0.35	0.20	0.06	0.06	0.45
Collo	0.00	0.01	0.00	0.00	1.00	0.14	0.00	0.00	0.00
(Constant)	0.01	0.04		0.12	0.91				
T1SC2Size	0.00	0.00	-0.09	-0.64	0.52	0.06	-0.04	-0.04	-0.35
Deri	0.01	0.00	0.14	1.25	0.21	0.11	0.08	0.08	1.00
Anto	0.00	0.01	0.00	0.01	0.99	0.06	0.00	0.00	0.01
Collo	0.00	0.01	0.06	0.62	0.54	0.09	0.04	0.04	0.34
(Constant)	0.00	0.03		0.01	0.99				
T1SC3Size	0.00	0.00	-0.08	-0.60	0.55	0.09	-0.04	-0.04	-0.34
Deri	0.00	0.00	0.16	1.41	0.16	0.14	0.09	0.09	1.01
Anto	0.00	0.00	0.02	0.19	0.85	0.09	0.01	0.01	0.10
Collo	0.00	0.00	0.05	0.51	0.61	0.11	0.03	0.03	0.24
(Constant)	5.77	0.46		12.50	0.00				
T1SC4Size	0.00	0.00	0.28	2.15	0.03	0.27	0.14	0.14	1.00
Deri	0.03	0.05	0.06	0.57	0.57	0.23	0.04	0.04	0.18
Anto	-0.01	0.07	-0.02	-0.15	0.88	0.21	-0.01	-0.01	-0.05
Collo	-0.04	0.06	-0.07	-0.71	0.48	0.15	-0.05	-0.05	-0.13
(Constant)	2.93	0.13		22.29	0.00				
T1LC1Size	0.00	0.00	0.00	0.01	0.99	0.32	0.00	0.00	0.00
Deri	0.02	0.01	0.17	1.64	0.10	0.33	0.11	0.10	0.42
Anto	0.04	0.02	0.22	2.01	0.05	0.34	0.13	0.13	0.56
Collo	0.00	0.02	0.01	0.09	0.93	0.25	0.01	0.01	0.01
(Constant)	3.90	0.32		12.37	0.00				
T1LC2Size	0.00	0.00	0.17	1.31	0.19	0.17	0.09	0.09	0.74
Deri	0.05	0.04	0.14	1.32	0.19	0.18	0.09	0.09	0.64
Anto	-0.05	0.05	-0.12	-1.04	0.30	0.10	-0.07	-0.07	-0.31
Collo	-0.01	0.04	-0.03	-0.29	0.77	0.11	-0.02	-0.02	-0.07
(Constant)	6.20	0.70		8.83	0.00				
T1LC3Size	0.00	0.00	-0.15	-1.09	0.28	-0.12	-0.07	-0.07	0.81
Deri	0.09	0.08	0.12	1.12	0.27	-0.05	0.08	0.07	-0.28
Anto	-0.08	0.11	-0.09	-0.74	0.46	-0.12	-0.05	-0.05	0.47
Collo	0.00	0.09	0.00	0.00	1.00	-0.07	0.00	0.00	0.00
(Constant)	0.38	0.03		14.82	0.00				
T1LC4Size	0.00	0.00	-0.28	-2.10	0.04	-0.23	-0.14	-0.14	1.05
Deri	0.00	0.00	0.10	0.95	0.34	-0.13	0.06	0.06	-0.23
Anto	0.00	0.00	-0.09	-0.79	0.43	-0.20	-0.05	-0.05	0.30
Collo	0.00	0.00	0.07	0.72	0.47	-0.11	0.05	0.05	-0.12
(Constant)	1.66	0.12		14.25	0.00				
T1LC5Size	0.00	0.00	-0.23	-1.70	0.09	0.05	-0.11	-0.11	-0.39
Deri	0.02	0.01	0.15	1.37	0.17	0.11	0.09	0.09	0.61
Anto	0.03	0.02	0.18	1.54	0.13	0.11	0.10	0.10	0.72
Collo	0.00	0.01	0.02	0.23	0.82	0.08	0.02	0.02	0.06
(Constant)	2.16	0.19		11.60	0.00				
T1LC6Size	0.00	0.00	-0.01	-0.11	0.91	0.05	-0.01	-0.01	-0.08
Deri	0.03	0.02	0.14	1.28	0.20	0.09	0.09	0.09	1.28
Anto	-0.01	0.03	-0.03	-0.28	0.78	0.04	-0.02	-0.02	-0.12
Collo	-0.01	0.02	-0.03	-0.29	0.77	0.03	-0.02	-0.02	-0.09
(Constant)	3.48	0.42		8.22	0.00				
T1LC7Size	0.00	0.00	-0.24	-1.82	0.07	-0.17	-0.12	-0.12	1.17
Deri	0.06	0.05	0.13	1.22	0.22	-0.08	0.08	0.08	-0.30
Anto	-0.02	0.06	-0.03	-0.27	0.79	-0.14	-0.02	-0.02	0.12
Collo	0.00	0.05	0.00	-0.04	0.97	-0.10	0.00	0.00	0.01

(Appendix 12.11 continues)

(Appendix 12.11 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.56	0.03		20.42	0.00				
T1LC8Size	0.00	0.00	-0.29	-2.17	0.03	-0.19	-0.14	-0.14	1.41
Deri	0.00	0.00	0.04	0.37	0.71	-0.12	0.02	0.02	-0.13
Anto	0.00	0.00	0.08	0.70	0.48	-0.12	0.05	0.05	-0.25
Collo	0.00	0.00	0.01	0.13	0.90	-0.11	0.01	0.01	-0.03
(Constant)	0.10	0.01		9.13	0.00				
T1LC11XSize	0.00	0.00	-0.20	-1.50	0.14	-0.17	-0.10	-0.10	0.87
Deri	0.00	0.00	0.15	1.39	0.17	-0.09	0.09	0.09	-0.33
Anto	0.00	0.00	-0.06	-0.55	0.58	-0.16	-0.04	-0.04	0.25
Collo	0.00	0.00	-0.06	-0.66	0.51	-0.14	-0.04	-0.04	0.21
(Constant)	0.44	0.05		8.21	0.00				
T1LC12XSize	0.00	0.00	-0.15	-1.13	0.26	-0.02	-0.08	-0.08	0.20
Deri	0.01	0.01	0.18	1.64	0.10	0.05	0.11	0.11	0.65
Anto	0.00	0.01	0.05	0.43	0.67	0.01	0.03	0.03	0.03
Collo	0.00	0.01	-0.07	-0.74	0.46	-0.03	-0.05	-0.05	0.12
(Constant)	0.59	0.07		7.87	0.00				
T1LC13XSize	0.00	0.00	-0.03	-0.25	0.80	-0.03	-0.02	-0.02	0.08
Deri	0.01	0.01	0.15	1.36	0.18	0.02	0.09	0.09	0.25
Anto	-0.01	0.01	-0.06	-0.50	0.62	-0.04	-0.03	-0.03	0.20
Collo	-0.01	0.01	-0.10	-1.05	0.29	-0.06	-0.07	-0.07	0.48
(Constant)	0.82	0.11		7.64	0.00				
T1LC14XSize	0.00	0.00	-0.24	-1.82	0.07	-0.16	-0.12	-0.12	1.16
Deri	0.02	0.01	0.15	1.32	0.19	-0.07	0.09	0.09	-0.31
Anto	0.00	0.02	-0.02	-0.14	0.89	-0.13	-0.01	-0.01	0.06
Collo	0.00	0.01	-0.03	-0.30	0.77	-0.11	-0.02	-0.02	0.09
(Constant)	0.15	0.02		9.22	0.00				
T1LC15XSize	0.00	0.00	-0.16	-1.19	0.23	-0.12	-0.08	-0.08	0.89
Deri	0.00	0.00	0.13	1.17	0.24	-0.05	0.08	0.08	-0.31
Anto	0.00	0.00	-0.02	-0.13	0.90	-0.10	-0.01	-0.01	0.07
Collo	0.00	0.00	-0.07	-0.75	0.45	-0.11	-0.05	-0.05	0.34
(Constant)	0.37	0.02		15.57	0.00				
T1LC11YSize	0.00	0.00	-0.24	-1.83	0.07	-0.26	-0.12	-0.12	0.68
Deri	0.00	0.00	0.13	1.22	0.22	-0.16	0.08	0.08	-0.22
Anto	-0.01	0.00	-0.22	-2.00	0.05	-0.27	-0.13	-0.13	0.66
Collo	0.00	0.00	0.09	0.94	0.35	-0.13	0.06	0.06	-0.12
(Constant)	1.62	0.11		14.57	0.00				
T1LC12YSize	0.00	0.00	-0.20	-1.47	0.14	0.00	-0.10	-0.10	0.03
Deri	0.02	0.01	0.16	1.42	0.16	0.07	0.10	0.09	0.69
Anto	0.01	0.02	0.04	0.35	0.72	0.03	0.02	0.02	0.07
Collo	0.01	0.01	0.06	0.65	0.51	0.05	0.04	0.04	0.21
(Constant)	2.14	0.16		13.11	0.00				
T1LC13YSize	0.00	0.00	0.02	0.17	0.86	0.00	0.01	0.01	0.00
Deri	0.03	0.02	0.15	1.38	0.17	0.05	0.09	0.09	0.51
Anto	-0.03	0.02	-0.16	-1.40	0.16	-0.05	-0.09	-0.09	0.48
Collo	0.00	0.02	-0.01	-0.12	0.90	0.00	-0.01	-0.01	0.00
(Constant)	3.45	0.45		7.61	0.00				
T1LC14YSize	0.00	0.00	-0.25	-1.89	0.06	-0.18	-0.13	-0.12	1.12
Deri	0.07	0.05	0.14	1.30	0.20	-0.08	0.09	0.09	-0.29
Anto	-0.03	0.07	-0.06	-0.48	0.63	-0.15	-0.03	-0.03	0.21
Collo	0.01	0.06	0.02	0.19	0.85	-0.10	0.01	0.01	-0.04
(Constant)	0.55	0.03		20.02	0.00				
T1LC15YSize	0.00	0.00	-0.25	-1.92	0.06	-0.23	-0.13	-0.13	0.96
Deri	0.00	0.00	0.07	0.68	0.50	-0.15	0.05	0.04	-0.18
Anto	0.00	0.00	-0.10	-0.89	0.37	-0.21	-0.06	-0.06	0.35
Collo	0.00	0.00	0.07	0.73	0.47	-0.12	0.05	0.05	-0.13

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.43	0.18	0.17	14.60	12.17	0.00	1.66
T1F3	0.40	0.16	0.14	2.20	10.30	0.00	1.58
T1F4	0.30	0.09	0.07	5.44	5.52	0.00	1.82
T1F5	0.14	0.02	0.00	0.14	1.17	0.33	2.10
T1F6	0.18	0.03	0.01	1.10	1.76	0.14	2.02
T1A1	0.33	0.11	0.09	0.26	6.72	0.00	2.04
T1A2	0.33	0.11	0.09	0.23	6.84	0.00	2.04
T1SC1	0.22	0.05	0.03	0.28	2.75	0.03	1.78
T1SC2	0.12	0.02	0.00	0.17	0.85	0.50	2.03
T1SC3	0.15	0.02	0.00	0.10	1.21	0.31	2.01
T1SC4	0.28	0.08	0.06	1.76	4.57	0.00	1.58
T1LC1	0.36	0.13	0.12	0.50	8.44	0.00	1.96
T1LC2	0.20	0.04	0.02	1.20	2.32	0.06	1.91
T1LC3	0.15	0.02	0.00	2.67	1.22	0.30	1.85
T1LC4	0.24	0.06	0.04	0.10	3.49	0.01	1.95
T1LC5	0.16	0.03	0.01	0.44	1.54	0.19	2.25
T1LC6	0.10	0.01	-0.01	0.71	0.55	0.70	2.00
T1LC7	0.19	0.03	0.02	1.61	1.98	0.10	1.99
T1LC8	0.20	0.04	0.02	0.10	2.19	0.07	2.18
T1LC11X	0.20	0.04	0.02	0.04	2.30	0.06	1.90
T1LC12X	0.12	0.01	0.00	0.20	0.81	0.52	1.96
T1LC13X	0.11	0.01	-0.01	0.29	0.71	0.59	1.94
T1LC14X	0.18	0.03	0.02	0.41	1.94	0.11	1.86
T1LC15X	0.15	0.02	0.00	0.06	1.22	0.30	1.89
T1LC11Y	0.30	0.09	0.08	0.09	5.57	0.00	1.72
T1LC12Y	0.13	0.02	0.00	0.42	0.89	0.47	2.01
T1LC13Y	0.12	0.02	0.00	0.62	0.85	0.49	1.89
T1LC14Y	0.20	0.04	0.02	1.72	2.25	0.06	1.98
T1LC15Y	0.25	0.06	0.04	0.10	3.63	0.01	1.97

Note. n = 225. SEE = Standard Error of the Estimate; F (4, 220); DW = Durbin-Watson.

Appendix 12.12

Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (outliers excluded) in Task 1

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1							
T1F3	0.26	0.07	0.05	3.39	3.90	0.00	1.67
T1F4	0.23	0.05	0.03	4.66	2.94	0.02	1.73
T1F5	0.20	0.04	0.02	0.10	2.32	0.06	1.70
T1F6	0.18	0.03	0.01	0.57	1.72	0.15	1.71
T1A1	0.41	0.17	0.15	0.17	11.01	0.00	1.96
T1SC2	0.18	0.03	0.01	0.15	1.82	0.13	2.01
T1SC4	0.27	0.07	0.05	1.07	4.19	0.00	1.74
T1LC2	0.22	0.05	0.03	0.86	2.76	0.03	1.85
T1LC3	0.12	0.01	0.00	0.79	0.73	0.57	1.76
T1LC6	0.10	0.01	-0.01	0.59	0.57	0.69	1.79
T1LC7	0.16	0.03	0.01	0.56	1.47	0.21	1.70
T1LC11X	0.14	0.02	0.00	0.04	1.17	0.33	1.52
T1LC12X	0.14	0.02	0.00	0.22	1.04	0.39	1.60
T1LC13X	0.13	0.02	0.00	0.22	0.90	0.46	1.59
T1LC14X	0.12	0.02	0.00	0.19	0.84	0.50	1.56
T1LC15X	0.12	0.02	0.00	0.05	0.85	0.49	1.51
T1LC13Y	0.09	0.01	-0.01	0.55	0.47	0.76	1.83
T1LC14Y	0.05	0.00	-0.02	0.50	0.14	0.97	1.84

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Appendix 12.16

Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (outliers excluded) in Task 2

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F4	0.33	0.11	0.09	5.34	6.81	0.00	2.17
T1F6	0.25	0.06	0.04	0.81	3.57	0.01	2.24
T1A2	0.15	0.02	0.01	0.21	1.30	0.27	2.07
T1SC1	0.39	0.15	0.14	0.19	9.80	0.00	2.08
T1LC2	0.21	0.04	0.03	0.88	2.50	0.04	2.06
T1LC3	0.26	0.07	0.05	1.69	3.87	0.00	1.96
T1LC4	0.43	0.19	0.17	0.10	12.75	0.00	1.76
T1LC6	0.10	0.01	-0.01	0.56	0.53	0.71	2.16
T1LC7	0.27	0.07	0.06	1.08	4.28	0.00	1.93
T1LC8	0.36	0.13	0.11	0.09	7.90	0.00	1.98
T1LC11X	0.20	0.04	0.02	0.03	2.36	0.05	2.03
T1LC14X	0.16	0.03	0.01	0.23	1.41	0.23	2.14
T1LC11Y	0.41	0.17	0.15	0.10	10.82	0.00	1.67
T1LC13Y	0.08	0.01	-0.01	0.61	0.32	0.86	2.03
T1LC14Y	0.30	0.09	0.07	1.17	5.34	0.00	1.81

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Appendix 12.17

Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (outliers excluded) in Task 3

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.51	0.26	0.25	12.69	19.38	0.00	1.63
T1F3	0.41	0.17	0.15	2.30	11.15	0.00	1.64
T1F4	0.19	0.03	0.02	5.27	1.97	0.10	1.86
T1F5	0.12	0.01	0.00	0.15	0.75	0.56	2.05
T1F6	0.11	0.01	-0.01	0.95	0.69	0.60	1.94
T1SC1	0.21	0.04	0.03	0.17	2.43	0.05	2.19
T1SC2	0.16	0.03	0.01	0.14	1.43	0.22	2.18
T1SC3	0.17	0.03	0.01	0.10	1.60	0.17	2.20
T1SC4	0.28	0.08	0.06	1.31	4.57	0.00	1.93
T1LC1	0.43	0.19	0.17	0.50	12.57	0.00	1.72
T1LC2	0.13	0.02	0.00	1.04	0.91	0.46	1.88
T1LC3	0.17	0.03	0.01	1.09	1.63	0.17	1.96
T1LC4	0.44	0.19	0.18	0.09	13.07	0.00	1.65
T1LC5	0.14	0.02	0.00	0.38	1.07	0.37	1.84
T1LC6	0.08	0.01	-0.01	0.66	0.32	0.86	1.94
T1LC7	0.20	0.04	0.02	0.73	2.21	0.07	2.14
T1LC8	0.31	0.10	0.08	0.09	5.85	0.00	1.83
T1LC11X	0.25	0.06	0.04	0.05	3.56	0.01	1.85
T1LC14X	0.15	0.02	0.00	0.30	1.27	0.28	1.90
T1LC11Y	0.42	0.18	0.16	0.08	11.83	0.00	1.73
T1LC12Y	0.14	0.02	0.00	0.37	1.11	0.35	1.87
T1LC13Y	0.10	0.01	-0.01	0.61	0.57	0.69	1.98
T1LC14Y	0.22	0.05	0.03	0.66	2.84	0.03	2.04
T1LC15Y	0.26	0.07	0.05	0.10	3.90	0.00	1.77

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Appendix 12.18

Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (outliers excluded) in Task 4

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F1	0.53	0.28	0.27	13.00	21.19	0.00	1.53
T1F3	0.48	0.23	0.22	2.56	16.70	0.00	1.56
T1F4	0.25	0.06	0.04	4.29	3.51	0.01	1.87
T1F5	0.09	0.01	-0.01	0.11	0.42	0.79	1.93
T1F6	0.13	0.02	0.00	0.69	0.89	0.47	1.84
T1SC1	0.26	0.07	0.05	0.15	3.77	0.01	1.87
T1SC2	0.24	0.06	0.04	0.08	3.46	0.01	1.77
T1SC3	0.25	0.06	0.05	0.07	3.69	0.01	1.54
T1SC4	0.28	0.08	0.06	1.25	4.73	0.00	1.76
T1LC1	0.48	0.23	0.22	0.46	16.41	0.00	1.71
T1LC2	0.19	0.04	0.02	1.01	2.12	0.08	1.83
T1LC3	0.24	0.06	0.04	1.37	3.39	0.01	1.77
T1LC5	0.28	0.08	0.06	0.40	4.73	0.00	2.06
T1LC6	0.18	0.03	0.02	0.55	1.86	0.12	1.84
T1LC7	0.17	0.03	0.01	0.78	1.53	0.20	1.87
T1LC11X	0.30	0.09	0.07	0.04	5.35	0.00	1.82
T1LC12X	0.19	0.04	0.02	0.18	2.15	0.08	1.84
T1LC13X	0.15	0.02	0.00	0.25	1.21	0.31	1.72
T1LC14X	0.23	0.05	0.04	0.28	3.08	0.02	1.65
T1LC15X	0.25	0.06	0.04	0.05	3.59	0.01	1.75
T1LC11Y	0.28	0.08	0.06	0.07	4.61	0.00	2.16
T1LC13Y	0.21	0.05	0.03	0.53	2.59	0.04	1.90
T1LC14Y	0.18	0.03	0.02	0.79	1.91	0.11	2.01

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Appendix 12.19

Regression analysis summary for productive vocabulary knowledge predicting each Speaking Performance Measures (outliers excluded) in Task 5

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T1F4	0.30	0.09	0.08	5.31	5.54	0.00	1.85
T1F5	0.16	0.02	0.01	0.13	1.36	0.25	2.11
T1F6	0.25	0.06	0.05	0.94	3.63	0.01	1.91
T1SC1	0.29	0.09	0.07	0.25	5.07	0.00	1.78
T1SC2	0.20	0.04	0.02	0.08	2.29	0.06	2.07
T1SC3	0.20	0.04	0.02	0.06	2.18	0.07	2.07
T1SC4	0.32	0.10	0.08	1.65	6.08	0.00	1.56
T1LC2	0.22	0.05	0.03	1.08	2.76	0.03	1.93
T1LC3	0.23	0.05	0.03	1.75	2.89	0.02	1.74
T1LC4	0.22	0.05	0.03	0.09	2.84	0.03	1.92
T1LC6	0.14	0.02	0.00	0.63	1.07	0.37	2.09
T1LC7	0.21	0.05	0.03	1.03	2.57	0.04	1.72
T1LC11X	0.21	0.04	0.02	0.04	2.41	0.05	1.80
T1LC12X	0.16	0.03	0.01	0.19	1.41	0.23	1.84
T1LC13X	0.09	0.01	-0.01	0.26	0.50	0.74	1.85
T1LC14X	0.23	0.05	0.04	0.32	3.10	0.02	1.89
T1LC11Y	0.31	0.10	0.08	0.09	5.89	0.00	1.80
T1LC13Y	0.12	0.02	0.00	0.58	0.83	0.51	1.91
T1LC14Y	0.25	0.06	0.05	1.08	3.68	0.01	1.69

Note. SEE = Standard Error of the Estimate; DW = Durbin-Watson. X = J1001 = The criterion of considering 1,001 to 8,000 words in the JACET8000 as sophisticated. Y = J101 = The criterion of considering 101 to 8,000 words in the JACET8000 as sophisticated.

Appendix 12.20

Regression analysis summary for productive vocabulary knowledge predicting main tokens and types, component scores, and Speaking Performance Measures for each task (n = 225) when size was entered first

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T2 Tokens	0.40	0.16	0.15	11.96	41.31	0.00	1.45
T2 Types	0.35	0.12	0.12	4.42	30.68	0.00	1.57
T3 Tokens	0.44	0.20	0.19	9.93	54.75	0.00	1.58
T3 Types	0.43	0.19	0.18	5.62	50.71	0.00	1.60
T4 Tokens	0.47	0.23	0.22	10.21	64.86	0.00	1.48
T4 Types	0.50	0.25	0.24	5.35	72.42	0.00	1.48
T5 Tokens	0.38	0.14	0.14	11.14	36.63	0.00	1.64
T5 Types	0.40	0.16	0.16	5.26	42.92	0.00	1.61
PF1	0.45	0.21	0.20	0.89	57.96	0.00	1.39
PA1	0.36	0.13	0.12	0.94	32.83	0.00	2.00
PSC5	0.48	0.23	0.23	0.88	66.66	0.00	2.01
PLC1	0.41	0.17	0.16	0.91	44.88	0.00	1.61
T1A1	0.36	0.13	0.12	0.19	32.52	0.00	2.03
T1A2	0.35	0.13	0.12	0.19	31.90	0.00	1.97
T2F1	0.40	0.16	0.15	15.94	41.31	0.00	1.45
T2SC1	0.35	0.12	0.12	0.20	31.26	0.00	2.17
T2SC4	0.37	0.14	0.13	1.62	34.93	0.00	2.01
T2LC4	0.39	0.15	0.15	0.10	40.55	0.00	1.69
T2LC11J101	0.37	0.14	0.13	0.10	35.79	0.00	1.60
T3F1	0.44	0.20	0.19	13.24	54.76	0.00	1.58
T3F3	0.37	0.13	0.13	2.39	34.48	0.00	1.61
T3LC1	0.35	0.13	0.12	0.53	32.12	0.00	1.73
T3LC4	0.36	0.13	0.13	0.09	33.75	0.00	1.78
T3LC11J101	0.35	0.13	0.12	0.09	31.91	0.00	1.63
T4F1	0.47	0.23	0.22	13.62	64.85	0.00	1.48
T4F3	0.43	0.18	0.18	2.71	49.41	0.00	1.57
T4LC1	0.42	0.18	0.17	0.48	48.36	0.00	1.71
T5F1	0.38	0.14	0.14	14.85	36.62	0.00	1.64
T5F3	0.34	0.11	0.11	2.24	28.49	0.00	1.56

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (1, 223); DW = Durbin-Watson.

Appendix 12.21

Regression analysis summary for productive vocabulary knowledge predicting main tokens and types, component scores, and Speaking Performance Measures for each task (n = 225) when depth was entered first

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
T2 Tokens	0.42	0.18	0.17	11.84	16.09	0.00	1.40
T2 Types	0.40	0.16	0.14	4.35	13.64	0.00	1.56
T3 Tokens	0.47	0.22	0.21	9.82	20.96	0.00	1.59
T3 Types	0.45	0.21	0.19	5.58	19.06	0.00	1.59
T4 Tokens	0.51	0.26	0.25	10.01	26.15	0.00	1.47
T4 Types	0.52	0.27	0.26	5.30	26.59	0.00	1.43
T5 Tokens	0.42	0.18	0.17	10.94	15.99	0.00	1.64
T5 Types	0.45	0.20	0.19	5.15	18.89	0.00	1.64
PF1	0.49	0.24	0.23	0.88	22.89	0.00	1.35
PA1	0.29	0.08	0.07	0.96	6.57	0.00	2.03
PSC5	0.46	0.21	0.20	0.89	19.85	0.00	2.00
PLC1	0.45	0.20	0.19	0.90	18.67	0.00	1.61
T1A1	0.28	0.08	0.07	0.19	6.41	0.00	2.05
T1A2	0.29	0.08	0.07	0.20	6.53	0.00	2.00
T2F1	0.42	0.18	0.17	15.79	16.10	0.00	1.40
T2SC1	0.33	0.11	0.10	0.20	9.08	0.00	2.09
T2SC4	0.40	0.16	0.15	1.60	14.17	0.00	2.03
T2LC4	0.40	0.16	0.15	0.10	14.44	0.00	1.68
T2LC11J101	0.38	0.14	0.13	0.10	12.34	0.00	1.58
T3F1	0.47	0.22	0.21	13.10	20.96	0.00	1.59
T3F3	0.41	0.16	0.15	2.36	14.51	0.00	1.60
T3LC1	0.37	0.14	0.13	0.53	11.99	0.00	1.69
T3LC4	0.37	0.14	0.12	0.09	11.66	0.00	1.83
T3LC11J101	0.39	0.15	0.14	0.09	13.28	0.00	1.69
T4F1	0.51	0.26	0.25	13.35	26.15	0.00	1.47
T4F3	0.47	0.22	0.21	2.65	21.14	0.00	1.51
T4LC1	0.42	0.18	0.17	0.48	15.87	0.00	1.64
T5F1	0.42	0.18	0.17	14.59	15.98	0.00	1.64
T5F3	0.39	0.16	0.14	2.20	13.52	0.00	1.57

Note. *n* = 225. SEE = Standard Error of the Estimate; *F* (3, 221); DW = Durbin-Watson.

Appendix 12.22

*Regression analysis summary for productive vocabulary knowledge and the questionnaire items predicting each component score (*n = *225)*

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-1.22	0.44		-2.79	0.01				
PF1Size	0.00	0.00	0.27	2.12	0.04	0.46	0.17	0.14	0.34
Deri	-0.02	0.03	-0.08	-0.77	0.44	0.35	-0.06	-0.05	-0.08
Anto	0.07	0.04	0.21	1.77	0.08	0.48	0.14	0.11	0.27
Collo	-0.01	0.03	-0.03	-0.34	0.73	0.34	-0.03	-0.02	-0.03
Q2_1	0.00	0.07	0.00	0.03	0.97	0.06	0.00	0.00	0.00
Q2_2	0.20	0.06	0.25	3.11	0.00	0.27	0.24	0.20	0.18
Q2_3	0.01	0.05	0.01	0.12	0.90	0.12	0.01	0.01	0.00
Q2_9	0.01	0.05	0.02	0.29	0.77	-0.02	0.02	0.02	0.00
Q2_10	0.04	0.05	0.06	0.89	0.37	0.10	0.07	0.06	0.02
Q2_11	0.10	0.07	0.13	1.40	0.16	0.13	0.11	0.09	0.04
Q2_12	0.13	0.10	0.17	1.24	0.22	-0.04	0.10	0.08	-0.02
Q2_13	-0.16	0.10	-0.22	-1.54	0.13	-0.12	-0.12	-0.10	0.07
Q2_14	-0.13	0.06	-0.17	-2.09	0.04	-0.15	-0.17	-0.14	0.07
Q2_15	0.03	0.07	0.04	0.48	0.63	-0.02	0.04	0.03	0.00
Q2_16	-0.14	0.06	-0.19	-2.47	0.01	-0.23	-0.20	-0.16	0.12

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.14	0.46		0.32	0.75				
PF2Size	0.00	0.00	-0.06	-0.39	0.69	0.19	-0.03	-0.03	-0.07
Deri	0.01	0.03	0.04	0.31	0.75	0.18	0.03	0.02	0.05
Anto	0.04	0.04	0.12	0.93	0.35	0.22	0.08	0.07	0.18
Collo	0.06	0.03	0.19	1.80	0.07	0.24	0.14	0.13	0.31
Q2_1	-0.15	0.07	-0.18	-2.17	0.03	-0.13	-0.17	-0.16	0.15
Q2_2	0.07	0.07	0.09	0.99	0.33	0.04	0.08	0.07	0.03
Q2_3	-0.01	0.06	-0.01	-0.16	0.88	0.04	-0.01	-0.01	0.00
Q2_9	0.03	0.05	0.06	0.67	0.50	-0.01	0.05	0.05	0.00
Q2_10	0.00	0.05	-0.01	-0.06	0.95	-0.01	-0.01	0.00	0.00
Q2_11	-0.06	0.07	-0.09	-0.84	0.40	-0.07	-0.07	-0.06	0.04
Q2_12	0.17	0.11	0.26	1.62	0.11	-0.06	0.13	0.12	-0.10
Q2_13	-0.12	0.11	-0.19	-1.14	0.26	-0.13	-0.09	-0.08	0.16
Q2_14	-0.05	0.06	-0.08	-0.85	0.40	-0.19	-0.07	-0.06	0.10
Q2_15	-0.09	0.07	-0.13	-1.33	0.18	-0.17	-0.11	-0.10	0.15
Q2_16	-0.02	0.06	-0.02	-0.26	0.80	-0.12	-0.02	-0.02	0.02
(Constant)	-0.21	0.50		-0.42	0.67				
PF3Size	0.00	0.00	0.02	0.16	0.87	0.23	0.01	0.01	0.03
Deri	-0.03	0.03	-0.10	-0.82	0.41	0.16	-0.07	-0.06	-0.08
Anto	0.09	0.04	0.28	2.13	0.03	0.29	0.17	0.16	0.45
Collo	-0.02	0.04	-0.04	-0.42	0.67	0.16	-0.03	-0.03	-0.04
Q2_1	-0.08	0.08	-0.09	-1.08	0.28	-0.04	-0.09	-0.08	0.02
Q2_2	0.13	0.07	0.16	1.73	0.09	0.15	0.14	0.13	0.13
Q2_3	0.01	0.06	0.02	0.23	0.82	0.07	0.02	0.02	0.01
Q2_9	-0.05	0.05	-0.07	-0.85	0.40	-0.04	-0.07	-0.06	0.01
Q2_10	0.09	0.05	0.14	1.72	0.09	0.12	0.14	0.13	0.09
Q2_11	0.14	0.08	0.17	1.72	0.09	0.09	0.14	0.13	0.09
Q2_12	0.05	0.12	0.06	0.40	0.69	-0.10	0.03	0.03	-0.03
Q2_13	-0.17	0.12	-0.24	-1.48	0.14	-0.15	-0.12	-0.11	0.19
Q2_14	0.06	0.07	0.08	0.89	0.37	-0.03	0.07	0.07	-0.01
Q2_15	-0.13	0.08	-0.17	-1.68	0.10	-0.11	-0.13	-0.12	0.10
Q2_16	-0.06	0.07	-0.07	-0.85	0.40	-0.15	-0.07	-0.06	0.06
(Constant)	0.01	0.50		0.01	0.99				
PF4Size	0.00	0.00	0.10	0.66	0.51	0.12	0.05	0.05	0.20
Deri	0.02	0.03	0.07	0.54	0.59	0.13	0.04	0.04	0.14
Anto	-0.03	0.04	-0.10	-0.68	0.50	0.09	-0.06	-0.05	-0.14
Collo	0.00	0.04	0.01	0.10	0.92	0.09	0.01	0.01	0.02
Q2_1	-0.06	0.08	-0.07	-0.80	0.42	-0.05	-0.06	-0.06	0.05
Q2_2	0.04	0.07	0.06	0.58	0.56	0.00	0.05	0.05	0.00
Q2_3	-0.06	0.06	-0.08	-0.92	0.36	-0.07	-0.07	-0.07	0.09
Q2_9	0.03	0.05	0.05	0.59	0.56	0.03	0.05	0.05	0.02
Q2_10	-0.01	0.05	-0.02	-0.18	0.85	0.01	-0.01	-0.01	0.00
Q2_11	0.14	0.08	0.19	1.77	0.08	0.11	0.14	0.14	0.33
Q2_12	0.04	0.12	0.06	0.35	0.73	-0.02	0.03	0.03	-0.02
Q2_13	-0.11	0.12	-0.16	-0.93	0.36	-0.06	-0.08	-0.07	0.14
Q2_14	-0.01	0.07	-0.01	-0.11	0.91	-0.03	-0.01	-0.01	0.01
Q2_15	-0.08	0.08	-0.11	-1.03	0.31	-0.08	-0.08	-0.08	0.14
Q2_16	-0.02	0.07	-0.02	-0.23	0.82	-0.07	-0.02	-0.02	0.02
(Constant)	0.12	0.59		0.20	0.84				
PF5Size	0.00	0.00	0.08	0.53	0.60	0.03	0.04	0.04	0.11
Deri	-0.01	0.04	-0.05	-0.39	0.70	0.00	-0.03	-0.03	0.01
Anto	-0.02	0.05	-0.04	-0.30	0.77	0.00	-0.02	-0.02	0.00
Collo	0.01	0.04	0.02	0.18	0.86	0.00	0.01	0.01	0.00
Q2_1	-0.09	0.09	-0.09	-0.98	0.33	-0.07	-0.08	-0.08	0.32
Q2_2	0.01	0.09	0.01	0.08	0.94	0.00	0.01	0.01	0.00
Q2_3	-0.01	0.07	-0.01	-0.12	0.91	0.00	-0.01	-0.01	0.00
Q2_9	-0.04	0.06	-0.05	-0.58	0.56	-0.03	-0.05	-0.05	0.08
Q2_10	0.04	0.06	0.05	0.57	0.57	0.03	0.05	0.05	0.07
Q2_11	0.03	0.09	0.04	0.35	0.72	0.05	0.03	0.03	0.10
Q2_12	-0.01	0.14	-0.01	-0.04	0.96	0.04	0.00	0.00	-0.02
Q2_13	0.03	0.14	0.04	0.21	0.83	0.05	0.02	0.02	0.09
Q2_14	0.06	0.08	0.07	0.71	0.48	0.05	0.06	0.06	0.20
Q2_15	-0.05	0.09	-0.07	-0.61	0.54	-0.01	-0.05	-0.05	0.02
Q2_16	0.01	0.08	0.01	0.09	0.93	0.01	0.01	0.01	0.00

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-0.96	0.51		-1.88	0.06				
PF6Size	0.00	0.00	0.25	1.63	0.10	0.25	0.13	0.13	0.66
Deri	-0.02	0.03	-0.09	-0.76	0.45	0.15	-0.06	-0.06	-0.15
Anto	0.01	0.05	0.02	0.11	0.91	0.20	0.01	0.01	0.03
Collo	0.03	0.04	0.09	0.85	0.40	0.19	0.07	0.07	0.19
Q2_1	0.01	0.08	0.01	0.09	0.93	0.05	0.01	0.01	0.00
Q2_2	-0.01	0.07	-0.01	-0.07	0.94	0.01	-0.01	-0.01	0.00
Q2_3	0.00	0.06	0.00	0.05	0.96	0.00	0.00	0.00	0.00
Q2_9	-0.02	0.06	-0.03	-0.30	0.76	-0.03	-0.02	-0.02	0.01
Q2_10	0.02	0.05	0.03	0.37	0.71	0.04	0.03	0.03	0.01
Q2_11	0.00	0.08	0.00	0.00	1.00	0.03	0.00	0.00	0.00
Q2_12	0.13	0.12	0.19	1.12	0.27	0.03	0.09	0.09	0.05
Q2_13	-0.10	0.12	-0.14	-0.83	0.41	-0.04	-0.07	-0.06	0.06
Q2_14	0.07	0.07	0.10	1.04	0.30	0.02	0.08	0.08	0.02
Q2_15	-0.11	0.08	-0.14	-1.39	0.17	-0.08	-0.11	-0.11	0.12
Q2_16	0.01	0.07	0.02	0.18	0.86	-0.05	0.01	0.01	-0.01
(Constant)	-1.27	0.52		-2.45	0.02				
PA1Size	0.00	0.00	0.45	3.10	0.00	0.34	0.24	0.23	0.77
Deri	-0.05	0.03	-0.19	-1.68	0.10	0.17	-0.14	-0.12	-0.17
Anto	-0.01	0.05	-0.04	-0.30	0.76	0.24	-0.02	-0.02	-0.05
Collo	0.03	0.04	0.09	0.92	0.36	0.25	0.07	0.07	0.12
Q2_1	0.01	0.08	0.01	0.15	0.88	0.05	0.01	0.01	0.00
Q2_2	-0.02	0.08	-0.02	-0.20	0.84	0.04	-0.02	-0.01	0.00
Q2_3	0.09	0.06	0.12	1.43	0.15	0.12	0.12	0.10	0.07
Q2_9	-0.05	0.06	-0.07	-0.92	0.36	-0.10	-0.07	-0.07	0.04
Q2_10	0.02	0.05	0.04	0.45	0.66	0.03	0.04	0.03	0.01
Q2_11	-0.04	0.08	-0.05	-0.50	0.62	0.02	-0.04	-0.04	-0.01
Q2_12	0.24	0.12	0.31	1.96	0.05	0.07	0.16	0.14	0.11
Q2_13	-0.10	0.12	-0.13	-0.81	0.42	-0.01	-0.07	-0.06	0.01
Q2_14	0.03	0.07	0.03	0.36	0.72	-0.04	0.03	0.03	-0.01
Q2_15	-0.08	0.08	-0.10	-1.01	0.31	-0.09	-0.08	-0.07	0.04
Q2_16	-0.08	0.07	-0.09	-1.10	0.27	-0.14	-0.09	-0.08	0.06
(Constant)	-1.27	0.49		-2.60	0.01				
PA2Size	0.00	0.00	0.33	2.35	0.02	0.31	0.19	0.17	0.44
Deri	0.00	0.03	-0.01	-0.12	0.91	0.24	-0.01	-0.01	-0.01
Anto	0.00	0.04	0.00	0.00	1.00	0.28	0.00	0.00	0.00
Collo	-0.02	0.04	-0.07	-0.69	0.49	0.18	-0.06	-0.05	-0.05
Q2_1	-0.08	0.08	-0.08	-1.07	0.28	0.03	-0.09	-0.08	-0.01
Q2_2	0.23	0.07	0.29	3.31	0.00	0.31	0.26	0.24	0.39
Q2_3	0.04	0.06	0.05	0.66	0.51	0.17	0.05	0.05	0.04
Q2_9	0.01	0.05	0.01	0.19	0.85	0.03	0.02	0.01	0.00
Q2_10	0.03	0.05	0.05	0.66	0.51	0.09	0.05	0.05	0.02
Q2_11	-0.02	0.08	-0.03	-0.30	0.76	0.14	-0.02	-0.02	-0.02
Q2_12	0.22	0.11	0.30	1.98	0.05	0.12	0.16	0.14	0.15
Q2_13	-0.20	0.11	-0.28	-1.79	0.08	0.04	-0.14	-0.13	-0.05
Q2_14	0.13	0.07	0.17	1.86	0.06	0.15	0.15	0.13	0.11
Q2_15	-0.04	0.07	-0.06	-0.58	0.56	0.08	-0.05	-0.04	-0.02
Q2_16	-0.05	0.06	-0.06	-0.75	0.46	-0.04	-0.06	-0.05	0.01
(Constant)	-1.07	0.52		-2.07	0.04				
PA3Size	0.00	0.00	0.21	1.52	0.13	0.35	0.12	0.11	0.33
Deri	0.00	0.03	-0.02	-0.14	0.89	0.27	-0.01	-0.01	-0.02
Anto	0.07	0.05	0.20	1.57	0.12	0.36	0.13	0.11	0.32
Collo	0.01	0.04	0.02	0.23	0.81	0.26	0.02	0.02	0.03
Q2_1	0.00	0.08	0.00	0.01	0.99	0.01	0.00	0.00	0.00
Q2_2	0.03	0.07	0.04	0.45	0.65	0.06	0.04	0.03	0.01
Q2_3	0.08	0.06	0.10	1.26	0.21	0.11	0.10	0.09	0.05
Q2_9	0.14	0.06	0.21	2.60	0.01	0.14	0.21	0.18	0.13
Q2_10	-0.04	0.05	-0.06	-0.77	0.44	0.03	-0.06	-0.06	-0.01
Q2_11	-0.06	0.08	-0.07	-0.68	0.50	-0.03	-0.06	-0.05	0.01
Q2_12	-0.02	0.12	-0.02	-0.16	0.87	-0.10	-0.01	-0.01	0.01
Q2_13	0.03	0.12	0.03	0.22	0.82	-0.12	0.02	0.02	-0.02
Q2_14	0.01	0.07	0.01	0.13	0.90	-0.09	0.01	0.01	0.00
Q2_15	-0.15	0.08	-0.19	-1.94	0.05	-0.16	-0.16	-0.14	0.13
Q2_16	-0.03	0.07	-0.04	-0.47	0.64	-0.18	-0.04	-0.03	0.03

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-0.96	0.49		-1.95	0.05				
PA4Size	0.00	0.00	0.44	3.22	0.00	0.40	0.25	0.22	0.61
Deri	-0.03	0.03	-0.10	-0.95	0.34	0.23	-0.08	-0.06	-0.08
Anto	0.05	0.04	0.15	1.19	0.23	0.36	0.10	0.08	0.18
Collo	-0.04	0.04	-0.11	-1.16	0.25	0.21	-0.09	-0.08	-0.08
Q2_1	0.02	0.08	0.02	0.29	0.77	0.04	0.02	0.02	0.00
Q2_2	0.16	0.07	0.18	2.17	0.03	0.19	0.17	0.15	0.12
Q2_3	0.11	0.06	0.14	1.78	0.08	0.15	0.14	0.12	0.07
Q2_9	-0.04	0.05	-0.06	-0.74	0.46	-0.11	-0.06	-0.05	0.02
Q2_10	-0.03	0.05	-0.04	-0.59	0.55	-0.01	-0.05	-0.04	0.00
Q2_11	-0.08	0.08	-0.10	-1.02	0.31	-0.11	-0.08	-0.07	0.04
Q2_12	-0.12	0.11	-0.15	-1.01	0.32	-0.18	-0.08	-0.07	0.09
Q2_13	-0.01	0.11	-0.01	-0.08	0.94	-0.20	-0.01	-0.01	0.01
Q2_14	0.10	0.07	0.13	1.43	0.15	-0.04	0.12	0.10	-0.02
Q2_15	-0.12	0.08	-0.15	-1.60	0.11	-0.10	-0.13	-0.11	0.05
Q2_16	0.05	0.07	0.06	0.78	0.43	-0.09	0.06	0.05	-0.02
(Constant)	0.45	0.52		0.87	0.38				
PA5Size	0.00	0.00	0.19	1.21	0.23	0.10	0.10	0.09	0.22
Deri	-0.01	0.03	-0.03	-0.24	0.81	0.03	-0.02	-0.02	-0.01
Anto	-0.01	0.05	-0.04	-0.28	0.78	0.03	-0.02	-0.02	-0.01
Collo	0.00	0.04	0.01	0.07	0.94	0.00	0.01	0.01	0.00
Q2_1	-0.13	0.08	-0.13	-1.59	0.11	-0.13	-0.13	-0.12	0.21
Q2_2	-0.05	0.08	-0.06	-0.61	0.54	-0.10	-0.05	-0.05	0.08
Q2_3	-0.01	0.06	-0.01	-0.16	0.87	-0.05	-0.01	-0.01	0.01
Q2_9	-0.05	0.06	-0.08	-0.89	0.38	-0.09	-0.07	-0.07	0.08
Q2_10	0.01	0.05	0.02	0.25	0.80	-0.03	0.02	0.02	-0.01
Q2_11	-0.08	0.08	-0.11	-0.98	0.33	-0.05	-0.08	-0.08	0.06
Q2_12	0.08	0.12	0.11	0.64	0.53	0.05	0.05	0.05	0.06
Q2_13	0.05	0.12	0.06	0.38	0.71	0.04	0.03	0.03	0.03
Q2_14	0.09	0.07	0.13	1.29	0.20	0.02	0.10	0.10	0.04
Q2_15	-0.13	0.08	-0.18	-1.70	0.09	-0.11	-0.14	-0.13	0.25
Q2_16	0.02	0.07	0.03	0.33	0.74	-0.02	0.03	0.03	-0.01
(Constant)	-0.74	0.56		-1.33	0.19				
PSC1Size	0.00	0.00	0.16	1.06	0.29	0.14	0.09	0.08	0.24
Deri	-0.02	0.03	-0.06	-0.52	0.61	0.07	-0.04	-0.04	-0.05
Anto	0.00	0.05	0.00	-0.03	0.97	0.11	0.00	0.00	-0.01
Collo	0.03	0.04	0.08	0.70	0.48	0.11	0.06	0.05	0.08
Q2_1	0.04	0.09	0.04	0.52	0.60	0.05	0.04	0.04	0.02
Q2_2	0.04	0.08	0.05	0.54	0.59	0.10	0.04	0.04	0.05
Q2_3	0.03	0.07	0.04	0.47	0.64	0.05	0.04	0.04	0.02
Q2_9	-0.07	0.06	-0.10	-1.18	0.24	-0.12	-0.10	-0.09	0.12
Q2_10	-0.09	0.06	-0.13	-1.58	0.12	-0.13	-0.13	-0.12	0.17
Q2_11	-0.10	0.09	-0.12	-1.15	0.25	-0.01	-0.09	-0.09	0.01
Q2_12	-0.03	0.13	-0.04	-0.24	0.81	0.04	-0.02	-0.02	-0.01
Q2_13	0.08	0.13	0.11	0.63	0.53	0.05	0.05	0.05	0.05
Q2_14	0.12	0.08	0.16	1.59	0.11	0.12	0.13	0.12	0.19
Q2_15	0.06	0.08	0.07	0.70	0.48	0.10	0.06	0.05	0.07
Q2_16	-0.07	0.07	-0.08	-0.91	0.36	-0.04	-0.07	-0.07	0.04
(Constant)	-0.29	0.47		-0.61	0.54				
PSC2Size	0.00	0.00	0.29	1.90	0.06	0.25	0.15	0.14	0.56
Deri	0.02	0.03	0.07	0.60	0.55	0.21	0.05	0.05	0.12
Anto	-0.02	0.04	-0.07	-0.54	0.59	0.19	-0.04	-0.04	-0.11
Collo	-0.02	0.03	-0.06	-0.59	0.56	0.09	-0.05	-0.04	-0.05
Q2_1	-0.09	0.07	-0.10	-1.27	0.20	-0.08	-0.10	-0.10	0.06
Q2_2	0.08	0.07	0.10	1.11	0.27	0.05	0.09	0.08	0.04
Q2_3	-0.10	0.06	-0.15	-1.69	0.09	-0.12	-0.14	-0.13	0.14
Q2_9	-0.03	0.05	-0.05	-0.54	0.59	-0.05	-0.04	-0.04	0.02
Q2_10	-0.01	0.05	-0.02	-0.24	0.81	-0.02	-0.02	-0.02	0.00
Q2_11	0.07	0.07	0.09	0.87	0.38	0.13	0.07	0.07	0.09
Q2_12	-0.09	0.11	-0.13	-0.81	0.42	0.02	-0.07	-0.06	-0.02
Q2_13	0.06	0.11	0.09	0.54	0.59	0.04	0.04	0.04	0.03
Q2_14	0.09	0.07	0.13	1.33	0.18	0.12	0.11	0.10	0.12
Q2_15	-0.02	0.07	-0.02	-0.23	0.82	0.06	-0.02	-0.02	-0.01
Q2_16	-0.02	0.06	-0.02	-0.24	0.81	-0.04	-0.02	-0.02	0.01

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	-0.24	0.48		-0.51	0.61				
PSC3Size	0.00	0.00	-0.06	-0.39	0.70	0.14	-0.03	-0.03	-0.10
Deri	0.01	0.03	0.04	0.33	0.74	0.15	0.03	0.03	0.07
Anto	0.06	0.04	0.19	1.41	0.16	0.19	0.11	0.11	0.40
Collo	-0.01	0.03	-0.04	-0.34	0.73	0.10	-0.03	-0.03	-0.04
Q2_1	-0.01	0.07	-0.01	-0.11	0.91	0.05	-0.01	-0.01	0.00
Q2_2	0.11	0.07	0.15	1.57	0.12	0.11	0.13	0.12	0.18
Q2_3	-0.09	0.06	-0.13	-1.51	0.13	-0.07	-0.12	-0.12	0.10
Q2_9	-0.05	0.05	-0.08	-0.97	0.34	-0.05	-0.08	-0.07	0.05
Q2_10	0.11	0.05	0.18	2.13	0.03	0.15	0.17	0.16	0.29
Q2_11	-0.02	0.08	-0.02	-0.23	0.82	0.01	-0.02	-0.02	0.00
Q2_12	0.04	0.11	0.06	0.33	0.74	-0.04	0.03	0.03	-0.02
Q2_13	-0.03	0.11	-0.04	-0.23	0.82	-0.05	-0.02	-0.02	0.02
Q2_14	0.00	0.07	-0.01	-0.07	0.94	-0.04	-0.01	-0.01	0.00
Q2_15	-0.03	0.07	-0.05	-0.46	0.65	-0.04	-0.04	-0.04	0.02
Q2_16	-0.03	0.06	-0.05	-0.50	0.62	-0.07	-0.04	-0.04	0.04
(Constant)	-1.53	0.52		-2.93	0.00				
PSC4Size	0.00	0.00	0.31	2.12	0.04	0.32	0.17	0.16	0.64
Deri	0.03	0.03	0.12	1.03	0.30	0.28	0.08	0.08	0.22
Anto	-0.04	0.05	-0.11	-0.80	0.43	0.23	-0.06	-0.06	-0.15
Collo	0.00	0.04	0.01	0.13	0.90	0.22	0.01	0.01	0.02
Q2_1	0.11	0.08	0.11	1.32	0.19	0.11	0.11	0.10	0.08
Q2_2	0.03	0.08	0.04	0.44	0.66	0.12	0.04	0.03	0.03
Q2_3	0.08	0.06	0.11	1.24	0.22	0.10	0.10	0.09	0.07
Q2_9	-0.02	0.06	-0.03	-0.38	0.70	-0.06	-0.03	-0.03	0.01
Q2_10	-0.07	0.05	-0.11	-1.31	0.19	-0.04	-0.11	-0.10	0.03
Q2_11	-0.06	0.08	-0.07	-0.68	0.50	0.03	-0.06	-0.05	-0.01
Q2_12	-0.02	0.12	-0.03	-0.19	0.85	0.03	-0.02	-0.01	-0.01
Q2_13	0.08	0.12	0.11	0.68	0.50	0.02	0.06	0.05	0.02
Q2_14	0.04	0.07	0.05	0.54	0.59	0.04	0.04	0.04	0.01
Q2_15	0.01	0.08	0.02	0.17	0.86	0.05	0.01	0.01	0.01
Q2_16	-0.06	0.07	-0.07	-0.84	0.40	-0.10	-0.07	-0.06	0.05
(Constant)	-0.79	0.47		-1.70	0.09				
PSC5Size	0.00	0.00	0.28	2.09	0.04	0.49	0.17	0.14	0.47
Deri	0.01	0.03	0.03	0.24	0.81	0.39	0.02	0.02	0.03
Anto	0.07	0.04	0.21	1.76	0.08	0.49	0.14	0.12	0.36
Collo	-0.01	0.03	-0.03	-0.30	0.76	0.31	-0.02	-0.02	-0.03
Q2_1	-0.10	0.07	-0.11	-1.42	0.16	-0.09	-0.11	-0.10	0.03
Q2_2	0.06	0.07	0.08	0.92	0.36	0.10	0.07	0.06	0.03
Q2_3	0.00	0.06	0.01	0.07	0.95	0.03	0.01	0.00	0.00
Q2_9	0.00	0.05	0.00	0.06	0.95	-0.04	0.00	0.00	0.00
Q2_10	-0.03	0.05	-0.05	-0.61	0.54	-0.02	-0.05	-0.04	0.00
Q2_11	0.09	0.07	0.11	1.18	0.24	0.13	0.10	0.08	0.05
Q2_12	0.01	0.11	0.02	0.11	0.92	-0.04	0.01	0.01	0.00
Q2_13	-0.05	0.11	-0.06	-0.42	0.68	-0.08	-0.03	-0.03	0.02
Q2_14	-0.02	0.07	-0.02	-0.27	0.78	-0.06	-0.02	-0.02	0.00
Q2_15	-0.02	0.07	-0.02	-0.24	0.81	-0.02	-0.02	-0.02	0.00
Q2_16	-0.06	0.06	-0.08	-0.95	0.34	-0.17	-0.08	-0.06	0.04
(Constant)	-0.76	0.47		-1.63	0.11				
PLC1Size	0.00	0.00	0.24	1.80	0.07	0.42	0.14	0.12	0.36
Deri	0.00	0.03	0.00	0.00	1.00	0.35	0.00	0.00	0.00
Anto	0.06	0.04	0.17	1.42	0.16	0.44	0.11	0.10	0.27
Collo	-0.03	0.03	-0.08	-0.80	0.43	0.28	-0.06	-0.05	-0.08
Q2_1	-0.06	0.07	-0.06	-0.77	0.44	-0.02	-0.06	-0.05	0.00
Q2_2	0.14	0.07	0.17	1.99	0.05	0.17	0.16	0.14	0.10
Q2_3	-0.01	0.06	-0.01	-0.11	0.92	0.06	-0.01	-0.01	0.00
Q2_9	0.04	0.05	0.07	0.89	0.37	0.03	0.07	0.06	0.01
Q2_10	0.01	0.05	0.02	0.26	0.80	0.07	0.02	0.02	0.00
Q2_11	0.14	0.07	0.18	1.94	0.05	0.14	0.16	0.13	0.09
Q2_12	0.02	0.11	0.03	0.22	0.83	-0.09	0.02	0.01	-0.01
Q2_13	-0.14	0.11	-0.20	-1.32	0.19	-0.13	-0.11	-0.09	0.09
Q2_14	-0.06	0.07	-0.09	-0.96	0.34	-0.08	-0.08	-0.07	0.02
Q2_15	0.04	0.07	0.05	0.53	0.60	0.00	0.04	0.04	0.00
Q2_16	-0.13	0.06	-0.17	-2.05	0.04	-0.22	-0.16	-0.14	0.13

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	1.03	0.50		2.07	0.04				
PLC2Size	0.00	0.00	-0.17	-1.07	0.28	0.00	-0.09	-0.08	0.01
Deri	0.02	0.03	0.10	0.81	0.42	0.04	0.07	0.06	0.06
Anto	0.03	0.04	0.08	0.59	0.56	0.05	0.05	0.05	0.05
Collo	-0.01	0.04	-0.02	-0.16	0.87	-0.01	-0.01	-0.01	0.00
Q2_1	-0.15	0.08	-0.17	-1.99	0.05	-0.15	-0.16	-0.16	0.35
Q2_2	0.08	0.07	0.11	1.18	0.24	0.04	0.10	0.09	0.06
Q2_3	-0.07	0.06	-0.11	-1.20	0.23	-0.07	-0.10	-0.09	0.10
Q2_9	-0.07	0.05	-0.11	-1.23	0.22	-0.13	-0.10	-0.10	0.20
Q2_10	-0.02	0.05	-0.03	-0.32	0.75	-0.10	-0.03	-0.02	0.04
Q2_11	-0.01	0.08	-0.02	-0.17	0.87	-0.02	-0.01	-0.01	0.00
Q2_12	0.09	0.12	0.13	0.76	0.45	-0.02	0.06	0.06	-0.04
Q2_13	-0.09	0.12	-0.14	-0.81	0.42	-0.04	-0.07	-0.06	0.08
Q2_14	-0.03	0.07	-0.04	-0.38	0.70	-0.06	-0.03	-0.03	0.03
Q2_15	0.06	0.08	0.09	0.82	0.41	0.01	0.07	0.06	0.02
Q2_16	-0.06	0.07	-0.08	-0.84	0.40	-0.05	-0.07	-0.07	0.06
(Constant)	-0.47	0.51		-0.93	0.35				
PLC3Size	0.00	0.00	0.07	0.47	0.64	-0.03	0.04	0.04	-0.02
Deri	-0.04	0.03	-0.14	-1.15	0.25	-0.09	-0.09	-0.09	0.10
Anto	-0.01	0.04	-0.02	-0.17	0.86	-0.05	-0.01	-0.01	0.01
Collo	-0.03	0.04	-0.07	-0.69	0.49	-0.08	-0.06	-0.05	0.05
Q2_1	0.10	0.08	0.10	1.23	0.22	0.11	0.10	0.09	0.09
Q2_2	0.03	0.07	0.04	0.40	0.69	0.08	0.03	0.03	0.03
Q2_3	0.03	0.06	0.04	0.49	0.63	0.04	0.04	0.04	0.01
Q2_9	-0.11	0.05	-0.18	-2.11	0.04	-0.10	-0.17	-0.16	0.15
Q2_10	0.02	0.05	0.02	0.28	0.78	0.00	0.02	0.02	0.00
Q2_11	0.09	0.08	0.12	1.17	0.24	0.12	0.09	0.09	0.11
Q2_12	-0.02	0.12	-0.02	-0.15	0.88	0.07	-0.01	-0.01	-0.01
Q2_13	-0.02	0.12	-0.04	-0.21	0.83	0.07	-0.02	-0.02	-0.02
Q2_14	0.16	0.07	0.22	2.26	0.03	0.21	0.18	0.17	0.37
Q2_15	0.03	0.08	0.04	0.44	0.66	0.10	0.04	0.03	0.04
Q2_16	-0.13	0.07	-0.18	-1.97	0.05	-0.07	-0.16	-0.15	0.10
(Constant)	0.50	0.44		1.15	0.25				
PLC4Size	0.00	0.00	-0.14	-0.91	0.36	-0.14	-0.07	-0.07	0.25
Deri	0.00	0.03	-0.02	-0.16	0.87	-0.13	-0.01	-0.01	0.03
Anto	0.04	0.04	0.16	1.12	0.26	-0.10	0.09	0.09	-0.18
Collo	-0.04	0.03	-0.13	-1.18	0.24	-0.19	-0.10	-0.09	0.30
Q2_1	-0.03	0.07	-0.04	-0.51	0.61	-0.09	-0.04	-0.04	0.05
Q2_2	-0.11	0.06	-0.16	-1.67	0.10	-0.11	-0.13	-0.13	0.22
Q2_3	0.03	0.05	0.05	0.52	0.60	-0.02	0.04	0.04	-0.01
Q2_9	-0.04	0.05	-0.06	-0.74	0.46	-0.02	-0.06	-0.06	0.01
Q2_10	0.02	0.05	0.03	0.39	0.70	-0.02	0.03	0.03	-0.01
Q2_11	-0.05	0.07	-0.07	-0.67	0.51	-0.02	-0.05	-0.05	0.01
Q2_12	-0.05	0.10	-0.09	-0.51	0.61	0.05	-0.04	-0.04	-0.05
Q2_13	0.10	0.10	0.17	0.98	0.33	0.10	0.08	0.08	0.21
Q2_14	0.02	0.06	0.04	0.40	0.69	0.08	0.03	0.03	0.04
Q2_15	0.07	0.07	0.11	1.04	0.30	0.09	0.08	0.08	0.13
Q2_16	0.00	0.06	0.01	0.07	0.95	0.06	0.01	0.01	0.00
(Constant)	0.29	0.55		0.52	0.60				
PLC5Size	0.00	0.00	-0.20	-1.30	0.20	-0.21	-0.10	-0.10	0.44
Deri	0.01	0.03	0.05	0.41	0.69	-0.12	0.03	0.03	-0.06
Anto	-0.04	0.05	-0.10	-0.76	0.45	-0.19	-0.06	-0.06	0.21
Collo	0.03	0.04	0.07	0.65	0.52	-0.09	0.05	0.05	-0.07
Q2_1	-0.01	0.08	-0.01	-0.14	0.89	0.00	-0.01	-0.01	0.00
Q2_2	0.04	0.08	0.05	0.54	0.59	0.02	0.04	0.04	0.01
Q2_3	-0.08	0.07	-0.10	-1.15	0.25	-0.07	-0.09	-0.09	0.07
Q2_9	-0.01	0.06	-0.02	-0.23	0.82	0.02	-0.02	-0.02	0.00
Q2_10	0.03	0.06	0.04	0.49	0.63	0.03	0.04	0.04	0.01
Q2_11	0.09	0.09	0.12	1.09	0.28	0.06	0.09	0.08	0.08
Q2_12	-0.07	0.13	-0.10	-0.58	0.56	0.05	-0.05	-0.04	-0.05
Q2_13	0.05	0.13	0.06	0.36	0.72	0.07	0.03	0.03	0.05
Q2_14	-0.04	0.08	-0.05	-0.50	0.62	0.02	-0.04	-0.04	-0.01
Q2_15	-0.05	0.08	-0.06	-0.60	0.55	0.03	-0.05	-0.05	-0.02
Q2_16	0.14	0.07	0.18	1.94	0.05	0.19	0.16	0.15	0.34

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.28	0.51		0.56	0.58				
PLC6Size	0.00	0.00	-0.23	-1.57	0.12	-0.28	-0.13	-0.12	0.36
Deri	0.06	0.03	0.22	1.88	0.06	-0.15	0.15	0.14	-0.18
Anto	-0.06	0.05	-0.16	-1.24	0.22	-0.29	-0.10	-0.09	0.26
Collo	-0.04	0.04	-0.11	-1.09	0.28	-0.22	-0.09	-0.08	0.14
Q2_1	0.22	0.08	0.22	2.73	0.01	0.17	0.22	0.20	0.20
Q2_2	-0.05	0.07	-0.07	-0.72	0.47	-0.07	-0.06	-0.05	0.02
Q2_3	-0.01	0.06	-0.02	-0.18	0.85	-0.08	-0.01	-0.01	0.01
Q2_9	-0.03	0.06	-0.04	-0.46	0.65	-0.03	-0.04	-0.03	0.01
Q2_10	-0.07	0.05	-0.11	-1.34	0.18	-0.09	-0.11	-0.10	0.06
Q2_11	0.03	0.08	0.04	0.41	0.68	-0.03	0.03	0.03	-0.01
Q2_12	0.03	0.12	0.04	0.26	0.80	0.06	0.02	0.02	0.01
Q2_13	-0.04	0.12	-0.06	-0.36	0.72	0.04	-0.03	-0.03	-0.01
Q2_14	-0.04	0.07	-0.06	-0.59	0.56	0.01	-0.05	-0.04	0.00
Q2_15	-0.02	0.08	-0.03	-0.26	0.79	0.00	-0.02	-0.02	0.00
Q2_16	0.12	0.07	0.15	1.75	0.08	0.17	0.14	0.13	0.14
(Constant)	-0.41	0.52		-0.77	0.44				
PLC7Size	0.00	0.00	0.21	1.41	0.16	-0.09	0.11	0.11	-0.15
Deri	0.03	0.03	0.12	1.02	0.31	-0.03	0.08	0.08	-0.03
Anto	-0.14	0.05	-0.42	-3.06	0.00	-0.21	-0.24	-0.23	0.74
Collo	-0.04	0.04	-0.10	-0.96	0.34	-0.10	-0.08	-0.07	0.09
Q2_1	0.06	0.08	0.06	0.77	0.44	0.09	0.06	0.06	0.05
Q2_2	0.03	0.08	0.04	0.42	0.68	0.02	0.03	0.03	0.01
Q2_3	0.00	0.07	0.00	-0.04	0.97	0.02	0.00	0.00	0.00
Q2_9	0.05	0.06	0.07	0.82	0.41	0.09	0.07	0.06	0.06
Q2_10	0.08	0.05	0.12	1.42	0.16	0.17	0.11	0.11	0.17
Q2_11	0.04	0.08	0.06	0.54	0.59	0.04	0.04	0.04	0.02
Q2_12	-0.01	0.12	-0.02	-0.11	0.91	0.07	-0.01	-0.01	-0.01
Q2_13	0.01	0.12	0.02	0.12	0.90	0.08	0.01	0.01	0.01
Q2_14	-0.08	0.07	-0.10	-1.04	0.30	0.02	-0.08	-0.08	-0.02
Q2_15	0.02	0.08	0.02	0.24	0.81	0.05	0.02	0.02	0.01
Q2_16	0.06	0.07	0.07	0.82	0.42	0.11	0.07	0.06	0.07
(Constant)	-0.32	0.51		-0.62	0.53				
PLC8Size	0.00	0.00	-0.16	-1.09	0.28	-0.10	-0.09	-0.08	0.12
Deri	0.03	0.03	0.12	0.97	0.33	-0.02	0.08	0.07	-0.02
Anto	-0.05	0.04	-0.14	-1.07	0.29	-0.13	-0.09	-0.08	0.14
Collo	0.09	0.04	0.26	2.48	0.01	0.05	0.20	0.19	0.11
Q2_1	-0.03	0.08	-0.03	-0.37	0.71	0.01	-0.03	-0.03	0.00
Q2_2	-0.04	0.07	-0.05	-0.49	0.62	-0.03	-0.04	-0.04	0.01
Q2_3	0.00	0.06	-0.01	-0.07	0.94	0.00	-0.01	-0.01	0.00
Q2_9	0.01	0.05	0.01	0.16	0.87	0.02	0.01	0.01	0.00
Q2_10	-0.04	0.05	-0.06	-0.77	0.44	-0.06	-0.06	-0.06	0.03
Q2_11	-0.15	0.08	-0.19	-1.81	0.07	-0.02	-0.15	-0.14	0.03
Q2_12	0.22	0.12	0.31	1.90	0.06	0.18	0.15	0.14	0.41
Q2_13	-0.03	0.12	-0.04	-0.27	0.79	0.13	-0.02	-0.02	-0.04
Q2_14	0.11	0.07	0.15	1.50	0.14	0.09	0.12	0.11	0.10
Q2_15	-0.12	0.08	-0.15	-1.50	0.14	-0.03	-0.12	-0.11	0.03
Q2_16	0.07	0.07	0.09	0.99	0.32	0.11	0.08	0.07	0.07
(Constant)	0.06	0.55		0.10	0.92				
PLC9Size	0.00	0.00	-0.33	-2.17	0.03	-0.15	-0.17	-0.16	0.38
Deri	0.00	0.03	-0.01	-0.10	0.92	-0.09	-0.01	-0.01	0.01
Anto	0.03	0.05	0.10	0.71	0.48	-0.10	0.06	0.05	-0.07
Collo	0.08	0.04	0.22	2.05	0.04	0.01	0.16	0.16	0.02
Q2_1	-0.04	0.09	-0.04	-0.52	0.60	-0.01	-0.04	-0.04	0.00
Q2_2	-0.13	0.08	-0.15	-1.67	0.10	-0.09	-0.13	-0.13	0.11
Q2_3	0.01	0.07	0.01	0.15	0.88	0.00	0.01	0.01	0.00
Q2_9	-0.02	0.06	-0.03	-0.37	0.71	0.04	-0.03	-0.03	-0.01
Q2_10	0.09	0.06	0.14	1.63	0.10	0.11	0.13	0.12	0.12
Q2_11	-0.13	0.09	-0.15	-1.47	0.14	-0.10	-0.12	-0.11	0.11
Q2_12	-0.07	0.13	-0.09	-0.55	0.58	-0.08	-0.04	-0.04	0.05
Q2_13	0.00	0.13	0.00	-0.01	0.99	-0.04	0.00	0.00	0.00
Q2_14	0.16	0.08	0.20	2.06	0.04	0.12	0.17	0.16	0.18
Q2_15	0.08	0.08	0.10	0.95	0.34	0.09	0.08	0.07	0.06
Q2_16	0.05	0.07	0.06	0.62	0.53	0.09	0.05	0.05	0.04

(Appendix 12.22 continues)

(Appendix 12.22 continued)

Variable	B	*SEB*	*β*	*t*	*p*	*r*	Parti	Part	*d*
(Constant)	0.35	0.55		0.63	0.53				
PLC10Size	0.00	0.00	-0.08	-0.56	0.58	-0.10	-0.05	-0.04	0.07
Deri	0.07	0.03	0.25	2.04	0.04	0.01	0.16	0.16	0.03
Anto	-0.04	0.05	-0.12	-0.87	0.39	-0.12	-0.07	-0.07	0.12
Collo	-0.05	0.04	-0.14	-1.33	0.19	-0.12	-0.11	-0.10	0.15
Q2_1	0.07	0.09	0.07	0.86	0.39	0.04	0.07	0.07	0.02
Q2_2	0.00	0.08	0.00	0.03	0.98	-0.04	0.00	0.00	0.00
Q2_3	-0.05	0.07	-0.07	-0.75	0.46	-0.10	-0.06	-0.06	0.06
Q2_9	0.00	0.06	-0.01	-0.06	0.95	0.03	0.00	0.00	0.00
Q2_10	0.03	0.06	0.04	0.47	0.64	0.07	0.04	0.04	0.02
Q2_11	0.07	0.09	0.08	0.78	0.44	-0.02	0.06	0.06	-0.02
Q2_12	-0.37	0.13	-0.48	-2.88	0.00	-0.13	-0.23	-0.22	0.53
Q2_13	0.21	0.13	0.28	1.66	0.10	-0.03	0.13	0.13	-0.07
Q2_14	0.02	0.08	0.03	0.26	0.79	0.03	0.02	0.02	0.01
Q2_15	-0.04	0.08	-0.05	-0.46	0.64	0.00	-0.04	-0.04	0.00
Q2_16	0.11	0.07	0.13	1.45	0.15	0.08	0.12	0.11	0.09
(Constant)	1.07	0.54		1.97	0.05				
PLC11Size	0.00	0.00	-0.35	-2.32	0.02	-0.25	-0.18	-0.18	0.77
Deri	0.03	0.03	0.11	0.89	0.37	-0.14	0.07	0.07	-0.14
Anto	0.00	0.05	0.00	-0.03	0.97	-0.21	0.00	0.00	0.01
Collo	0.04	0.04	0.11	1.00	0.32	-0.11	0.08	0.08	-0.11
Q2_1	-0.11	0.08	-0.11	-1.31	0.19	-0.12	-0.11	-0.10	0.11
Q2_2	-0.03	0.08	-0.04	-0.41	0.69	-0.06	-0.03	-0.03	0.02
Q2_3	-0.02	0.07	-0.02	-0.28	0.78	-0.03	-0.02	-0.02	0.01
Q2_9	-0.06	0.06	-0.09	-1.06	0.29	-0.07	-0.09	-0.08	0.06
Q2_10	-0.01	0.06	-0.02	-0.25	0.81	-0.08	-0.02	-0.02	0.01
Q2_11	-0.09	0.09	-0.11	-1.04	0.30	-0.06	-0.08	-0.08	0.06
Q2_12	0.05	0.13	0.06	0.38	0.70	0.07	0.03	0.03	0.04
Q2_13	0.07	0.13	0.09	0.52	0.61	0.09	0.04	0.04	0.07
Q2_14	-0.01	0.08	-0.01	-0.11	0.91	-0.01	-0.01	-0.01	0.00
Q2_15	-0.01	0.08	-0.02	-0.16	0.87	0.01	-0.01	-0.01	0.00
Q2_16	0.07	0.07	0.09	1.00	0.32	0.13	0.08	0.08	0.10
(Constant)	0.69	0.52		1.33	0.18				
PLC12Size	0.00	0.00	-0.21	-1.38	0.17	-0.23	-0.11	-0.11	0.43
Deri	-0.01	0.03	-0.03	-0.21	0.84	-0.20	-0.02	-0.02	0.04
Anto	0.00	0.05	0.01	0.04	0.97	-0.21	0.00	0.00	-0.01
Collo	0.00	0.04	-0.01	-0.11	0.91	-0.17	-0.01	-0.01	0.02
Q2_1	0.07	0.08	0.07	0.89	0.37	0.07	0.07	0.07	0.05
Q2_2	-0.05	0.08	-0.07	-0.71	0.48	-0.10	-0.06	-0.05	0.06
Q2_3	-0.03	0.06	-0.04	-0.49	0.63	-0.09	-0.04	-0.04	0.03
Q2_9	-0.08	0.06	-0.12	-1.43	0.16	-0.09	-0.12	-0.11	0.10
Q2_10	0.04	0.05	0.06	0.75	0.45	0.00	0.06	0.06	0.00
Q2_11	-0.04	0.08	-0.05	-0.52	0.61	-0.12	-0.04	-0.04	0.06
Q2_12	0.07	0.12	0.10	0.61	0.54	-0.04	0.05	0.05	-0.03
Q2_13	-0.12	0.12	-0.17	-1.03	0.31	-0.07	-0.08	-0.08	0.11
Q2_14	0.08	0.07	0.10	1.05	0.29	0.02	0.09	0.08	0.02
Q2_15	-0.05	0.08	-0.06	-0.58	0.56	-0.06	-0.05	-0.04	0.03
Q2_16	0.08	0.07	0.10	1.14	0.26	0.10	0.09	0.09	0.09

Note. n = 225. r = Zero-order correlation; Parti = Partial correlation; Part = Part correlation; d = relative Pratt index.

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
PF1	0.60	0.36	0.30	0.80	5.74	0.00	1.43
PF2	0.39	0.15	0.07	0.83	1.79	0.04	1.62
PF3	0.43	0.18	0.10	0.92	2.28	0.01	2.05
PF4	0.25	0.06	-0.03	0.90	0.67	0.81	2.12
PF5	0.14	0.02	-0.08	1.07	0.19	1.00	1.97
PF6	0.31	0.09	0.01	0.93	1.06	0.40	1.98
PA1	0.45	0.20	0.12	0.94	2.53	0.00	2.01
PA2	0.48	0.23	0.16	0.88	3.07	0.00	2.36
PA3	0.48	0.23	0.15	0.94	3.00	0.00	1.99
PA4	0.54	0.29	0.22	0.90	4.11	0.00	2.02
PA5	0.28	0.08	-0.01	0.94	0.89	0.58	2.22
PSC1	0.31	0.10	0.01	1.01	1.10	0.36	2.06
PSC2	0.36	0.13	0.04	0.86	1.49	0.11	2.21
PSC3	0.30	0.09	0.00	0.87	1.01	0.45	1.95
PSC4	0.39	0.16	0.07	0.95	1.87	0.03	1.81
PSC5	0.54	0.29	0.23	0.85	4.23	0.00	1.75
PLC1	0.53	0.29	0.21	0.85	4.04	0.00	1.78
PLC2	0.27	0.07	-0.02	0.90	0.81	0.67	1.98
PLC3	0.35	0.12	0.04	0.92	1.44	0.13	1.62
PLC4	0.29	0.08	-0.01	0.80	0.91	0.55	1.75
PLC5	0.31	0.10	0.01	0.99	1.08	0.38	1.98
PLC6	0.43	0.18	0.10	0.93	2.24	0.01	2.12
PLC7	0.34	0.12	0.03	0.95	1.36	0.17	1.93
PLC8	0.37	0.13	0.05	0.92	1.57	0.09	2.00
PLC9	0.36	0.13	0.05	1.01	1.53	0.10	1.86
PLC10	0.34	0.12	0.03	1.00	1.33	0.19	1.93
PLC11	0.34	0.12	0.03	0.99	1.32	0.20	2.00
PLC12	0.33	0.11	0.02	0.95	1.28	0.22	2.14

Note. n = 225. SEE = Standard Error of the Estimate; F (15, 152); DW = Durbin-Watson.

Appendix 12.24

Correlations between size, depth (derivations, antonyms, and collocations), and the questionnaire items

	Size	Deri	Anto	Collo	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Size	--	.74	.80	.62	.05	.07	-.01	-.09	.09	.10	-.07	-.14	-.06	.00	-.19
Deri		--	.67	.60	.05	.12	-.01	.01	.15	.17	-.02	-.07	-.02	.07	-.15
Anto			--	.58	.01	.17	.01	.00	.03	.16	-.07	-.13	-.06	.05	-.17
Collo				--	.13	.18	.13	-.02	.13	.11	-.09	-.16	-.12	.03	-.11
Q1					--	.19	-.02	.07	.19	.02	.08	.00	.12	.07	.04
Q2						--	.42	.05	.10	.25	.20	.21	.18	.30	.16
Q3							--	.11	.13	.08	.10	.10	.04	.13	.03
Q4								--	.29	.23	.15	.21	.24	.14	.09
Q5									--	.10	-.01	.02	.07	.06	.03
Q6										--	.59	.57	.43	.43	.21
Q7											--	.87	.43	.40	.35
Q8												--	.47	.46	.33
Q9													--	.52	.28
Q10														--	.44
Q11															--

Note. n = 168. when r = | .13| to | .17|, p < .05; when r = | .18| or more, p < .01.

Appendix 12.25

Regression analysis summary for productive vocabulary knowledge and the questionnaire items predicting each component score (outliers excluded)

Variable	*R*	R^2	Adjusted R^2	SEE	*F*	*p*	DW
PF2	0.41	0.16	0.08	0.80	1.98	0.02	1.59
PF4	0.24	0.06	-0.03	0.86	0.63	0.84	2.05
PF5	0.17	0.03	-0.07	1.02	0.30	0.99	1.98
PSC4	0.40	0.16	0.08	0.54	1.89	0.03	1.95
PLC2	0.27	0.07	-0.02	0.87	0.78	0.70	1.92
PLC4	0.34	0.12	0.03	0.67	1.34	0.18	1.83
PLC9	0.39	0.15	0.07	0.95	1.77	0.04	1.88

Note. SEE = Standard Error of the Estimate; F (15, 151); DW = Durbin-Watson.

Errata (2006/1/21)

page	line	In the dissertaion	New
63	14	seems	seem
154	5	Appendix 7.2	Appendix 7.6
311	17	because it is basically based on “factually oriented talk”	because the former is a basis of the latter
312	1	covey	convey
324	5	p.	pp.
336	16	word	word”
338	3	Appendix ?	Appendix 7.11
341	7	Appendix ?	Appendix 7.5