Using the think aloud protocol to investigate vocabulary strategies used in two vocabulary tests

by

Ana Maria Fonseca Saravia

A Thesis Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS

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

Approved:  Signature redacted for privacy
           In Charge of Major Work  Signature redacted for privacy
           For the Major Department  Signature redacted for privacy
           For the Graduate College

Iowa State University
   Ames, Iowa
31995
To

Preta

and to the memory of

my grandfather, B.G.
# TABLE OF CONTENTS

INVESTIGATING VOCABULARY ABILITY: AN INTRODUCTION

- Background
- The Current Study
- Overview of the Contents

LITERATURE REVIEW

- Introduction
- Second Language Vocabulary
  - The context of vocabulary use
  - Vocabulary size
  - The organization of the mental lexicon
  - Word characteristics
  - Vocabulary processes
  - Vocabulary strategies
- Second Language Testing
  - Communicative language ability
  - Defining vocabulary ability
  - Process-oriented approaches to language testing
- Verbal Reports
  - Introduction
  - The think-aloud protocol
  - Verbal reports in L2 research

METHODS

- Introduction
- Participants
- Instruments
- Pilot Project
- Procedures
- Data Analysis

ANALYSIS AND RESULTS

- Introduction
  - The participants' approach to the tests
- Word Retrieval and the Vocabulary Abilities Used
  - Taxonomy
  - The abilities used
- Test Format, Individual Differences, and the Abilities Used
- Topic Familiarity and the Abilities Elicited

SUMMARY, CONCLUSIONS, FURTHER CONSIDERATIONS, AND IMPLICATIONS

- Summary
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implications</td>
<td>82</td>
</tr>
<tr>
<td>Methodological</td>
<td>82</td>
</tr>
<tr>
<td>Theoretical</td>
<td>85</td>
</tr>
<tr>
<td>Pedagogical</td>
<td>87</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>89</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>93</td>
</tr>
<tr>
<td>APPENDIX A SENTENCE TEST</td>
<td>94</td>
</tr>
<tr>
<td>APPENDIX B PASSAGES TEST</td>
<td>101</td>
</tr>
<tr>
<td>APPENDIX C TEST RESULTS</td>
<td>111</td>
</tr>
</tbody>
</table>
INVESTIGATING VOCABULARY ABILITY: AN INTRODUCTION

Background

"What's the word for...?" "What does ... mean?" Those are questions any second language learner has asked over and over again. It is a considerable undertaking to step down from the comfortable position of a competent native speaker of a language and take on the role of a second language learner trying to grasp the meaning and/or form of words one can use automatically in the first language.

Fortunately, vocabulary learning and teaching have received deserved attention from researchers over the last decade. Many questions are still unanswered, but any progress made towards a better understanding of how humans manage vocabulary knowledge will be beneficial to future second language learners.

My interest in vocabulary research probably stems from my personal struggle with English words. As an English as a second language learner I have had the frustrating experience of not finding the appropriate words for a given context. While I did not want to sound too formal or bookish, I was not certain about the degree of informality of some words. Preparing to become an English as a foreign language teacher, I became interested in strategies for teaching and learning vocabulary. I also became interested in language testing, especially in the social effects of tests on the lives of language learners.

Those two lines of interest were combined when I decided to investigate the aspects of vocabulary ability involved in taking different types of vocabulary tests. In this study ability (singular) is used in the sense of intellectual
capacity to use words for both recognition and production purposes. *Abilities* (plural) refer to the different aspects of the interactionalist definition of vocabulary ability, including context of use, knowledge of word characteristics, vocabulary processes (e.g., word retrieval), and metacognitive strategies. The word *strategies* refers to the strategic uses learners make of different bodies of knowledge (e.g., abilities) in order to retrieve the tested words.

According to some researchers (e.g., Cohen, 1992; Chapelle, 1994) the typical paper-and-pencil tests usually administered in educational settings will not provide a satisfactory answer to these questions. In looking at product data (e.g., test scores), the researcher or teacher cannot make inferences about strategies and processes. We can make inferences about the aspects of word knowledge our tests were intended to measure and attribute successful performance to the presence of the abilities involved, but the test results will not tell us how the students chose a correct or incorrect answer. If words are tested in the context of a reading passage, for example, we can hypothesize that the students used contextual clues. The extent to which they have used those clues and the strategies used to accomplish that cannot be inferred simply by looking at test scores.

An ideal vocabulary test would enable teachers and other test administrators to make inferences about the vocabulary knowledge of test-takers as related to a specific context of use, the metacognitive strategies required for vocabulary use in that context, the pragmatic and linguistic knowledge of those words, as well as the processes involved in retrieving them. However, a single test method will not elicit such complete performance. Paper and pencil tests may tap into the knowledge of word characteristics, but will not tell us much
about the processes and strategies learners use to retrieve words.

A process-oriented approach to testing has to be taken to gain more insights on the strategies students use to answer test items correctly. Such an approach will also help researchers and test users to better understand the meaning of test scores and enable them to make more informed decisions based on test results. In other words, such an approach will further justify the interpretations and uses (Messick, 1988) of a "given test in a given test administration with given respondents" (Cohen, 1992:27).

Verbal reports done by students while taking a test or doing other language tasks have begun to be used in second language research (Hosenfeld, 1981; Cohen, 1984, 1992) and were used in the study reported here. Students either think aloud while working on the test items or tasks or retrospect about the strategies they used. Think aloud protocols are obtained through instructions such as 'say everything that comes to your mind as you answer the questions, no matter how irrelevant they may seem' (Ericsson and Simon, 1987). The student's 'talk' is then audiotaped. The assumption researchers who use this method are making is that think aloud data will provide information about the cognitive processes used (e.g. processes related to word retrieval) as well as the metacognitive strategies used to carry them out successfully.

The Current Study

In this study I set out to investigate the aspects of vocabulary ability elicited by two types of vocabulary tests. More specifically, I asked: What do learners do to retrieve the words being tested? Other issues related to the use
of vocabulary strategies were investigated, namely the possible effects of contextual changes, topic familiarity, and individual differences on the types of strategies the participants use.

The main purpose of this study was not to present a conclusive answer to those issues, but to explore them in order to gain a preliminary understanding of what aspects of vocabulary ability are involved in taking vocabulary tests.

Two types of vocabulary tests were used in this study. The first type presents words in the context of a sentence followed by a series of four alternatives, whereas the second type presents words in a larger context (e.g., three passages of approximately 400 words each). In both tests the participants were asked to choose the alternative containing the word or expression that best kept the meaning of the tested word.

The basic difference between the two tests was the amount of context surrounding the words to be tested. The first test format was chosen due to its popularity in educational settings. On the other hand, this type of test has received negative criticism (e.g., Hughes, 1989; Weir, 1990) and it has been suggested that words to be tested should be presented in a context larger than a sentence. This suggestion led me to choose the more contextualized second test as a basis of comparison.

This study was exploratory. My goal was to set the stage for future research on the validation of vocabulary assessment instruments as called for by Cohen (1992). In defining vocabulary ability for testing purposes (cf. chapter two), a test designer should consider not only the elicitation of the pragmatic and linguistic knowledge of words, but also the processes and strategies used to access that knowledge for use
in a given context. Therefore it is necessary to investigate to what extent different vocabulary tests elicit the processes and strategies associated with word retrieval (cf. chapter 2).

Overview of the Contents of Remaining Chapters

Chapter two

This chapter provides an overview of issues on vocabulary ability, testing, and verbal reports, including how they relate to this study. The section on vocabulary includes several dimensions which are identified in applied linguistics as relevant to our understanding of second language learner's vocabulary development (e.g., linguistic and pragmatic knowledge of words).

The section on testing presents an overview of a current framework of language testing and issues involved in the definition of abilities to be measured by tests (Bachman, 1990). A definition of vocabulary ability (Chapelle, 1994) according to that framework is also presented, followed by some considerations on the use of verbal reports in language testing (Cohen, 1987, 1992).

The last section deals with verbal reports, beginning with the description of a model of information processing and an overview of verbal reports as used in psychology (Ericsson and Simon, 1984, 1987). Considerations on the use of verbal reports in second language research are presented.
Chapter three

This chapter provides a description of the procedures used for data collection, the selection of participants in this study, and the instruments used, as well as the methods used for analyzing the data obtained.

The steps followed in conducting the think aloud are those suggested in Ericsson and Simon (1984, 1987), including instructions, training of participants, and prodding.

Chapter four

This chapter presents and discusses the verbal report data. The abilities and strategies used by the participants in the two types of tests are classified and presented in different categories. The data are quantified and discussed in relation to the participants' characteristics (e.g. level of proficiency) and framed by the theoretical foundations on which this study rests.

This chapter also presents the participants' general approach to the test tasks, as well as considerations of the effects of test format, topic familiarity, and individual differences on the abilities/strategies elicited. Examples illustrating the abilities and strategies used and their functions were taken from the participants' reports and are discussed in relation to the participants and to the literature on vocabulary ability.

Chapter five

This chapter presents a summary of this study and its contributions and implications for future research. The implications identified relate to research in language testing
and vocabulary ability and the use of test results in educational settings. Limitations of the present study and suggestions for further research are also presented.
LITERATURE REVIEW

Introduction

The purpose of the study reported here was to investigate, using thinking aloud protocols, the aspects of vocabulary ability involved in taking two types of vocabulary tests. Therefore there are three major areas of second language research to which this study is particularly related, namely vocabulary learning, language testing, and the use of verbal report data.

The first section of this review of the literature deals with research concerning vocabulary learning, especially the dimensions identified by applied linguists as relevant to our understanding of second language vocabulary development. These dimensions range from the context of vocabulary use to vocabulary strategies and processes.

The next area to be considered is second language testing. First, a current framework of communicative language ability is described together with the aspects to be considered in defining language abilities to be measured by tests. Second, a communicative definition of vocabulary ability is presented, followed by some considerations on the use of introspective data as it begins to be used in language testing research.

The last section includes research on the theoretical aspects of verbal reports, as well as the methods used to elicit think alouds. The methods described on this section are the ones followed in gathering the data for this study. The emergent use of introspective data on second language research is also considered as it relates to the increased research interest in language learners' strategies.
Second Language Vocabulary

Applied linguists have identified several dimensions relevant to our understanding of second language learners' vocabulary development. For a review of this literature, see Chapelle (1994). These dimensions include the context of vocabulary use, the mental lexicon organization, vocabulary size, linguistic and pragmatic knowledge of word characteristics, vocabulary processes associated with lexical access, and the metacognitive strategies used to access and manage word knowledge. This section will deal with each of these aspects separately.

The context of vocabulary use

According to the interactionalist view of language ability (e.g., Bachman, 1990), vocabulary ability "will differ qualitatively depending on the context in which it is used, and therefore must be specified with reference to that context" (cited in Chapelle, 1994:164).

Chapelle (1994) uses Halliday and Hasan's (1989) theory of context to explain the effects of a particular context of use on "the linguistic choices a language user can make during linguistic performance" (Chapelle, 1994:164). Their theory of context includes three elements: field, tenor, and mode. Field includes the topic present and actions occurring in a particular context and the setting in which language is used. Tenor includes the participants and their relationships and objectives. Mode includes the channel and the genre of the language used in context. Differences in vocabulary occur due to differences in field, tenor, and mode. Therefore learners' vocabulary would differ "depending on whether they were
reading a newspaper at home, or listening to a chemistry lecture in a classroom, for example" (Chapelle, 1994:165).

Vocabulary size

Some of the problems that arise when estimating vocabulary size have to do with the definition of the notion 'word' (e.g., derivatives and compounds may or may not be counted as 'words') and the difficulty of finding a reliable procedure for assessing vocabulary knowledge (Aitchison, 1987). Dictionaries have been used to estimate vocabulary size (Seashore and Eckerson, 1940; Diller, 1978). Learners are tested on words randomly selected and the proportion of their knowledge is applied to the overall number of words in the whole dictionary, thus providing an estimate of their vocabulary size. Additional problems arise with this procedure: the bigger the dictionary, the more words people are found to know. Based on such procedures, estimates of the words known by educated adult native speakers range from 50,000 to 250,000 words (Aitchison, 1987).

The experiments above are based on the assumption that the number of words known by an individual can be assessed in absolute terms. This assumption is challenged by theories that consider vocabulary ability as definable only in face of the context in which words are used (e.g., Chapelle, 1994). Vocabulary size is then considered to change qualitatively depending on the context of use, and therefore should be assessed with reference to particular contexts.

Since vocabulary size estimates for native speakers vary greatly, "there is no absolute to compare with the size of [second language] learners' vocabularies" (Chapelle, 1994: 167). Some researchers believe this growth can be assessed in
absolute size (Nation, 1993), but in this case vocabulary would not be "defined with reference to particular contexts of vocabulary use" (Chapelle, 1994:167). In other words, since vocabulary knowledge is considered to change qualitatively depending on the context, absolute size estimates will not be indicative of performance in a given context.

The organization of the mental lexicon

The numbers of words known by an educated adult native speaker is "unlikely to be less than 50,000, and may be as high as 250,000... Also remarkable is the speed at which the mental lexicon works: native speakers can recognize a word of their language (e.g., access its meaning) in 200ms (one fifth of a second) from its beginning. In many cases this is well before all the word has even been heard" (Aitchison, 1987:7). These impressive facts suggest that the mental lexicon is arranged systematically.

Aitchison (1987) sets out to form an overall picture of the mental lexicon: "How do humans manage to store so many words, and how do they find the ones they want?" Several bodies of research on different aspects of the mental lexicon are examined in order to gain a better understanding of the nature of the "human word-store." Some of the conclusions reached are summarized below.

Words, according to Aitchison, are "like coins, with meaning and word class on the one side and sounds on the other... This link is easily broken, as shown by the common experience of knowing that a word exists, and being quite sure of its meaning, but being unable to "clothe it in sounds" (Aitchison, 1987:191). The fact that the sounds of words seem to be easily detached from meaning and word class is
interpreted as an indication that the mental lexicon might be arranged in two different ways: one component dealing with the semantics and word class and another phonologically organized. The semantic and word class arrangement is intended to facilitate production, whereas the phonological arrangement is suitable for "speedy recognition."

Humans link words together in their minds. Words seem to be organized in semantic fields. This organization aids production in that a speaker can "pick easily from a particular topic area, comparing several possible words which are linked closely together" (Aitchison, 1987: 85). From the perspective of word recognition, however, a more convenient arrangement would be "to have all instances of the same sound sequence closely bound as in a printed dictionary. This would enable a hearer to compare them fast" (Aitchison, 1987:191).

Words that have similar sounds at their beginnings and ends are those which are most closely linked. This arrangement makes it possible for listeners to "examine several of them together, and find the best fit for what they have heard" (Aitchison, 1987:191).

The mental lexicon organization then takes into consideration the needs of production and recognition, since "production begins with the semantics and syntax," whereas recognition "begins with phonology" (Aitchison, 1987:193). The needs of memory seem also to be taken into account in the lexicon organization: "Within each component, the set-up may have been modified not only to aid speedy retrieval but also to make words easier to remember. For example, the arrangement of words into word classes, and into clumps of co-ordinates, possibly occurs at least partly because the memory needs a more structured system to cope with the tens of thousands of words involved" (Aitchison, 1987:194).
The considerations about the mental lexicon organization presented above were drawn from research using data provided by native speakers (e.g., "slips of the tongue" and word searches), particularly those speakers of British and American English. The extent to which those conclusions are representative of other languages is yet not known. "The fact that different languages may require different organizations of the mental lexicon leads to a number of questions for the future", says Aitchison (1987:204) in the concluding chapter of her book.

Regarding the mental lexicon of second language speakers, many questions are still unanswered. Much of the research done has been on neurological and psychological aspects of bilingualism, "much of which still remains equivocal on the issue of lexical organization" (Channell, 1988:86). Some studies seem to point to separate listings for the two languages, while others argue that there is a single lexical store. Most studies seem to show that there is interaction between the lexicons of the two languages in one user (e.g., Albert and Obler, cited in Channell, 1988).

Does the second language learner's lexicon of a language resemble that of monolingual native speakers of that language? There is some evidence from word association tasks that they are different (Meara, 1982). The "relative stability of responses to many word association stimuli recorded for monolinguals is not found in second language learners" (Channell, 1988:86). Yet in the absence of definite research evidence either way, argues Channell, "it makes sense for L2 theorists to draw on L1 models, and hence to test their validity for L2 theory, until there is definite evidence that they should not do so" (Channell, 1988:86).
The lexicon of the low-proficiency second language learner has been described as loosely organized, with connections made on the basis of phonological features. "It has also been suggested that vocabulary organization would not be fixed at any stage of development, but the connections the language user makes between words would depend on contextual factors prompting those connections" (Chapelle, 1994:166).

**Word characteristics**

The knowledge of word characteristics includes phonemic, graphemic, morphemic, syntactic, semantic, pragmatic and collocational features (Chapelle, 1994).

Carter (1987) identifies several characteristics which define 'knowing a word' in a second language:

1. How to use it productively and having the ability to recall it.
2. The likelihood of encountering it in spoken or written contexts.
3. Its syntactic patterns, underlying forms, and derivations.
4. Its relations with other words in the target language and in a L1.
5. Its relative coreness as well as its more marked uses.
6. Its different meanings and its collocational patterns.
7. Knowing words as part of or wholly as fixed expressions.
Chapelle (1994) summarizes some instances of word knowledge characteristics as they occur during the process of acquisition. Knowledge of specific words can be incorrect, incomplete or unanalyzed (Bialystock and Sharwood Smith, 1985). Incorrect knowledge refers to lexical representations which do not correspond to the target language, such as incorrect spelling or semantic representations. Incomplete knowledge refers to the gaps in the learner's knowledge of a word, resulting in confusion among two words with similar forms (Chapelle, 1994). Unanalyzed knowledge is "what the learner knows as a unit but cannot break up or use creatively," that is, the learner "does not know how to change [words] morphologically or use them in other phrases" (Chapelle, 1994:167).

The characteristics a learner knows about individual words should be considered in relation to the context of use for those words: "the characteristics ... would be prompted by and would depend on the context in which those words are used" (Chapelle, 1994:168).

**Vocabulary processes**

This section will focus on models of retrieval for both word production and word recognition. Vocabulary processes associated with lexical access will also be identified.

"Humans behave like jugglers when they use the mental lexicon ... they have to deal with semantic, syntactic and phonological information at the same time" (Aitchison, 1987:165). The processes used for retrieving words for production (e.g., from meaning to form) seem to be different from those used for comprehension (e.g. from form to meaning). Using Aitchison's (1987:164) metaphor, "we cannot take it for
granted that [speakers] utilize the same processes in a different order ... just as we cannot automatically assume that going upstairs uses identical muscles to going down, but in the reverse sequence." Following is a summary of models that attempt to explain the processes of retrieving words for production and recognition purposes, found in Aitchison (1987).

According to Aitchison, any model of retrieval must recognize the fact that "humans seem to activate many more words than they need as they plan speech, words which occasionally pop into one's utterance quite inconveniently" (Aitchison, 1987:175), for example, "I'll give her a prescription (subscription) for a magazine", said by someone who was trying to keep a doctor's visit a secret. The occurrence of slips such as this is explained by the sound link between the two words.

A "spreading activation model" like the one presented in Figure 1 is an attempt to explain the process of retrieving a word (in that example, the word "beaver"). The situation is "likened to a complex electric circuitry, in which current flows backwards and forwards between particular points and in so doing excites numerous other points around. The relevant points and links get more and more excited and the irrelevant ones get suppressed, until finally one word wins out over the numerous ones activated" (Aitchison, 1987:173-4). As happens with word production, readers and speakers apparently consider many more words than they eventually select. A large number of words are also activated in the lexicon, then those that are not required are gradually suppressed.

Aitchison notes that (in speech production) "since words are spoken fast, and need to be interpreted fast, there may also be more attention to word beginnings than word endings,
Figure 1. An interactive activation model
(From Aitchison, 1987:174)

Figure 2. A spreading activation model.
(From Aitchison, 1987:174)
in the hope that the early sections of a word can lead to a decision" (Aitchison, 1987:187-188). Aitchison (1987:189) concludes that "the same type of spreading activation model found in word production is also found in word recognition--even though many details are still obscure."

Figure 2 illustrates the "spreading activation model" for recognition of an orally produced word. The perceived sequence "b l...s" "t" led to the activation of several words until the appropriate word was eventually selected. Figure 3 represents a score-board system that might keep track of the relative probability of different words candidates. Such a device would "from moment to moment assign to every item a relative probability rating for each of three factors, perceptual evidence, frequency weighting, and contextual evidence" (Norris, 1986, cited in Aitchison, 1987:188-89).

<table>
<thead>
<tr>
<th>Stage I Likelihood ranking</th>
<th>Stage II Likelihood rank.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(order of probability)</td>
<td></td>
</tr>
<tr>
<td>Blame</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Blade</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Brace</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Bracelet</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Blanket</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
</tr>
</tbody>
</table>

Key: P - perceptual evidence; F - frequency probability; C - context

Fig. 3. Possible score-board system. (Aitchison, 1987:188).
Vocabulary strategies

The metacognitive strategies used by developing second language learners to retrieve and manage vocabulary knowledge and processes are typically compensatory in nature. Those strategies are used in order to achieve communicative goals despite limited vocabulary knowledge (Chapelle, 1994).

Blum-Kulka and Levenson (1983) have identified the following strategies: circumlocution, paraphrase, language switch, appeal to authority, change of topic, and semantic avoidance. "As ability increases and the learner has access to a larger number of words, these compensatory plans become less essential, but metacognitive strategies remain important because any linguistic performance is the result of learners' successful metacognitive control" (Chapelle, 1994:167).

The strategies mentioned above are primarily used in language production, involving interaction between interlocutors. In the context of a test, however, the uses of these strategies would probably be restricted since their communicative functions would be different. In other words, in a test situation learners lose the kind of control over their performance when they engage in normal communication. The test situation is more similar to listening or reading in this respect, than to speaking or writing. The test format may also constrain the range of possible strategies.

Strategies which learners can use to "decode for themselves the meanings of words" have been identified, particularly those used in reference to reading (Carter, 1987:187). Nation (1983), cited in Carter (1987:166-167), views a strategy as "just a means of acquiring the unconscious skill than an efficient reader already has". He describes a series of strategies that can be followed in encountering unknown or "half-known" words in the course of reading:
(1) Identify the part of speech to which the word belongs.
(2) Examine the immediate context (e.g., clause or sentence containing the target word) for clues.
(3) Look for the relationship between the clause or sentence containing the word and other sentences or paragraphs (e.g., conjunctions).
(4) Use the knowledge gained from the previous strategies in order to guess the meaning of the word.
(5) Check if the guess is correct by (a) comparing the part of speech, (b) replacing the unknown word with the guess, and (c) breaking down the word into its prefix, root and suffix, if applicable.

Carter (1987:188) analyzes the strategies identified by Nation as "progressing from 'word building' to 'definition clues' and then to 'inference clues' which require higher levels of analytic skills." Carter also points out "the inevitability of learners encountering words in context (and increasingly so as the learner becomes more advanced)" (Carter, 1987:188), which underscores the importance of inferential strategies "which might put this increased learning load on a more systematic basis" (Carter, 1987:189).

Second Language Testing

Language tests have many uses in both education and research. In educational settings test results are used to make decisions about students in terms of entrance, placement, diagnosis, progress, and grading. In research, language tests are used to obtain information on the very nature of language proficiency, in the study of language acquisition and attrition, and in assessing the effects of instructional methods on language learning (Bachman, 1990).
Vocabulary tests play an important part in the decisions made in educational settings since they are part of most test batteries. Vocabulary tests are also used in language research for the purposes mentioned earlier. The use of a vocabulary test for placement purposes (e.g., vocabulary was the only ability tested) is reported on Meara and Buxton (1987).

From the various uses of vocabulary tests presented above, we can conclude that the interpretations of vocabulary test results, as well as the decisions that are taken based on them, have a potential impact on the lives of those involved in taking those tests. Therefore ensuring that the decisions made on the basis of vocabulary test scores are justified should be the main concern of those designing or using such a test.

Bachman (1990) examines some "fundamental considerations in language testing." This section starts with some of those considerations regarding the definition of abilities to be measured, then presents a communicative definition of vocabulary ability, and moves on to present some considerations on a process-oriented approach used to further justify the interpretation and uses of test results.

Communicative language ability

Bachman’s (1990) framework of communicative language ability is in line with earlier work on communicative competence (Canale and Swain, 1980; Savignon, 1983; Widdowson, 1983) in that it also "recognizes that the ability to use language communicatively involves both knowledge or competence in the language, and the capacity for implementing this competence" (Bachman, 1990:81). Bachman’s framework, however,
goes beyond earlier models in that it "attempts to characterize the processes by which the various components interact with each other and with the context in which language use occurs" (Bachman, 1990:81).

Bachman's framework includes three components: language competence, strategic competence, and psychophysiological mechanisms. Language competence comprises two types of competence: organizational and pragmatic. Organizational competence includes "the knowledge employed in creating or recognizing grammatically correct utterances, in comprehending their propositional content, and in organizing them to form oral or written texts" (Bachman, 1990:86). Pragmatic competence includes "knowledge of language functions, of sociolinguistic rules of appropriateness, and cultural references and figurative language (Bachman, 1990:89-90).

Strategic competence is seen as the "capacity that relates a user's knowledge structures and the features of the context in which communication takes place" (Bachman, 1990:98). It includes three components: assessment, planning, and execution. The assessment component enables us to (1) identify the information needed for achieving a communicative goal in a given context (e.g., language variety); (2) determine what language competences will be more effective (e.g., native language, foreign language); (3) assess the abilities and knowledge shared by our interlocutor; and (4) evaluate the extent to which the communicative goal has been achieved. The planning component "retrieves relevant items (e.g., grammatical, sociolinguistic) from language competence and formulates a plan whose realization is expected to achieve the communicative goal" (Bachman, 1990:100). The execution component "draws on the relevant psychophysiological mechanisms to implement the plan in the modality and channel"
appropriate to the communicative goal and the context” (Bachman, 1990:103).

Bachman’s (1990) construct definition of communicative language ability includes the language knowledge and fundamental processes required to access that knowledge as well as the context of language use. The component controlling the interaction between the two is defined as ‘strategic competence’, “the metacognitive strategies required for assessing contexts, setting goals, constructing plans and controlling execution of those plans” (Chapelle, 1994:164). Therefore context of language use, knowledge and processes, and metacognitive strategies are an integral part of the definition of any aspect of language ability.

Defining vocabulary ability

Chapelle (1994) defines vocabulary ability in terms of the context of language use, vocabulary knowledge and processes, and the metacognitive strategies required for vocabulary use in context. Different contexts of language use will cause vocabulary to “differ qualitatively,” and therefore it “must be specified with reference to [a] context” (Chapelle, 1994:164). For example, learners’ vocabulary will differ depending on whether they are reading a magazine or listening to an academic lecture.

Chapelle (1994) describes four dimensions pertinent to the definition of vocabulary knowledge and processes: (1) the number of content words a person knows within a particular context; (2) what a person knows about each content word (linguistic and pragmatic characteristics); (3) how morphemes are stored in the mental lexicon (organization); (4) the vocabulary processes associated with lexical access; and (5)
the metacognitive strategies used. Each of these areas "can change as learners' vocabulary develops" (Chapelle, 1994:165). (See previous sections on vocabulary size, mental lexicon organization, vocabulary processes and strategies).

Metacognitive strategies are particularly important for "developing learners who must devise and execute plans for achieving their communicative goals despite limited vocabulary knowledge" (Chapelle, 1994:165). Some of these strategic plans are: circumlocution, paraphrase, language switch, appeal to authority, change of topic and semantic avoidance (Blum-Kulka and Levenson, 1983). The use of these compensatory strategies will diminish as the learners' vocabulary knowledge develops and they have access to more words. Metacognitive strategies, however, remain important because they "call on and manage vocabulary knowledge and processes for all language users" (Chapelle, 1994:168). Native speakers usually accomplish cognitive processes and metacognitive strategies automatically, but when they are faced with unfamiliar or more demanding contexts, automaticity breaks down. For example, reading a legal document places different demands on the lay person, causing one to make a conscious use of strategies that are typically used automatically in less demanding situations. (See sections on vocabulary size, mental lexicon organization, vocabulary processes).

"All language tests must be based on a clear definition of language abilities and must utilize some procedure for eliciting language performance" (Bachman, 1990:9) that can be interpreted as an indication of either the presence or lack of those abilities. Regarding the definition of vocabulary ability presented above, what testing method would elicit all aspects of vocabulary ability? Chapelle (1994) argues that the typical paper and pencil tests usually administered to students in classrooms cannot tell us about the processes and
strategies students use to access their knowledge of words because those tests only give us product data. Although it can be argued that one could conceive of a test in which such processes could be revealed, such a test would be very difficult to construct (Andre, 1994, personal communication) and would not be likely to be used in more typical testing situations (e.g., in educational settings).

Considering this argument, Chapelle calls for an approach to assess the processes used to perform on vocabulary tests. The goal of such an approach would be to find out what aspects of vocabulary ability (word knowledge, processes, and strategies) are used to perform successfully in the context of different vocabulary tests.

Process-oriented approaches to language testing

Cohen (1992) examines test-taking strategies identified through verbal reports in several studies. These strategies were used in performing on various tests: multiple choice, cloze, open-ended questions, and compositions. He suggests that "second language acquisition researchers would probably want to consider validating the testing measures that they use through triangulation, which would include the collection of test-taking strategy data on subsamples of respondents" (Cohen, 1992:28).

Cohen suggests that potential shortcomings of multiple-choice items (Hughes, 1989; Weir, 1990) could be investigated using verbal report techniques--for instance, the role of guessing or simple matching of stems with alternatives, the extent to which items meant to test one thing test another (e.g., a grammar item that tests lexical knowledge as well), and the extent to which respondents eliminate choices as
absurd or through common knowledge. Cohen argues that "it would appear that test-taking strategy research can be used to substantiate or refute such claims with respect to a given test, in a given test administration, with given respondents" (Cohen, 1992:27).

In earlier studies (e.g., Cohen, 1984) verbal reports were used to elicit test-taking strategies. In a series of studies dealing with how language learners take reading comprehension tests, Cohen found that "due to flaws in the test or due to certain test-taking strategies, students may not be displaying a representative performance of their language competence" (Cohen, 1984:71). He adds that "the outcome of [some] strategies may be that respondents got items wrong for the right reasons or right for the wrong reasons" (Cohen, 1984:71). Cohen also points out that test designers "should strive for a closer fit between how [they] intend for their tests to be taken and how respondents actually take them. This may involve changing the format of the test or training the respondents to deal with that format more effectively" (Cohen, 1984:78).

Cohen (1984, 1992) focuses particularly on the types of strategies used in responding to test items as presented in different test formats. Since differences were found between the ways test designers intended their tests to be taken and the ways the learners actually took them, he urges test designers to ensure that both sides conceptualize the testing situation in the same way. In his 1992 manuscript Cohen examined several studies on learners' strategies in taking different test formats (e.g., multiple choice reading comprehension tests, open-ended questions, and compositions) and argued for using introspective methods (e.g., think-aloud) to elicit learners' strategies in a subsample of test items as a way of further validating a given test.
The process-oriented approach (e.g., verbal report techniques used to investigate test-taking strategies) used in the studies reviewed or conducted by Cohen can also be used in investigating the strategies and processes learners follow to access their knowledge of the linguistic material being tested. Such an approach could be used, for instance, to find out what aspects of vocabulary ability, as defined by Chapelle (1994), are involved in performing in different types of vocabulary tests.

The research conducted or reviewed by Cohen studies was concerned with more general test-taking strategies, that is, how the test-takers approach the tasks on the tests. For example, did they read the questions before reading the passage in which the questions were based? In the study reported here, however, I am particularly interested in assessing the participants' use of different aspects of their vocabulary ability, including vocabulary strategies, elicited by different tests.

Learning about the strategies learners use to take different tests and the strategies and processes they follow in accessing their knowledge of the material being tested could be used in combination to further support the "adequacy and appropriateness of inferences and actions based on test scores" (Messick 1988:13). In other words, stronger evidence for the validity of a given test would be obtained if we ensure that our tests tap into the aspects of the ability they are intended to measure (the present study) and that they are taken the way we intended them to be taken (Cohen, 1984).
Verbal Reports

Introduction

The previous section dealt with testing issues such as the importance of having a clear definition of the ability a language test is intended to elicit and the effects test format can have on test performance. It also covered test-taking strategies and a method used to investigate them, namely verbal reports.

This section will further discuss verbal reports on thinking from both theoretical and practical perspectives. Ericsson and Simon (1980, 1984, 1987) have extensively examined the use of verbal protocols. Some of their considerations on “verbal reports on thinking,” including their information processing model and methodology for eliciting think alouds, are presented below. The use of verbal reports in second language research is also considered.

According to Ericsson and Simon (1987:24), “After a long period of studying human performance and abilities, research in psychology is now seeking to understand the underlying cognitive processes. Researchers are looking for observations on thinking that would allow tracing the intermediate steps of the thought processes.” Subjects' reports on their thinking is one of the types of observations that provide data on cognitive processes. After an early period in which psychologists widely used introspective methods, “they fell into disrepute during the era of behaviourism” (Ericsson and Simon, 1987:24), but since the 1970s verbal reports have been used within an information-processing framework, especially in the study of problem solving.
In outlining an information processing model, Ericsson and Simon hypothesize that human cognition is "information processing": "a cognitive process can be seen as a sequence of internal states successively transformed by a series of information processes" (Ericsson and Simon, 1987:25).

It is assumed that information is stored in several memories having different capacities and accessing characteristics: "several sensory stores of very short duration, a short-term memory (STM) with limited capacity and/or intermediate duration, and a long-term memory (LTM) with very large capacity and relatively permanent storage, but with relatively slow fixation and access times compared with the other memories" (Ericsson and Simon, 1987:25-26).

What type of information can be accessed and verbally reported? According to the model described above, the information that is available for further processing (e.g. for producing verbal reports) is the recent acquired information kept in STM. The information from LTM must first be transferred to STM before it can be verbally reported.

Verbal reports are obtained when the participants verbalize the corresponding thought or thoughts while the information is attended to. The crucial aspect of this procedure, according to Ericsson and Simon (1987:27), is that "the sequence of states, i.e., the information contained in attention and STM, remains the same with the verbal report as it would be when the cognitive processes proceed silently".

Ericsson and Simon (1984) reviewed studies comparing silent participants' performance with the performance of "talk-aloud" and "think-aloud" participants. "Talk-aloud" refers to report in which the participant simply vocalizes "silent speech", i.e., heeded information, whereas "think-aloud" is the type of report in which participants "must
convert the heeded information into a verbalizable form to vocalize” (Ericsson and Simon, 1987:33). No consistent differences were found in the accuracy of the generated responses. Participants with think-aloud instructions, however, required more time to complete their solutions.

The think-aloud protocol

The type of instructions given to participants in a think-aloud experiment may affect the verbalization procedures the subjects use. Previous training and the types of “reminders” (e.g., “keeping talking”) given to subjects may also affect the data obtained.

Ericsson and Simon (1984, 1987) have developed and examined instructions for thinking aloud in terms of their effects on the verbal reports. They also looked at the effects of training and the effects different types of reminders can have on the subjects' verbalization. The following is a summary of Ericsson and Simon’s considerations. The studies mentioned are cited in Cohen (1987).

Instructions. Instructions of the type “try to think aloud” or “verbalize your thoughts” simply ask the participants to vocalize their thoughts which are presumed to have the form of inner speech. However, when the investigator adds to this type of instruction “tell me everything that passes through your head”, he/she is asking the subjects to vocalize their thoughts whether they are encoded orally or not. In this case subjects will have to first encode the content of attention in order to later verbalize their thoughts. These types of instruction will then determine whether or not the subjects will report only thoughts that have the form of inner speech.
Instructions may also contain other information or requirements, "usually a request for completeness" (Ericsson and Simon, 1984:81). For instance, subjects may be asked to say what they are thinking no matter how irrelevant it may seem or even if it represents wrong attempts to the solution of the problem.

To avoid changing the structure of the subjects' thoughts, instructions can make it clear that they should not plan what to say or speak after the thought, but rather talk while thinking, pretending no one is present but themselves.

White (1980, cited in Cohen, 1987:91) points out "the need for the subjects and the researcher to conceptualize the situation the same way." Failure to obtain the intended type of data may be the result of an inadequate understanding on the part of the subjects as to how they are supposed to report.

Training. Training can take on two forms. First subjects can be given initial warm-up problems to acquaint them with the experimental situation, microphones, and tape recorders. Second, more extensive warm-up procedures can be used to train the subjects to conform to the think-aloud instructions. Cavalcanti (1987) found that subjects asked to think aloud while reading would read large chunks of text and then retrospect. She trained them to give think-aloud data each time they noticed they had paused in the course of reading.

During the training sessions the investigator can interfere to make sure the subjects are complying with the instructions. During the actual experiment, however, he/she should avoid interfering, unless this interference has to do with reminding the subjects to keep on talking if they fall silent.
Reminders. The experimenter is generally present during the thinking aloud experiments and his/her role is to monitor the verbalizations by reminding the subjects to talk when they fall silent. These reminders, usually given after 15-second to 1-minute pauses, have the standardized form of "keep talking" or "what are you thinking about?"

Another instance in which the researcher may have to interfere is to make sure that the respondents are producing think aloud-data. Cohen and Rosenfeld (1981) found that their subjects needed prodding in order to produce think-aloud data rather than retrospective data.

The two types of reminders mentioned above may have different effects on verbalizations. "Keep talking" may have little effect on the subjects' processing. However, "what are you thinking about?" is more likely to induce subjects to provide "other-oriented" reports, i.e., descriptions or explanations instead of simply reporting their thoughts.

Verbal reports in L2 research

As researchers have become increasingly interested in investigating the strategies used in learning a second or foreign language, there has been a growing interest in using learner's reports of their own intuitions and insights as a complement to classroom observation and other measures (Cohen, 1987).

Cohen (1987) identifies three basic categories of learner report data. They are referred to as self-reports, self-observation, and self-revelation. Self-reports are usually statements "based on beliefs or concepts that the learners have about the way they learn language" (e.g. "I try to figure out word meaning from context") (Cohen, 1987:83). Self-
observation refers to the "inspection of specific language behavior, either introspectively or retrospectively" (Cohen, 1987:84). Self-revelation consists of reporting on thought processes while the information is being attended to.

Cohen also identifies six factors which characterize the data obtained from the three categories of verbal report described above. These include: the number of participants, the research context, the recency of the event, the mode of elicitation and response, the formality of elicitation, and the degree of external intervention. Those factors can vary in accordance to the type of verbal report and/or research purposes.

Table 1, taken from Cohen (1987:85), provides a summary of the types of data and their descriptors. The number of participants refers to how many investigators and subjects are involved in the data collection process. The research context refers to when, where, and to some extent how the data is collected. The recency of the event relates to the proximity of the verbal report to the actual learning event. Whereas think-aloud data is obtained at the time of the learning event, self-observational and self-report data can vary with respect to how soon after the event they take place. The mode of elicitation and response concerns whether the investigator elicits the verbal report data orally or by means of written instructions, and whether the informants respond orally or in writing. The formality of elicitation has to do with the degree of formal structure imposed on the elicitation by an
Table 1. Types of data and their descriptors

<table>
<thead>
<tr>
<th>TYPE OF DATA</th>
<th>DESCRIPTORS</th>
<th># Participants</th>
<th>Context</th>
<th>Mode</th>
<th>Degree of External Interven.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>introspection</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>LO-HI</td>
</tr>
<tr>
<td>retrospection</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>LO-MED</td>
</tr>
<tr>
<td>Self-Disclosure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>think-aloud</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>LO</td>
</tr>
</tbody>
</table>

KEY:
- # Participants: group, individual + investigator, individual alone
- Context: during class, other
- Mode: elicitation--oral, written; response--oral, written
outside investigator. For example, self-report data can be obtained either through a set of fixed questions or allowing for the respondents to provide information along the lines that they themselves determine. The degree of external intervention is the extent to which the investigator shapes the respondent's reporting process. The data gathered can be directed along certain lines according to the investigator's instructions.

Cohen (1987) summarizes the three instances identified by Ericsson and Simon (1980) in which poor methods of reporting can result in faulty reporting. First, they claim that faulty reporting can result if the information is not attended to, since for information to be available from short-term memory, it must be attended to. Thus, in the case of probes requesting subjects to produce verbal reports about information not attended to, it is possible that the subjects will infer missing information and generalize incomplete memories. Second, they suggest that faulty reporting can result if not all the information which is in short-term memory at the time of the reporting, is actually reported. For this reason, it is important to use elicitation procedures that obtain reports that are as complete as possible. Finally, faulty reporting can result when not all information previously available in short-term memory has been retained in long-term memory, or is retrievable from long-term memory.

From all the considerations above, it seems that "verbal report data can be a useful research tool under certain conditions and with certain limitations" (Cohen, 1987:91). Ericsson and Simon (1980), cited by Cohen (1987:91), put it as follows:

For more than half a century ... the verbal reports of human subjects have been thought suspect as a source of evidence about cognitive processes... Verbal reports, elicited with
about cognitive processes... Verbal reports, elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and roughly reliable source of information about cognitive processes... They describe human behavior that is as readily interpreted as any other human behavior.

Cohen (1987:92) comments on the potential value of verbal report data to the study of second language learning: "As we find out more about the processes that learners use, we are better equipped to test hypotheses about strategies that we would predict are likely to produce the greatest success for given types of learners."

Language testing is an area of second language research in which verbal reports have begun to be used. According to Cohen (1992:27), "we appear to have entered an era where it is more acceptable not only to look at process approaches to language testing, but also to use verbal report techniques to better understand these processes and the test-taking strategies that respondents use." He adds that this type of investigation will "provide important information as to which testing methods would be potentially more or less reliable and valid for SLA (Second Language Acquisition) research" (Cohen, 1992:27).
METHODS

Introduction

My purpose in this study was to investigate the aspects of vocabulary ability elicited by two types of English vocabulary tests. More specifically, in this study I have posed the question: What do learners do to retrieve the words which are being tested? Such a question cannot be answered by simply looking at test scores. Introspective methods that produce learners' reports on their language strategies while working on a given task are likely to provide a more complete answer to that question.

Cohen (1987) identifies three categories of learner report data: self-report (e.g., learners general statements about their learning behavior), self-observation (e.g., retrospection), and self-revelation (e.g., think aloud). (See Table 1 in chapter two for a detailed presentation of these types of data and their descriptors).

Both retrospection and think-aloud are methods that produce learners' report on the language strategies they use. In this study the think-aloud was chosen because this procedure is believed to disclose learners' "thought processes while the information is being attended to" (Cohen, 1987:84) and, contrary to retrospection, it does not allow time for elaborating on the thought afterwards (e.g., edit and/or analyze the thought after it occurs). Think-alouds also prevent the forgetting that takes place after the information is processed, which makes retrospective reports somehow incomplete.
The remainder of this chapter presents the steps followed in selecting the participants, constructing the instruments, piloting the study, conducting the experiment, and analyzing the verbal report data obtained.

Participants

In choosing the participants for this study the following criteria were observed: they should come from similar cultural and language backgrounds and be comfortable with the think-aloud procedures. To find out whether they were comfortable with the procedure, the three participants were asked to perform the task of sorting different picture cards into groups, using whatever criteria they found appropriate, and verbally report their thoughts while performing the task. All three reported being comfortable with the procedure. Table 2 presents the characteristics of the participants.

Three female native speakers of Brazilian Portuguese, ranging in age from 27 to 31, and residing in the US for over a year were selected. Their TOEFL scores ranged from 500 to 572; all three were college educated and intended to pursue graduate studies. Choosing Portuguese speakers was a "natural" choice because I am a native speaker of that

Table 2. Characteristics of participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participant A</th>
<th>Participant B</th>
<th>Participant C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>27</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>Background</td>
<td>Linguistics/Literature</td>
<td>Occupational Therapy</td>
<td>Medicine</td>
</tr>
<tr>
<td>Length of stay in the US (in months)</td>
<td>15</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>TOEFL scores</td>
<td>543</td>
<td>500</td>
<td>572</td>
</tr>
</tbody>
</table>
language and would be able to transcribe the data and to identify instances in which the participants used their first language knowledge in answering the test items (e.g., cognates).

Instruments

Two types of tests were used to elicit the think-aloud data. A description of each test, the rationale for selecting the formats, and the steps followed in constructing one of the tests are presented below.

Sentence test

The sentence test was taken from Reading for TOEFL (1992), a publication of ETS designed for learners preparing to take the TOEFL test. No modifications were made to its content.

The test was a typical multiple choice vocabulary test in which underlined words were presented in the context of a sentence followed by a set of four alternatives. Test-takers were to choose the one alternative containing the word or expression that best kept the meaning of the underlined word in the proposition. The test (shown in Appendix A) contained a total of 30 items.

The sentence test was chosen due to its popularity in educational settings. Though commonly used, these kinds of tests have received negative criticism (e.g., Hughes, 1989; Weir, 1990). One of the common criticisms of the typical multiple choice vocabulary test refers to the amount of context in which the words to be tested are presented—a single sentence.
Passages test

The passages test presented words in a larger context and represented an alternative to typical multiple choice tests which present words in the context of single sentences. It was constructed by the investigator and presented underlined words in the context of three passages of approximately 400 words each. The tested words were selected according to the following criteria: some words could only be retrieved by using global contextual clues, some words were cognates, some words had different levels of difficulty (e.g., some would be easy, and some would be difficult for the participants).

The content of the passages was selected according to the following criteria: the topic of one of the passages would be related to each of the three participants' background (e.g., passage one reported on a medical experiment, passage two dealt with issues on first language acquisition, and passage three presented uses of plants in therapy); and the passages would not contain highly specialized words and therefore would be accessible to the lay person. In other words, each participant would be favored with a passage related to her field of study, but would still be able to understand the other two passages.

In choosing the topic of the passages used in reading comprehension tests, it is standard procedure to avoid topics related to the test-takers' background to avoid bias in favor or against certain groups. The assumption behind such a procedure is that general familiarity with a topic is advantageous to test-takers in reading and affects their performance positively. The passages test used here is a vocabulary test and as such does not directly measure reading comprehension. However, the tested words are presented in the context of reading passages and I wanted to find out whether
effects, similar to those observed in reading comprehension tests, would also be observed in the participants' performance on this type of vocabulary test. Since each participant would relate to one passage in particular, the possible positive effects would be equally distributed among the three participants.

The passages were each followed by a set of items focusing on the underlined words. The test-takers were to choose the one alternative containing the word or expression that best maintained the meaning of the word in the text. To answer the questions the participants would have to refer back to the text in order to locate the target word. A total of 31 words were tested. The test is presented in Appendix B.

Pilot Project

The pilot project had two purposes. First, I wanted to check for any problems on the test construction that could hinder comprehension and ascertain the appropriateness of the content for the language proficiency level of the participants. I wanted to avoid the two extremes of using a test that was either too difficult or too easy. The second purpose was to enable me to become familiar with conducting think-alouds.

Two native speakers of English took the tests (without thinking aloud) and commented on the construction of the questions. Afterwards the tests were piloted with a male nonnative speaker of English of approximately the same level of proficiency as the subjects. The nonnative speaker provided think-aloud protocol.

Based on the native speakers' comments, some changes were made on the tests items (e.g., two alternatives were "too
close", some items could be answered without referring back to the passages).

The nonnative speaker used English in his verbal report. He did not spend much time on the test items and retrospected instead of saying out loud what he was thinking at the moment the thoughts occurred. He admitted having difficulty in working on the test and talking at the same time: "I couldn’t concentrate." He also interpreted prods as criticism and would either change his answer or check it more carefully. Regarding content, the tests seemed appropriate to the participants’ level of proficiency, that is, it was neither easy nor very difficult.

The experience I acquired in conducting this pilot project made me realize the need for carefully training the participants in using the same type of tasks in the tests.

**Procedures**

The subjects answered items from the two different types of vocabulary tests and reported their thoughts orally while working on the test items. Their "talk" was audiotaped. This section describes the procedures used in training the subjects and in conducting the actual experiment. The methods used to elicit the think aloud were those suggested in Ericsson and Simon (1984, 1987) including the types of instructions given to the participants, training sessions, and the role of the investigator during the experiment.

Thinking aloud while working on a test is not a common experience for most learners and thinking aloud in a second language can be even more challenging. On the other hand, the test was in a second language and switching back and forth between two languages might represent a potential problem for
the participants. For this reason I decided to allow them to choose whether to use Portuguese, English, or to code switch as they choose.

The participants had reported earlier that they were comfortable with thinking aloud while performing the task of sorting out picture cards (see section on participants). Since that activity was different from the tasks on the tests, I decided to conduct a warm-up session before having the participants work on the actual tests. It consisted of 20 questions in the same format of the ones presented in both tests.

The warm-up session served several purposes: (1) to get the subjects acquainted with the test types; (2) to check their ability to think aloud in such situation; (3) to give me an opportunity to check the effectiveness of my instructions; and (4) to ensure that both subjects and I "conceptualized the situation in the same way" (White, 1980, in Cohen, 1987:91).

Instructions were given in the participants' native language: "While working on the test, say out loud everything that goes through your mind no matter how irrelevant it may seem. Pretend you are alone and think out loud" (Ericsson and Simon, 1984:33-34). Subjects were also alerted that they did not need to explain their thoughts in order to make them understandable to an interlocutor.

Reminders of the type "keep talking" were chosen to be used because they have been found to be less intrusive than those of the type "tell me what you are thinking". The last type may generate an other-oriented verbalization if the subject interprets the prod as a question (Ericsson and Simon, 1987).
During the warm-up the subjects asked clarification questions, became more comfortable with the new situation, and reassured themselves they would be able to think aloud while answering the test items. I had the opportunity to give additional instructions and detect instances of retrospection (e.g., subjects answered an item first, then reported). Two of the subjects were concerned whether they would be able to "talk a lot" in order to help me by providing more data. It was necessary to explain again that they should say whatever came to their minds, but without forcing it.

The actual experiment was conducted immediately after the warm-up session. The participants were asked to follow the same procedures they had followed during that session.

While observing the participants working on the tests, I took notes. These proved helpful in the process of transcribing the tapes by providing valuable information that is inevitably lost in the recording, such as hand gestures, marks on the paper, and facial expressions.

Being present during the whole experiment also gave me the opportunity to have a more complete picture of the situation. For instance, it was possible to notice if the participant was looking at the text or at the question when she verbalized a thought.

The passages test was administered first. There was no time limit set to complete the tests. It took fifty minutes to one hour and a half for the participants to complete both tests. During the actual experiment it was not necessary to remind the subjects to keep talking.
Data Analysis

The data gathered included two hours and twenty minutes of recording (two participants spoke in Portuguese and one used English predominantly) and the notes I took while observing the participants taking the test. The tapes were transcribed within an hour or two after the recording. The two Portuguese recordings were first transcribed in Portuguese then translated into English.

After being transcribed and translated, the data corresponding to each test item were individually examined and a description of the steps that the participant followed in answering the item was added next to it.

The notes taken during the experiment were used in conjunction with the transcripts. They were used to facilitate the transcription process, that is, they helped me to reconstruct the test-taking context, providing information such as whether the participant located the word in the passage before attempting to answer the question or referred back to the text to evaluate her answer. The notes also helped in the identification of the abilities and strategies the participants used.

The next step was to identify the aspects of vocabulary ability (and other abilities) each participant appeared to use to answer each item. The abilities used and corresponding instances were listed together, resulting in two lists (one for each test type) for each participant.

Examining the lists corresponding to each test, I observed that some abilities/strategies were recurrent. These were grouped and some broad categories began to emerge after which categories were subdivided. The resulting taxonomy is presented in the next chapter.
ANALYSIS AND RESULTS

Introduction

This study was designed to investigate the aspects of vocabulary ability elicited by two types of vocabulary tests. In other words, what do learners do to retrieve the words being tested? In searching for an answer to this question, other related issues were also investigated: the possible effects of contextual changes, topic familiarity, and individual differences on the vocabulary strategies and processes used. In this chapter I present and discuss the findings of this study as they relate to my research question and its related issues.

Before presenting and discussing the data, however, I include a section on the general strategies the participants used in taking the test, that is, their general approach to the tasks on the tests.

The next section deals with the development of the taxonomy used and the rationale for adopting the categories chosen. Illustrative examples taken from the participants' reports in both tests, that is, what they did to retrieve the tested words, are then presented and discussed in light of the theory on vocabulary learning.

In the subsequent section the data is quantified and the findings related to the effects of test format, topic familiarity, and individual differences on the vocabulary abilities and strategies used are presented and discussed.
The participants' approach to the tests

The participants' general approach to the tasks on the two tests showed several kinds of similarity.

In taking the sentence test, participants A and C chose to read the whole sentence before considering the alternatives. Participant B first read the underlined words then read the sentences when formulating or testing her hypotheses.

After reading the whole sentence or the underlined word, all three of the participants would usually make a hypothesis, i.e., consider a particular alternative, and evaluate it - "I think it's..."; "It can't be ...". The result of the evaluation process was either acceptance or rejection of the alternative under consideration. After an alternative was rejected, another hypothesis was usually made and the cycle would continue until an answer was chosen or the participant moved on to the next item.

On a few occasions, the participants delayed their answer returning to the item later on. Guessing was the strategy used when no hypothesis was formulated or the participant had considered all the alternatives unsuccessfully ("I have no idea", "I'm going to guess").

Although approaching the test tasks in a similar manner, the three participants' scores varied. This variation was in line with their scores on the TOEFL test, that is, the participant with the highest TOEFL score also obtained the highest scores in both vocabulary tests, whereas the one with the lowest score on the TOEFL obtained the lowest score in the vocabulary tests (cf. Appendix C).
Figures 4 and 5 present a summary of the participants' approach to the sentence test and to the passages test, respectively. The process of retrieving words ("word recall"), as represented in Figures 4 and 5, is identified as taking place after the participants read both the context in which the word occurs and the corresponding questions, though I am aware that this process may have started earlier, possibly when the participants first read the target word. The decision to consider the recall process after the reading was based on the moment the participants initiated their verbal reports, i.e., immediately after they finished reading the questions. In other words, only after the participants started thinking aloud, was it possible to have access to the processes taking place in retrieving the words.

Figure 4. Approach to the sentence test. 
(Adapted from Feldman and Stemmer, 1987:257)
The main difference between Figures 4 and 5 occurs before the process of selecting an answer and can be attributed to test format. After reading the questions in the passages test, the participants needed to refer back to the text to locate the target word before considering the alternatives (cf. Appendix B). The participants read the texts, read the questions, then located the word in the passage and reread the surrounding context (e.g., the sentence containing the target word or a few words before and after it). This procedure was adopted by all three participants and remained constant throughout the test.

**WORD RECALL**

(selection of answer)

![Diagram of the word recall process](image)

Figure 5. Approach to the passages test. (Adapted from Feldman and Stemmer, 1987:257)
Word Retrieval and the Vocabulary Abilities Used

As stated earlier, this was an exploratory study designed to elicit test-takers' verbal reports on their thinking processes while performing on vocabulary tests. My main purpose was to identify the aspects of vocabulary ability involved in performing on two types of vocabulary tests. In this section I present the taxonomy used, explain how it emerged from the data gathered, and how it reflected the aspects of vocabulary ability identified in the literature (e.g., Chapelle, 1994). Illustrative examples taken from the participants's verbal reports are then presented and discussed.

Taxonomy

Vocabulary ability is defined in chapter two according to the interactionalist definition of language ability, which attributes performance to the learners' characteristics (including knowledge, processes and strategies), to contextual factors, as well as to interactions among them (Bachman, 1990; Chapelle, 1994). Vocabulary ability is then defined in terms of the context of vocabulary use, vocabulary knowledge and processes, and the metacognitive strategies required for vocabulary use in context. In analyzing the data, several of these aspects were identified.

The data gathered provided a vast array of aspects of vocabulary ability and other abilities. While identifying the steps followed by the participants in answering the tests items, I noticed that they had used their knowledge of word characteristics (e.g., meaning features, word formation and derivation), their first language knowledge (e.g. cognates, translation), and contextual clues (e.g., clues in the
sentences containing the target words). I decided then to use those broad categories as a starting point and divided the data under three categories: knowledge of word characteristics, knowledge of the context, and knowledge of the first language. Some abilities and strategies did not fit into these categories and several categories would have to be created to accommodate them. Since they occurred sparingly, I decided to group them under "other abilities".

New groupings emerged from each category, prompting me to refine my classification by subdividing the categories. The resulting taxonomy follows:

A. Knowledge of word characteristics
   1. Semantic characteristics
      a. Meaning features
      b. Polysemy
      c. Synonymy
      d. Antonymy
   2. Morphemic/Syntactic characteristics
      a. Word formation and derivation
      b. Part of speech
   3. Phonemic/Graphemic characteristics
      a. word form
      b. synforms
      c. initial sounds
      d. pronunciation

B. Knowledge of the context
   1. Contextual clues
      a. global clues
      b. local clues
      c. syntactic clues
C. Knowledge of the first language
   1. translation
   2. cognates
   3. false cognates

D. Other abilities
   1. cultural/world knowledge
   2. imagery
   3. mimicry and sound

Now I turn to a discussion of examples illustrating the categories above.

The abilities used

The examples are grouped according to the categories or subcategories in which they were classified; they are shown in tables together with the functions they served in the tests. In the discussion I relate the examples (and their uses) to the literature on vocabulary learning, as well as to some characteristics of the participants.

Interestingly, similar aspects of vocabulary ability were elicited by the two types of tests and were used for similar purposes, so I decided to present and discuss them together. The functions presented in the tables were the ones accomplished through the examples presented. It does not mean, however, that an aspect of vocabulary knowledge was used to accomplish the same purposes by all three participants in every occurrence.

Although the majority of the examples presented led to successful performance on the items that elicited them, several of them did not necessarily lead to correct answers. Other factors, such as proficiency level, may have interfered.
(These will be discussed in the section on individual differences).

**Use of knowledge of word characteristics**

The category 'knowledge of word characteristics' was subdivided into semantic, morphemic, syntactic, phonemic, and graphemic characteristics. Each will be considered in turn.

**Semantic characteristics (Table 3).** I will first focus on the semantic characteristics of words the participants used in their attempts to answer the test items. Knowledge of meaning features was necessary to decide what word or expression was more appropriate in a given context. Knowledge of different meanings (polysemy), equivalent meanings (synonymy), and opposite meanings (antonymy) was also necessary for answering some items correctly.

**Meaning features.** Words have different semantic features and knowledge of some of these features plays an important part in choosing words with equivalent meanings. For example, participants A and B chose *breaking* as the best substitute for *smashing* because they thought it had "a stronger sense" than the other alternatives (cutting, chopping, slicing). Participant C in evaluating *impressive* as substitute for *remarkable* focused on a feature she called "strong", present in both words.

**Polysemy.** Knowledge that one word form can take on different meanings was also an aspect of vocabulary ability showed by participants A and B. 'A' recalled one meaning of the word "submarine" and, though not actually saying it, recalled that the same form had another common meaning. 'B' first retrieved one meaning ('ship'), then changed her mind ('underwater of course') after reading the alternatives.
Synonymy. Humans link words together in their minds with words apparently organized in semantic fields (Aitchison, 1987:191). Semantic fields can be formed with words of equivalent meanings (synonyms) or with words of opposite meanings (antonyms). Therefore synonyms and antonyms are likely to be closely linked in our mental lexicon.

Part of the data presented in Table 3 appears to be an indication that we indeed establish certain links between

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>meaning features</td>
<td>&quot;here the stronger sense must be breaking&quot;</td>
<td>selecting a word among related words</td>
</tr>
<tr>
<td>polysemy</td>
<td>&quot;it has the meaning of under water it can also be...&quot;</td>
<td>previewing (before looking at alternatives)</td>
</tr>
<tr>
<td>synonymy</td>
<td>&quot;occasionally and sometimes has the same meaning&quot;</td>
<td>checking answer</td>
</tr>
<tr>
<td></td>
<td>&quot;bright luminous brilliant radiant&quot;</td>
<td>retrieving a word (radiance)</td>
</tr>
<tr>
<td>antonymy</td>
<td>&quot;altruistic no it's the opposite&quot;</td>
<td>rejecting an alternative</td>
</tr>
</tbody>
</table>

words, which partially accounts for our ability to retrieve them fast (Aitchison, 1987; Chapelle, 1994). Evidence from Participant B’s report is a good example of that. She uttered several related words ("bright, luminous, brilliant, radiant") in trying to retrieve a synonym for the word "brightness".

Antonymy. The relationship of opposition in meaning was used successfully by participant C; she ruled out an
answer containing a word with opposite meaning of the target word. Participants A and B, however, failed in getting the correct answer to a question and chose the alternative containing the word or expression with the opposite meaning: "kept away from" for "hooked up"; "out of reach" for "tangible". Their choosing an antonym can also be interpreted as an indication that words with opposite meanings may also be linked in our mental lexicons.

Morphemic and syntactic characteristics (Table 4). This section deals with the participants' use of their knowledge of morphological and syntactic characteristics of words to accomplish different functions in the two tests.

Word parts. The participants showed that they were aware that certain words are formed by meaningful parts (morphemes) and used this knowledge to analyze those parts in order to interpret their meanings. In analyzing the word "remarkable", one of the participants focused on its root - "mark" - as containing the basic meaning: "remarkable is something that marks." After looking at the alternative containing "impressive", she related it to "something that marks" and concluded they were equivalent ("something that leaves an impression").

Participant B used her knowledge of the word "aid" and the prefix "un-" in analyzing the word "unaided" ("with no help"). Participant C, in checking her answer related the word "stronger" to the word "strength" in the passage, evidencing her knowledge of word derivation processes.

Part of speech. This illustrates the learners' use of their knowledge of a word's part of speech to answer a question. Only one of the participants made such attempts (participant A).
While reading the sentence containing the word "emergence", participant A paraphrased it as "something that appeared" then decided that the synonym would also have to be a noun, "appearance". At first I interpreted that as an attempt to find a clue to the correct answer, that is, the correct alternative would contain a noun. Later I considered the possibility that the participant was elaborating on her thought. In other words, she was conforming it to grammatical rules: a substitute for a noun should be another noun. In another instance it was clear that this participant was using part of speech as a clue: she read a sentence with a blank to be completed and said "an adjective is missing here".

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>word parts</td>
<td>&quot;remarkable is something that marks&quot;</td>
<td>uncovering word meaning by analysing its parts</td>
</tr>
<tr>
<td>formation / derivation</td>
<td>&quot;unaided with no help&quot;</td>
<td>uncovering word meaning by analyzing its parts</td>
</tr>
<tr>
<td>part of speech</td>
<td>&quot;emerged is something that appeared so here it’s a noun-appearance&quot;</td>
<td>using part of speech as clue</td>
</tr>
</tbody>
</table>

Phonemic and graphemic characteristics (Table 5). The participants' reports contained evidence that graphemic and phonemic features of words were aspects of word knowledge necessary to answer or to interpret some test items correctly. Absence or incomplete knowledge of these characteristics hindered one participant's rate of progress through test items.

Word form. This illustrates the common phenomenon of having a concept or even an image of a word in our mind, but being unable to "clothe it in sounds" or spelling
(Aitchison, 1987:191). In one instance, participant B was unable to come up with the word form for the concept she had in mind. This seems to be a situation particularly common to second language speakers due to one of the following: either the form of that word was not known or could not be retrieved at that moment.

**Synforms.** Synforms, confusion between two words with similar forms (Laufer, 1990), is the result of incomplete knowledge of words. Gaps in the learners' knowledge of a word lead to that confusion. Participant B was able to distinguish between the similar forms of "harm" and "ham", one of which may be more familiar to her. She confused, however, the words "emergence" and "emergency", interpreting "emergence" as "urgency". The word "emergency" seems to be more common in both Portuguese and English. In Portuguese, however, there is a single form, "emergencia", for the two meanings. In addition to that, Portuguese speakers tend to pronounce the two forms in the same way. In other words, it seems that first language interference combined with incomplete knowledge of English words caused the participant to confuse two similar forms.

**Initial sounds.** These refer to the retrieval of words that have similar initial sounds (usually the first three phonemes) to aid the retrieval of the target word. All three participants were found to use words with similar sounding beginnings. Some of them were also semantically related. These words were from both Portuguese and English.

It is claimed that words in our mental lexicon can be accessed through one of two channels, namely, semantic/syntactic or phonological (e.g., Channell, 1988). Within the phonological organization of the mental lexicon, words with similar initial and final sounds seem to be closely
linked (Aitchison, 1987). Word initial and final sounds have been found to be prominent features in identifying spoken words and in retrieving words, as well (see Aitchison, 1987 for a review of several studies).

Pronunciation. This refers to a single instance in which participant B had problems in pronouncing the word "depicts", which slowed down her work on that item. She tried several ways of pronouncing it until she was satisfied with a way.

The participants were thinking aloud and, probably as a consequence, tended to read parts of the alternatives loudly. We can only speculate whether or not being unable to pronounce a word would have hindered her performance had she not been asked to think aloud while taking the test.

Table 5. Phonemic/graphemic characteristics.

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>word form</td>
<td>&quot;I know it but I don't know how to say it&quot;</td>
<td>(incomplete knowledge of a word)</td>
</tr>
<tr>
<td>synforms</td>
<td>&quot;harm, it's not ham&quot;</td>
<td>avoiding confusion between two similar forms</td>
</tr>
<tr>
<td>emergence/emergency</td>
<td></td>
<td>(failure in avoiding synforms)</td>
</tr>
<tr>
<td>initial sounds</td>
<td>&quot;entail inteirar interaction involvement&quot;[In]</td>
<td>retrieving a word (involve)</td>
</tr>
<tr>
<td>pronunciation</td>
<td>&quot;what's this word I don't know how to say it&quot;</td>
<td>(pronunc. problem)</td>
</tr>
</tbody>
</table>
Use of Context

Contextual clues (Table 6). The words tested in both types of tests were presented in the context of a sentence or in the context of a reading passage. Therefore it was expected that the participants would use that context together with their knowledge of other aspects of words. This section presents an analysis of samples of the contextual clues used by the three participants while working on the items from the two tests.

Global clues. Global clues refer to the use of the whole context in order to answer a question. It was found that the participants used global clues, i.e., they were able to infer meaning of words based on the global context. The words 'grinning' and 'chuckling' were unknown by the three participants but they were able to figure out their meanings based on the passage topic - the effects of laughter on the human body. The use of such clues was possible only in the passages test.

Local clues. Local clues refer to those instances in which the participants focused on the context surrounding the target word; it included the whole sentence, parts of it, or another word in the sentence. The sentence test due to its format only allowed these types of clues.

Local clues appeared to be the most frequently used in the passages test as well. After reading the passage and a question, the participants located the word in the text, read the surrounding context, then tried to answer it. In some instances it was possible to identify the words or expressions that led to an answer. In many cases, however, the participants read silently and did not mention any portions of the sentence when answering.
The theory that claims that the true meaningful unit is a phrase or sentence instead of longer pieces of discourse (e.g., Steinberg, 1978) seems to be partially supported by the participants' use of local clues. Only after considering the sentence in which the target word is located did the participants try to use global clues. Participant A's report illustrates this point: "Here it's not possible to know the meaning of this word... but because of the sense of the text it must be something to do with laughter".

**Syntactic clues.** Only one of the participants explicitly focused on syntactic clues present in the target word sentence. In the sentence "Some critics have praised J. Michener's epic novels for their facts but deplored their characterizations" participant C interpreted 'deplored' as somehow opposed to 'praised' due to the presence of the adversative conjunction 'but'.

Table 6. Contextual clues.

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>global clues</strong></td>
<td>&quot;but because of the sense of the text it must be something to do with laughter&quot;</td>
<td>guessing the meaning of unknown words</td>
</tr>
<tr>
<td><strong>local clues</strong></td>
<td>&quot;in the beginning of spring so...&quot;</td>
<td>using parts of the sentence as clue to the target word meaning</td>
</tr>
<tr>
<td></td>
<td>&quot;distinguished because of controversial&quot;</td>
<td>evaluating answer using another word in the target word sentence</td>
</tr>
<tr>
<td><strong>syntactic clues</strong></td>
<td>(but) &quot;it must be something opposite to praised&quot;</td>
<td>using syntactic knowledge to guess word meaning</td>
</tr>
</tbody>
</table>
Use of first language knowledge (Table 7)

The participants used the knowledge of their first language, Portuguese, to understand word meanings and/or to recall words with equivalent meanings. These goals were sometimes obtained through the translation of English words into Portuguese or through the use of their knowledge of cognates.

**Translation.** Translation was widely used during the tests. The frequent use of translation can be partially explained by the fact that the subjects were allowed to use their first language in their verbal reports. Two participants used Portuguese almost exclusively to report on their thoughts. These were also the ones who used translation most often. The third participant chose English as the language to report on her thoughts and, perhaps as a consequence, translated considerably less.

The questions and alternatives were the parts of the tests that were translated. During the first reading of the passages none of the participants attempted to translate parts of it. However, when locating the target word in the passage, the context surrounding the word (e.g., within the sentence limits) was usually translated.

It was difficult to determine to what extent translation was simply a consequence of the methods and instruments used in collecting the data. However, some strategic uses of translation were identified. These included (a) using translation to evaluate and/or confirm an answer; (b) using translation to better understand the meaning of a word; and (c) using translation to better interpret the context surrounding the word. The use of cognates also implies some degree of translation since the similarity of the target language word prompts the equivalent in the first language.
Participant A tried to use translation to answer a question and concluded it was not the appropriate strategy. In another instance participant A used translation to check her answer; after considering 'spell' as the correct answer she translated it and decided it could not be correct. Participant C translated the word she chose as the correct answer in order to check/confirm her choice: "volition" (target word), "will" (answer chosen), "vontade" (translation).

In some instances, especially in the sentence test, immediate translation seemed to signal automatic retrieval of words. The participants would read the English word and immediately translate it, sometimes only pronouncing the Portuguese word loudly. In those instances, the participants usually moved on immediately to the alternatives.

**Cognates and false cognates.** The participants also showed an awareness that Portuguese and English are related languages, containing several words with similar spelling and meaning (cognates) and used this knowledge as a tool in selecting their answers.

In one example a participant related "dual" to the Portuguese "dois" (two). The same participant also showed an awareness of false cognates (e.g., words with similar spelling but with different meaning or uses in the two languages). "Sensitive" and "sensitivo" are false cognates, as well as "sensible" and "sensivel". (The correct translation of "sensitive" is "sensivel", while "sensible" corresponds to "sensato" and "sensitivo" means extra-sensory perception).

Participant B used her first language knowledge of the word "estritamente" to guess the meaning of the English word "strictly". This illustrates her awareness that both
Table 7. L1 knowledge.

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>translation</td>
<td>&quot;here it's not possible to know the meaning of this word with translation&quot;</td>
<td>using translation to uncover word meaning</td>
</tr>
<tr>
<td></td>
<td>&quot;spell soletrar can't be&quot;</td>
<td>using translation to evaluate answer</td>
</tr>
<tr>
<td>cognates</td>
<td>&quot;dual in the sense that it has two&quot; (dual-dois)</td>
<td>using cognate knowledge to interpret word meaning</td>
</tr>
<tr>
<td></td>
<td>&quot;never seen in English but if it's like Portuguese...&quot;</td>
<td>relating L1 to L2 to guess word meaning</td>
</tr>
<tr>
<td>false cognates</td>
<td>translated sensitive as sensitivo, changed quickly to sensivel</td>
<td>avoiding confusion among false cognates</td>
</tr>
</tbody>
</table>

languages are related and she could make an educated guess based on her first language.

Use of other abilities (Table 8)

In categorizing the data, I identified three broad categories, namely, use of knowledge of word characteristics, use of context, and use of first language knowledge. Other types of knowledge and strategies were also used by the participants. These include world knowledge, images, and physical representations as used by the participants in retrieving the meaning of some words.

Cultural/world knowledge. Participant B used her knowledge that magic is commonly used for entertainment purposes in our culture. She also evoked her knowledge of the newness of a newborn's environment and interpreted it as a potentially painful situation.
Use of images. Two of the participants seemed to use images of the target word or of its context. "Shelter" was described as a house. In order to get a synonym for "pick up" in the sentence "The Red River gains its name from the color of the soil it picks up as it flows through rich prairie land", one participant focused on the image of the water running and taking part of the soil with it.

Table 8. Other abilities.

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>cultural/world knowledge</td>
<td>&quot;well... magic amuses people&quot;</td>
<td>using cultural knowledge in selecting an answer</td>
</tr>
<tr>
<td></td>
<td>&quot;it's a little of painful because everything is new&quot; (newborn's cry)</td>
<td>using the word context of use in selecting an answer</td>
</tr>
<tr>
<td>use of images</td>
<td>&quot;shelter is as if it were a house&quot;</td>
<td>&quot;visualizing&quot; functions of the word</td>
</tr>
<tr>
<td></td>
<td>&quot;as the water runs it takes...&quot;</td>
<td>interpreting the context adjacent to the target word</td>
</tr>
<tr>
<td>mimicry/sounds</td>
<td>mimicked cut, chop, and slice with hand gestures and sounds</td>
<td>clarifying different aspects of word meaning or uses</td>
</tr>
</tbody>
</table>

Mimicry/sounds. The participants also use mimicry and sounds to illustrate some words. Words such as "chop", "slice", and "cut" were mimicked with hand gestures and sounds.

To summarize, this study was based on the assumption that product data from tests (e.g., test scores) do not provide enough evidence on how test-takers use their vocabulary (and other) abilities to answer test items. It was found that the
participants in this study strategically used several different aspects of their vocabulary ability and other abilities, as well, to retrieve the tested words.

Test Format, Individual Differences, and the Abilities Used

Before proceeding to a separate discussion of the effects of test format, topic familiarity, and individual differences on the vocabulary abilities used, I will quantify the data and discuss the frequency of the abilities used in relation to the types of tests and the participants' characteristics.

The data is quantified and presented in tables containing their occurrence in absolute numbers and percentages. The numbers are then discussed in relation to test format and participants' characteristics.

The number of abilities presented refer to those instances that were verbally reported and, therefore, may not represent the actual numbers of instances in which abilities were drawn upon. However, I present the number of occurrences reported to provide a clearer understanding regarding the relative frequency of the abilities reported in this study.

Frequency of the abilities/strategies used

Tables 9-12 present all of the instances of vocabulary ability and other abilities found in the participants' verbal reports. The numbers in those tables refer to the different abilities used by the three participants in the two types of tests.
Table 9 presents the occurrences of the different characteristics of word knowledge as used by the three participants in the two tests. Table 10 presents the participants' use of different types of contextual clues.

<table>
<thead>
<tr>
<th>Word Characteristics</th>
<th>Number of Occurrences</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sent.</td>
<td>Pas.</td>
<td>Sent.</td>
<td>Pas.</td>
</tr>
<tr>
<td>meaning features</td>
<td></td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>polysemy</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>synonymy</td>
<td></td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>antonymy</td>
<td></td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>word formation / derivation</td>
<td></td>
<td>-</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>part of speech</td>
<td></td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>word form</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>synforms</td>
<td></td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>initial sounds</td>
<td></td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>pronunciation</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 10. Knowledge of the context.

<table>
<thead>
<tr>
<th>Contextual Clues</th>
<th>Number of Occurrences</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>Totals</td>
</tr>
<tr>
<td></td>
<td>Sent.</td>
<td>Pas.</td>
<td>ent.</td>
<td>Pas.</td>
<td>Sent.</td>
<td>Pas.</td>
<td></td>
</tr>
<tr>
<td>global clues</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>local clues</td>
<td>-</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>syntactic clues</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Totals</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 11. First language knowledge.

<table>
<thead>
<tr>
<th>First language knowledge</th>
<th>Number of Occurrences</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>Totals</td>
</tr>
<tr>
<td>translation*</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>cognates</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>false cognates</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>43</td>
</tr>
</tbody>
</table>

The numbers representing the use of translation by the three participants (Table 11) refer to those instances in which translation was used strategically. Participants A and B used translation in virtually every test item, but in many occasions it was not possible to identify any strategic use.

Table 12 presents the number of occurrences of the other abilities the participants used in retrieving the tested words.
Table 12. Other abilities.

<table>
<thead>
<tr>
<th>Other Abilities</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>world knowledge</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>use of images</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>mimic/sounds</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Totals</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

Table 13. Total number of abilities/strategies used.

<table>
<thead>
<tr>
<th>Participants</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Sentence</td>
<td>18</td>
<td>24</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Passag.</td>
<td>26</td>
<td>23</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>44</td>
<td>47</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

Table 13 presents the total numbers of the abilities and strategies used by each participant in the two tests.

Examining Table 13, we can see that the total number of abilities and strategies used by participants A and B do not vary greatly. Participant C, however, reported using fewer abilities and strategies than the other two. This point will be further explored in the next paragraphs and in the section on individual differences.

Now I turn to examine the percentages of the total abilities reported for each of the four aspects of vocabulary ability found in the three participants' reports in each test. Table 14 contains the percentage of use of the four categories mentioned.
In the following paragraphs I will comment on the percentages presented in Table 14 as they refer to each of the four categories identified.

**Knowledge of word characteristics.** Participants A and B used their knowledge of word characteristics (e.g., semantic, syntactic, morphemic, phonemic, and graphemic) to approximately the same degree in both tests. Participant C, however, used it to a larger extent in the passages test.

**Knowledge of context.** Participants A and B made a larger use of contextual clues in the passages test. This was expected since the tested words were presented in the context of longer passages and supposedly would provide a larger variety of clues (e.g., global clues). Participant C, on the other hand, used contextual clues to a greater extent on the sentence test, which accounted for a large percentage of the total of the abilities/strategies she used.

Participant C differs from the others in reporting the use of more contextual clues provided in the sentences in the Table 14. Percentage of use.

<table>
<thead>
<tr>
<th>Abilities Used</th>
<th>Percentage of Use</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sentence Passag.</td>
<td>Sentence Passag.</td>
<td>Sentence Passag.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of word characteristics</td>
<td>28.0</td>
<td>34.6</td>
<td>41.7</td>
<td>39.2</td>
<td>9.1</td>
</tr>
<tr>
<td>Knowledge of the context</td>
<td>5.5</td>
<td>19.2</td>
<td>8.3</td>
<td>17.3</td>
<td>36.4</td>
</tr>
<tr>
<td>Knowledge of first language</td>
<td>55.5</td>
<td>34.6</td>
<td>41.7</td>
<td>30.5</td>
<td>36.4</td>
</tr>
<tr>
<td>Other abilities</td>
<td>11.0</td>
<td>11.6</td>
<td>8.3</td>
<td>13.0</td>
<td>18.1</td>
</tr>
</tbody>
</table>
sentence test. This does not necessarily mean that she used the context to a smaller degree in the passages test.

Perhaps she simply did not verbally report doing so because it was somehow automatic for her. During the experiment she would locate the target word in the passages before attempting to answer the corresponding question. Sometimes she would check her answer against the text. I observed that there was no pause in her reading the text (to locate the word) and working on the test items. This was easily observable because she read both the text and the questions loudly. In short, though relying on the context to answer the questions, participant C did not frequently report on the steps she followed nor the clues she probably used.

**Knowledge of the first language.** The use of first language knowledge accounted for a high percentage of the abilities the three participants used in both tests. The use of their knowledge of cognates was not reported on as frequently as we might expect considering that English and Portuguese are related languages. However, some instances in which they translated sentences or expressions containing cognates could also be interpreted as revealing their knowledge of those related words. Since translation was frequently used, it was difficult to distinguish the occasions in which they were particularly focusing on cognates from those in which the cognates were translated simply because they were part of the sentence or expression being translated.

Translation was largely used by participants A and B. Although several instances in which they used translation strategically were identified, its frequent use may be considered a consequence of thinking aloud in Portuguese. Both participants chose their first language to report on their thoughts. Participant C used English and her use of
translation was predominantly strategic (e.g., to check or confirm an answer).

Other abilities. This category included the use of world/cultural knowledge, the use of images, and the use of mimic/sounds to interpret words and/or their surrounding contexts. Participants A and B’s percentage of use of those strategies did not vary greatly from one test type to the other. Participant C’s accounted for a higher percentage of the total of her abilities/strategies in the sentence test. She did not, however, report on any such occurrence in the passages test.

As we examine the data, it becomes more apparent that the reports of participants A and B have several points in common, whereas participant C seems to differ from them in both the amount of abilities and strategies reported and the uses to which she put them. These differences will be further discussed in the section on individual differences.

Test format and the abilities elicited

The participants were more familiar with the format of the sentence test, but admitted having taken tests similar to my passages test in their native language. They did not mention any particular problems with the test format during the warmup session nor afterwards. They did express their opinion, though, that they preferred the typical sentence test because it takes less time to finish.

Although none of the participants commented on one test type being more demanding than the other, it was found that all three had lower scores on the passages test, and B scored much lower on that test (see Table 16). Presenting words in a larger context may have the advantage of providing more
contextual clues, but it may be more demanding to some learners, requiring additional language skills and thus not allowing them to use those clues effectively. This may have been the case with participant B who was the least proficient. Another explanation would be that the words tested in the passages were more difficult than those in the sentence test. There is no guarantee that the words tested in each test were of equivalent difficulty, since no attempt was made to select words of equal level of difficulty for the two tests.

Interestingly, the types of abilities and strategies the three participants reported in both tests were similar and were used to accomplish similar goals. The context richer test, however, provided the participants with global clues that made it possible for them to guess the meaning of three unknown words. The verb "mushroom" (second passage) was understood due to clues in the paragraph containing it. The words "grinning" and "chuckling", also unknown, were understood because the participants evoked their understanding of the passage as a whole (first passage).

However, when the data were quantified, it showed that some abilities/strategies were more frequently used in one type of test than the other. This use varied among the three subjects, though subjects A and B showed similar tendencies (cf. Table 14). Table 15 presents the total percentage of use of the four categories in each test.

The percentages in Table 15, however, were greatly affected by C’s use of abilities and strategies. As shown in Table 13, she reported fewer abilities and strategies than the others. However, when we transform her numbers into percentages and they are added to the others, they alter the whole picture, and the totals in each test no longer reflect the similarity between participants A and B. Therefore, in
interpreting the percentages presented in Table 15, I will refer back to Table 14, which presents the total percentage of use of the four categories of abilities by each participant in the two tests.

Table 15. Total percentage of use of the four categories of abilities.

<table>
<thead>
<tr>
<th>Abilities used</th>
<th>Percentage of use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
</tr>
<tr>
<td>knowledge of word characteristics</td>
<td>26.3</td>
</tr>
<tr>
<td>knowledge of the context</td>
<td>16.7</td>
</tr>
<tr>
<td>first language knowledge</td>
<td>44.5</td>
</tr>
<tr>
<td>other abilities</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Participants A and B used their knowledge of word characteristics to roughly the same degree in both tests. Participant C used that knowledge to a much larger extent in the passages test (cf. Table 14). When their totals were added and averaged, the use of word characteristics became higher in the passages tests due to C's high percentage of use.

Regarding knowledge of the context, both participants A and B used contextual clues to a larger extent in the passages text (cf. Table 14). Participant C, however, did not show such variation and used more contextual clues in the sentence test. Again when the three participants' totals were added and averaged, the tendency showed by A and B was no longer apparent.

The use of first language knowledge was the only category that maintained the same tendency in Table 15 as that observed in Table 14 where the participants' totals in each test were considered separately.
Regarding the use of world/cultural knowledge, images, and mimicry/sounds (referred to as 'other abilities') the tendency observed in A and B's individual totals (cf. Table 14) was lost when C's was added to them.

In conclusion, I believe no definite claims can be made regarding the effects of test format on the aspects of vocabulary ability used in the two types of tests. Although a quantitative difference was observed, no qualitative differences could be identified; the participants used similar abilities and strategies to accomplish similar goals in both tests. The numerical difference may be caused by other factors not directly related to test format. Learners may have a preference for given strategies or some strategies may be more used depending on the learners' level of proficiency.

**Individual differences and the abilities used**

I will now compare the three participants' performance and relate it to the number of abilities/strategies reported (see Table 16).

The most successful participant, C, scored highest on both tests, spent the least time (50 minutes), and used more English than did A and B (Table 16).

Some of the characteristics presented in Table 16 link participants A and B together. Both used their first language to report on their thoughts, reported on approximately the same number of abilities/strategies, and spent almost the same amount of time in answering the test items. In terms of scores, however, they differ, especially in the passages test.

Participant B's score in the passages test is closer to C's. Participant C differs from the others in that she used
Table 16. Comparison of the participants' performance.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Tests</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
<td>80%</td>
<td>76%</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>Passages</td>
<td>74%</td>
<td>52%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time spent</td>
<td>01:30 min.</td>
<td>01:20 min.</td>
<td>00:50 min.</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>Portuguese</td>
<td>Portuguese</td>
<td>English</td>
</tr>
<tr>
<td></td>
<td>Abilit./strateg.</td>
<td>18</td>
<td>24</td>
<td>11</td>
</tr>
</tbody>
</table>

English to verbalize her thoughts, reported on fewer abilities and strategies, and spent less time to complete both tests. Yet she had the highest scores.

The fact that participant C reported on fewer abilities and strategies, yet obtaining the highest scores, may be explained in light of the literature on vocabulary strategies (e.g., Chapelle, 1994) and the work on verbal reports by Ericsson and Simon (1984, 1987). The vocabulary knowledge and strategies used by a nonnative speaker tend to become more automatic the more proficient he/she becomes. Therefore these are not consciously attended to when a proficient learner works on a test, which makes them unavailable to be verbally reported (Ericsson and Simon, 1987). This appears to have been the case with participant C; due to her proficiency she did not need to spend the same amount of time as the others in processing the information in the tests.

Participants A and B spent approximately the same amount of time on the tests and reported on using roughly the same number of abilities and strategies. This seem to have helped A in obtaining higher scores. Participant B, however, did not
profit from her use of abilities to the same degree as A, probably due to her lower proficiency.

In previous sections I have compared the participants’ performance and use of abilities and strategies. In some occasions individual differences, particularly those related to level of proficiency, seemed to provide plausible explanations for differences observed among the three participants’ performance.

The most proficient participant, C, differed from both A and B in terms of the number of strategies and abilities used, language used for verbalizing her thoughts, and the total scores obtained. However, the strategies she used (except for syntactic clues) were of the same type as those used by the others. The main difference was in number of occurrences. Highly proficient learners are believed to use their abilities and strategies automatically (e.g., Chapelle, 1994) which cause them not to be readily available for verbalization through the think-aloud protocol, that is, only information being attended to is reportable (Ericsson and Simon, 1987).

Participant B, the least proficient, differed from the others in that she appeared to favor strategies at the word level (see Table 11), that is, she would first try to analyze what she believed to be their meaningful parts (e.g. pin point, for “pinpoint”; far fetched, for “farfetched”). She was the only one to report paying particular attention to word form (both successfully and unsuccessfully) as in “it’s harm not ham”, having problems with pronunciation, and the appropriate form of words (e.g., “I know it but I don’t know how to say”).

Participant A was the most “talkative”; at first I thought she had reported using more strategies than the others, but after looking more carefully at her data I realized that she
"expressed" her thoughts with more words, using expressions such as "it's something that...", "I think it's...", "it can't be". She used approximately the same number of strategies as B, but she was more successful. She spent more time on the tests than C did and was almost as successful. A's performance seems to be an indication of the compensatory use of strategies; what she may have lacked in knowledge was compensated for with the use of a variety of strategies.

In conclusion, though the participants formed a relatively homogeneous group, it was still possible to observe individual differences at work. I have focused on the ones I believed were caused by proficiency levels for two reasons: (1) they were more prominent and more readily identifiable and (2) they seemed to have occurred more frequently.

Regarding the relationship between individual differences in the vocabulary abilities and strategies used and the format of the vocabulary tests, I believe it is still early to draw any conclusions. It appears that the differences I have pointed out are somewhat too common in general language use to be attributed to particular types of tests.

**Topic Familiarity and the Abilities Elicited**

In choosing the topic of the three passages used, I selected one passage on a topic that would be of particular interest to one participant according to their background. Participant B referred to one of the passages on the following terms: "I liked these texts; this one has to do with my major, but there were lots of words I didn't know". Participant A commented that she especially liked one of the passages, but that was not the one related to her field of
study. Participant C did not make any comments about the passages.

The scores in each passage did not show any tendency that would indicate that a participant scored higher on the passage related to her field of study. The scores are presented in Table 17.

Table 17. Scores on the passages test.

<table>
<thead>
<tr>
<th>Participants/Related Passage</th>
<th>Passage 1</th>
<th>Passage 2</th>
<th>Passage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A / Passage 2</td>
<td>100%</td>
<td>45%</td>
<td>70%</td>
</tr>
<tr>
<td>B / Passage 3</td>
<td>80%</td>
<td>45%</td>
<td>40%</td>
</tr>
<tr>
<td>C / Passage 1</td>
<td>90%</td>
<td>55%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Participants A and B had their lowest scores on the passage that presented the topic on their area. Participant C had a high score on her passage, but she did equally well in another passage. One of the passages (passage B - first language acquisition) seemed to be difficult for all three participants, whereas passage A (medical experiment) seemed to be the easiest for all of them.

Contrary to B's favorable comments on the topic of the passages, she scored much lower in that test (52%) than she did on the sentence test (76%). It seems that the topic of the passages did not have any particular impact on the participants' performance. All three had their lowest score on the same passage of the test (passage two).

The participants seemed to have focused on the passages as part of a vocabulary test in which they were supposed to know the meaning of certain words. Of course, this experiment was insufficient to allow for claims about the effect of general
familiarity with a topic may have on test-takers' performance on the vocabulary tests used.

To summarize, in this chapter I presented the data obtained and analyzed them both qualitatively and quantitatively. It was found that the participants used a varied number of abilities and strategies to retrieve the tested words. There was no striking differences between the abilities elicited by the two tests (with the exception, of course, that only the passages test allowed for global clues). The effects of topic familiarity on the learners' performance on vocabulary tests were inconclusive.
SUMMARY, CONCLUSIONS, FURTHER CONSIDERATIONS, AND IMPLICATIONS

In this chapter I summarize the study, review its overall conclusions, and examine its methodological, theoretical, and pedagogical implications.

Summary

This study was designed to uncover information about the aspects of vocabulary ability learners use in performing on vocabulary tests. Two different types of vocabulary tests were selected. The main difference between the two tests was in the amount of context in which the words were presented. These tests reflected my interest in finding out more about the effects of context on the abilities, processes, and strategies used in retrieving the tested words. General familiarity with a topic was another aspect of contextual effects also investigated (each of the three passages was related to one of the participants' background).

The key findings of this study are listed below and discussed in the remaining paragraphs of this section.

Key findings

(1) It was found that the participants used a wide variety of strategies based on their vocabulary ability and other abilities, as well, to retrieve the tested words.

(2) The difference in the format of the two tests used (e.g., amount of context) did not elicit substantial differences on the strategies used.
(3) The effect of familiarity with a topic on the performance on the vocabulary items in a passage was inconclusive. It did not seem to have influenced the performance.

(4) Individual differences in the use of strategies were observed even though the participants had similar characteristics.

Regarding vocabulary ability, the participants in this study reported using a wide variety of strategies. These strategies cannot be identified by simply looking at the scores on the tests (Chapelle, 1994). Although in a controlled test we may select items that would tap into certain abilities, we still would not be able to infer how and what aspects of an ability were used successfully or not in answering those items. Using the think aloud protocol, however, made it possible to have a more complete picture of what goes on when learners work on tests, that is, how and how much of the aspects of the ability tested and of other abilities as well they utilize to answer the questions. The information obtained can provide the basis for later designing experimental studies on the effectiveness of particular strategies, for instance.

Individual differences were identified in the uses of strategies (especially on the number used). These appeared to be related to different levels of proficiency. For instance, the participant of highest proficiency reported using fewer strategies. The other two participants reported using comparable number of strategies, but the more proficient of the two was more successful.

Another important aspect of this study was the use of a taxonomy that emerged from the data gathered. Three broad categories were identified relating to the participants' use
of three kinds of knowledge, namely, knowledge of word characteristics, knowledge of the context, and knowledge of first language. Other abilities were used as well (e.g., imagery, mimic, world knowledge); these were included in a fourth category, "other abilities." The taxonomy emerged from the data and reflected several of the aspects of vocabulary ability identified in the literature (e.g., Chapelle, 1994) and found in the participants' reports.

The amount of context (e.g., the main difference in the format of the tests) does not seem to produce different strategies or elicit different abilities. With a few exceptions (e.g. the use of global clues), the participants focused on the context surrounding the words (usually a context no longer than a sentence), which appears to have diminished the difference between the two tests. The passages test was taken first and therefore the participants' approach to it cannot be attributed to influence from the approach to the sentence test. In other words, the context used to answer the majority of the questions in both tests was at the sentence level, which may have contributed to the two test formats not eliciting substantially different strategies and abilities.

This study was successful in tapping into a variety of abilities and strategies used by the three participants, which confirms the effectiveness of the think aloud protocol (Ericsson and Simon, 1987) in uncovering learners' processes and strategies.

The call to uncover the types of strategies used in tests (Cohen, 1992) and the aspects of vocabulary ability (and other abilities) elicited by vocabulary tests (Chapelle, 1994) is still far from being fully answered, but according to the
results in this study the think aloud protocol appears to tap into the type of information called for.

Implications

In this section I present some methodological, theoretical, and pedagogical implications of this study.

Methodological implications

Case study. The participants in this study formed a small and relatively homogeneous group (see table 2). There are several advantages to this. Small number of subjects allows for in-depth probe of strategies. It was possible for me to examine each item from the two tests after the data was transcribed. Had there been a larger number of participants it would not have been possible to examine the data as carefully. Larger numbers of subjects usually mean researchers are forced to use questionnaires or interviews in which subjects report what they think they do.

A small number of subjects also allows for more control of the variables involved. The participants in this study matched according to gender, age, language and cultural background, academic goals. This allows us to gain a more coherent picture of strategy use for a particular kind of learner. For example, with this group we can observe that they used strategies involving their first language knowledge (e.g., cognates). It was possible to observe differences among the strategies used, for example, one participant favored the use of strategies involving her knowledge of word parts (e.g., breaking words into meaningful parts), whereas another focused more on the context. The differences observed cannot be attributed to any of the variables mentioned, which
narrow them down to perhaps personality factors or learning style preferences.

The think aloud protocol. The think aloud protocol has proven to be a useful tool in eliciting the participants' vocabulary strategies and processes. Procedures suggested in Cohen (1987) and in Ericsson and Simon (1987) were carefully followed making it possible to tap into a variety of uses of the different aspects of vocabulary ability identified in the literature (e.g., Carter, 1988; Chapelle, 1994). Some advantages of this method over approaches such as inventories of learners' strategies using questionnaires (O'Malley and Chamot, 1990; Oxford, 1990) can be identified. The participants report what they are thinking while doing a task, whereas in answering questions from a questionnaire they report what they believe they normally do. Factors such as what they think is advisable to do, but do not necessarily do, can be reported as a procedure normally used (Cohen and Hosenfeld, 1981).

Thinking aloud in a related language may have inflated the occurrence of translation. This is a point worthy of further investigation; would the participants have used translation to the same degree if they were asked to report their thoughts in English? Would that cause any changes on the other types of strategies they would report? This could be investigated by having participants think-aloud in English while working on similar tests.

Role of first language. The participants used their knowledge of the first language to answer several items on the tests. This was expected since Portuguese and English are related languages and usually there is a considerable number of cognates especially in academic texts. However the extent
to which the participants normally use cognate knowledge and translation may not be actually reflected.

The effects of thinking aloud in the learners native language or in their second language needs more investigation. The participants (A and B, in particular) were switching back and forth from one language to the other and may have used translation to a larger extent than they typically do. When asked about her use of translation in regular tests, one of the participants said she usually uses translation ("the Portuguese word just pops up in my mind") and that she used it the same way in my tests. Of course, her claim needs to be further investigated. This could be done by having the same participants perform think-alouds in parallel tests using their first language in one test and their second language in the other.

An interesting aspect related to the use of translation was observed particularly in C's report. She used translation as "a last resource" when the context and other strategies would not work. She also used it to check and confirm her answers, clearly using it strategically. Again this may be explained by the fact that Portuguese and English are related. However, C did not use Portuguese in her report and translated to a considerably smaller degree when compared to the others.

In my experience as a foreign language learner translation was strongly discouraged by teachers and viewed as a negative strategy. Yet, this study suggests that it may not necessarily be a hindrance (e.g., slowing down performance) and some learners may use it successfully as a strategy, especially when dealing with related languages.

The strategic use of translation in retrieving words and checking answers would also be interesting to investigate. The participant who used it to evaluate her answers was highly
proficient and she used it successfully. A study could be
designed to investigate the relationship between translation
and proficiency level, for example. It might not be
appropriate to all levels, and it would probably be more
effective if the learners' first and second languages are
related.

Theoretical implications

Organization of the mental lexicon. Research on the
organization of second language speakers' mental lexicon (see
Channell, 1998 for a review of studies) has pointed to a
phonological arrangement (e.g., Meara, 1984) in low ability
speakers, whereas higher ability speakers are assumed to have
a semantic organization similar to that of native speakers
(Channell, 1988). In this study I found some indication for
both phonological and semantic relationships. While trying to
retrieve a word, the participants sometimes came up with
similar sounding words from both languages. In some instances
these words were also semantically related (forming semantic
fields), though not necessarily equivalent in meaning and/or
use to the target word.

Relationship of strategies and proficiency. Some
indication was found that points to the relationship between
the use of strategies and level of proficiency. The
participant was the highest score reported using fewer
strategies. The other two participants used roughly the same
number of strategies, but the most proficient of the two was
more successful, scoring higher.

It would be interesting to find out more about the
relationship between level of proficiency and the types of
vocabulary strategies learners use. A study similar to this
but including subjects with more marked differences in levels of proficiency (e.g., low, intermediate, advanced levels), could be designed in order to gain a clearer understanding of the effects of the relationship of vocabulary strategies and language proficiency.

Another interesting point for further investigation would be the types of strategies or combination strategies that result in successful performance. In this study there was some indication that using "good" strategies did not necessarily lead to correct answers (e.g., participants A and B used similar strategies but with different degree of success). My hypothesis is that the least proficient participant, B, though using the same kinds of strategies that worked well for A, was unable to carry them successfully due to her lower level of proficiency. Of course, this also needs to be further investigated.

Role of context. The main difference between the tests used in this study was that of the amount of context in which the tested words were presented. Both tests involved word recognition and choosing among words with equivalent meaning. No substantial differences were observed in the types of strategies elicited in both tests.

Perhaps to obtain substantial differences between the abilities and strategies used in two different test formats the tests would have to contain more marked differences. For instance, one test might involve recognition and another production. Since the models for word production and word recognition (Aitchison, 1987) suggest differences in the process of retrieving word for production and for recognition (see chapter two, p. 19), we would predict greater differences in strategies under such circumstances.
Pedagogical implications

Variety of strategies. A large variety of strategies involving different aspects of word knowledge were used by the participants in this study. This seems to point to the importance of exposing students to different techniques for learning vocabulary. Some techniques might be more appropriate to certain kinds of words or suit different learning styles.

In some instances a combination of different strategies may prove more effective and learners need to be alerted to that, that is, flexibility in the use of strategies should be encouraged. Participant A and B, for example, broke the word "farfetched" (far - fetched) and tried to get its meaning through the analysis of the two parts. However, it was not sufficient for them to arrive at the correct answer without also understanding the surrounding context. It appears that relying on a single strategy is not enough to retrieve the correct word in some instances. Knowledge of word characteristics combined with contextual clues seemed to work more efficiently.

A point worthy of further research would be investigate (in a similar study using the think-aloud protocol) the types of strategies or combination of strategies that lead to successful performance. I believe, however, that we first need more evidence on the kinds of strategies elicited by different vocabulary tests.

Translation. This study provided some evidence that translation is not necessarily detrimental as it is often suggested by many foreign language teachers. Teachers of English to speakers of Romance languages, for example, could allow students to use their first language knowledge
strategically instead of treating it as detrimental to the learners' development into a proficient speaker.

**Testing.** Although no substantial differences was found between the strategies used in the two tests in this study, the possibility of having different formats eliciting different strategies cannot be ruled out. As the research on vocabulary strategies used in different vocabulary tests accumulates, it might be possible for language teachers to design tests that would tap into types of strategy consonant to the teaching approach adopted.

To conclude, this study was exploratory; it was designed in response to the call for more research using process-oriented methods to investigate learners' performance on vocabulary tests (Chapelle, 1994; Cohen, 1992). Some of its findings still need to be further investigated, but the methodology used seems to tap into learners' processes and strategies and shows the appropriateness of this approach to uncover the aspects of vocabulary ability used in vocabulary tests.
References


ACKNOWLEDGEMENTS

I wish to first thank Dr. Vann for her expert and friendly guidance and my committee members, Dr. Abraham and Dr. Andre, for their insightful comments and suggestions.

I wish also to acknowledge Dr. Carol Chapelle for her suggestions and for providing me with materials.

My special thanks to three friends who agreed in participating in my experiment.

A big thank you to my husband who took on another job to support me through graduate school. Thanks also for the moral support.

Last but not least, my very special thanks to my son for putting up with a busy mom. And to my big family in Brazil, "obrigada pela torcida."
APPENDIX A
SENTENCE TEST

Directions: In the questions below each sentence has an underlined word or phrase. Below each sentence are four other words or phrases, marked a., b., c., and d. You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase.

1. The number of women working on cattle ranches increases every year.
   a. grows
   b. meets
   c. reorganizes
   d. vacillates

2. A guyot is a submarine mountain that does not reach the surface of the sea.
   a. a tall
   b. an underwater
   c. a huge
   d. a volcanic

3. Because some magic is intended to harm and some is not, distinctions are drawn between various types of magic.
   a. do damage
   b. fool people
   c. create wealth
   d. grant wishes

4. Poultry raising in the United States has become a multibillion-dollar industry.
   a. sport
   b. problem
   c. deficit
   d. business
5. Billboard painting is sometimes described as a dying art.

a. graphically
b. appropriately
c. occasionally
d. regrettably

6. Like many other swamp birds, blue herons use wild grasses for shelter.

a. finding mates
b. laying eggs
c. nourishment
d. cover

7. Revolutionary developments in both observational and theoretical astronomy took place in the 1960s.

a. commenced
b. occurred
c. were disputed
d. were exchanged

8. In the legends of the American West, Paul Bunyan’s most treasured possession was Babe the Blue Ox, whose horns were said to span a distance of 42 ax handles.

a. comical
b. valued
c. expensive
d. colorful

9. A supernova is a star that explodes and then slowly fades to less than its original brightness.

a. size
b. weight
c. radiance
d. importance
10. Over six million citizens of the United States collect benefits from private pension plans each year.

a. unauthorized  
b. confidential  
c. nontransferable  
d. nongovernmental

11. Since the 1950s, literary critics have attempted to answer the question: When did children’s literature first emerge as a distinct literary genre?

a. improve as  
b. appear as  
c. conform to  
d. respond to

12. Granite crystals are large enough to be seen with the naked eye.

a. wary  
b. candid  
c. unaided  
d. discerning

13. Having enlisted in the Marine Corps in August 1918, Opha May Johnson was made a provisional sergeant a month later.

a. considered  
b. joined  
c. written to  
d. risen in

14. College admissions officers use high school grades as one important factor in predicting a student’s success in college.

a. experiment  
b. benefit  
c. topic  
d. consideration
15. Plays that entail direct interaction between actor and audience present no unusual difficulties for actors.

a. advocate  
b. involve  
c. elicit  
d. exaggerate

16. An allergy results when the body reacts adversely to certain substances introduced to it.

a. negatively  
b. spontaneously  
c. purposefully  
d. quickly

17. The Red River gains its name from the color of the soil it picks up as it flows through rich prairie land.

a. nourishes  
b. produces  
c. gathers  
d. reaps

18. There has been no twentieth-century anthropologist more celebrated or more controversial than Margaret Mead.

a. imaginative  
b. hardworking  
c. distinguished  
d. strong-minded

19. The American antelope possesses remarkable powers of sight and can pinpoint potential danger at tremendous distances.

a. spot  
b. relay  
c. attach  
d. sustain
20. Despite notions to the contrary, a great deal of technical writing is at best awkward and at worst actually unclear.

a. denials
b. laws
c. attempts
d. ideas

21. The fragrant scent of the lilac is said to herald the beginning of spring.

a. symbolize
b. hasten
c. exult
d. announce

22. Communication satellites transmit information more reliably than do ordinary shortwave radios.

a. conveniently
b. dependably
c. accessibly
d. concisely

23. The near extinction of the leopard is the direct result of wanton destruction by humans.

a. ongoing
b. senseless
c. ignorant
d. accelerated

24. During the Civil War, the Southern states fought many desperate battles to uphold their asserted right to secede from the Union.

a. improve
b. preserve
c. detect
d. supply
25. The issue of loose construction versus strict construction of the United States Constitution contributed to the emergence of political parties.

   a. joining
   b. urgency
   c. appearance
   d. activity

26. Even animals that live strictly on land are capable of swimming.

   a. exclusively
   b. happily
   c. rigidly
   d. occasionally

27. Some critics have praised James Michener's epic novels for their facts but deplored their characterizations.

   a. emulated
   b. ridiculed
   c. complimented
   d. lamented

28. The upper atmosphere is believed to consist of a number of roughly concentric layers, which include the troposphere and stratosphere.

   a. plainly
   b. repeatedly
   c. changeably
   d. nearly

29. Materials such as clay, wax, glass, and rubber are widely used in industry today because they are malleable.

   a. easy to manufacture
   b. readily available
   c. pliable
   d. buoyant
30. Athletes who compete in the Olympic Games are supposed to be *amateurs*.

a. well trained
b. novices
c. physically fit
d. nonprofessionals
Appendix B
Passages Test

Directions: Read each of the three passages below and answer the set of questions following each passage. The questions focus on the underlined words. The number under the underlined words corresponds to the number of the question focusing on that word.

The Effect of Laughter on the Human Body

An increasing number of scientists have found that anything that gives you pleasure may be more than just fun. Pleasure and humor might help us live longer, fight off illness, and cure drug addiction. “Science has generally neglected pleasure and humor to concentrate on negative things like depression and illness,” said Dr. William Fry, a psychiatry professor at Stanford University Medical Center in California. “But there’s a growing realization of just how powerful pleasure is.”

For example, research shows that smiling and laughing can actually strengthen your immune system, reduce stress and physical pain, and even help cancer patients. One research project at Loma Linda University showed these beneficial effects of laughter on the body. In the study, ten medical students were hooked up to several machines and spent an hour watching a funny videotape of a man smashing watermelons with a hammer.

Researchers were not trying to prove that smashing watermelons was a funny thing to do. They already knew that the volunteers would probably howl with laughter. What the researchers wanted to find out was how grinning, chuckling, and howling affected the volunteers as they watched the video.

The researchers were delighted to discover that small but important changes were taking place in the volunteer’s bloodstream as they giggled or roared with laughter. Their blood showed increased production of white blood cells (the blood cells that defend us against infection and disease). At the same time, a decrease was noticed in the levels of chemicals that stop the immune system from working properly.

Studies such as this seem to show that you should laugh, chortle, and howl as much as possible! You will increase our chances of living a long, healthy life. That’s what one researcher means when he says, “He who laughs, lasts.” For
years we've thought that laughter and other forms of pleasure were good for you. Now it's a physiological reality.


1. The effects of pleasure and humor on the human body have---
   ------------
   a. been fully investigated
   b. never been studied
   c. not received much attention
   d. been considered a priority

2. Which of the following sentences best describes the effect laughter on our immune system?
   a. laughter is harmful to our immune system
   b. laughter controls our immune system
   c. laughter makes our immune system collapse
   d. laughter makes our immune system stronger

3. Smiling and laughing can ----------- pain and stress.
   a. cure
   b. intensify
   c. worsen
   d. diminish

4. Reduced stress and physical pain are --------- effects of smiling and laughing.
   a. advantageous
   b. neutral
   c. unheard
   d. dreadful
5. In the experiment reported on the text, medical students were ________ machines.
   a. cured by
   b. kept away from
   c. instructed by
   d. attached to

6. In the sentence A man smashing watermelons, the word smashing implies that the man was___________.
   a. cutting watermelons
   b. chopping watermelons
   c. breaking watermelons
   d. slicing watermelons

7. Grinning and chuckling refer to ways people___________.
   a. cry
   b. scold
   c. reacts to pain
   d. laugh

8. The scientists were __________ the experiment results.
   a. distressed by
   b. pleased with
   c. biased towards
   d. misled by

9. Some chemicals may stop our immune system from _________.
   a. functioning rightly
   b. shutting down
   c. performing slowly
   d. overworking

10. In the sentence He who laughs, lasts, the word lasts could be replaced by ___________.
    a. lives longer
    b. lives the moment
    c. lives forever
    d. lives comfortably
First Language Acquisition

There is no one who has not at some time witnessed the remarkable ability of children to communicate. As small babies, children babble and coo and cry and vocally or nonvocally send an extraordinary number of messages. As they reach the end of their first year, specific attempts are made to imitate words and speech sounds heard around them, and about this time they utter their first "words." By about age 3, children can comprehend an incredible quantity of linguistic behavior; their speech capacity mushrooms as they become generator of nonstop chattering and incessant conversation.

This fluency continues into school age as children internalize increasingly complex structures, expand their vocabulary, and sharpen communicative skills. At school age children not only learn what to say, but what not to say, as they learn the social functions of their language.

How can we explain this fantastic journey from that first anguish cry at birth to adult competence in a language? From the first word to tens of thousands? It is these sorts of questions that theories of language acquisition attempt to answer.

In principle we can adopt one of two extremist polarized positions in the study of first language acquisition. The extreme behavioristic position would be that children come into the world with a tabula rasa, a clean slate bearing no preconceived notions about the world or about language, and these children are then shaped by their environment. At the other extreme, you would find a position that claims that children come into this world with very specific innate knowledge, knowledge which includes not only general predispositions and tendencies but also knowledge of the nature of language and of the world. Then, through their own volition, they act upon their environment by developing these bodies of knowledge.

Clearly both of the extreme positions are too farfetched for respectable credibility; but they do represent opposites on a continuum with many possible positions in between.

1. According to the text, children's ability to communicate is ________________.
   a. ordinary
   b. impressive
   c. innate
   d. unexpected

2. By the end of their first year, children begin to ______ their first words.
   a. internalize
   b. speak
   c. discover
   d. spell

3. By age three, children's speech capacity ____________ as they become very talkative.
   a. develops gradually
   b. develops slowly
   c. increases rapidly
   d. decreases slowly

4. Chattering is a kind of _______.
   a. talk
   b. babbling
   c. cry
   d. learning

5. A baby's first cry is described in the text as a(n) _______ cry.
   a. happy
   b. angry
   c. painful
   d. loud

6. Two extreme ____________ regarding how children acquire language are presented in the text.
   a. views
   b. biases
   c. assumptions
   d. concerns
7. According to the behavioristic position, children come into the world with a clean slate —— no preconceived notions about language.
   a. understanding
   b. developing
   c. representing
   d. carrying

8. Behaviorists believe children are —— by their environment.
   a. programmed
   b. sharpened
   c. molded
   d. brought about

9. Opposed to the behavioristic position, there is a position that —— children come into this world with knowledge of the nature of language.
   a. reinforces
   b. complains
   c. maintains
   d. ignores

10. Through their own ——, children act upon their environment by developing their innate bodies of knowledge.
    a. will
    b. knowledge
    c. nature
    d. need

11. The author considers both theories of language acquisition too —— to deserve credibility.
    a. elementary
    b. ill-conceived
    c. forced
    d. powerful
Plants as Therapy

Gardening has been used in the treatment of certain emotional and physical problems for nearly two hundred years, but horticultural therapy as we know it today did not come into general practice until the late 1950s.

Rhea R. McCandliss, a horticultural therapist at the Menninger Clinic in Topeka, Kansas, describes a horticultural therapist as "one who uses the knowledge of plants and gardening, greenhouse and floristry skills as a tool to develop a relationship with a patient for the dual purpose of helping that patient with the problem of adjustment, and encouraging the patient to develop a broader interest in his or her surroundings as a result of increased knowledge of the plant world."

To this definition Ms. Burlingame adds, "In horticultural therapy you develop a program of working with flowers and plants, with the primary objective being to raise the level of motivation for the patient—whether his or her problem is mental or physical. Response will come from the patient in a renewed confidence, a warm feeling of achievement, and a greater interest in tomorrow than yesterday."

One reason plants are therapeutic is that caring for their needs helps us to become less self-involved. Another is that the rewards are tangible—a new leaf, a fragrant flower, an offspring. When you adopt a plant as your responsibility and it responds well to you, it increases your confidence not only as a gardener but as a person.

Basic to horticultural therapy is its tremendous adaptability. Work with plants, flowers, and gardens can be adjusted to every age, every illness or need. It has proven so beneficial that several universities now offer degrees in horticultural therapy.

The adaptability and success of horticultural therapy are based on the fact that human being and plants exist side by side in nature. Thus contact with orderly cycles of nature has a curative effect on the development of the individual intellectually, emotionally, socially, and physically. Through horticultural therapy patients learn to appreciate natural surroundings. Powers of observation are sharpened. People respond to watching things grow and feel a sense of responsibility to a plant that depends on them. They learn at first hand that people need plants and plants need people.
1. In the second paragraph, Ms. Candliss ______________ the role of a horticultural therapist.
   a. labels  
b. depicts  
c. discriminates  
d. creates

2. An horticultural therapist may use his/her floristry ______ ______ as a tool to help patients.
   a. interests  
b. hobbies  
c. talents  
d. facilities

3. The purpose of horticultural therapy is _____________.
   a. twofold  
b. ambiguous  
c. two-faced  
d. multiple

4. The primary objective of a horticultural therapy program is to ________________ the patient’s motivation.
   a. restrain  
b. degrade  
c. increase  
d. control

5. In the phrase a warm feeling of achievement, the word achievement can be replaced by ________________.
   a. accomplishment  
b. comfort  
c. knowledge  
d. peace
6. Taking care of plants may help us to become less _________.
   a. altruistic
   b. wrapped up in ourselves
   c. confident of our own abilities
   d. self-righteous

7. A new leaf, a fragrant flower, an offspring are ________ rewards we receive when caring for plants.
   a. out of reach
   b. delayed
   c. illusory
   d. concrete

8. Adaptability in the text refers to the capability that horticultural therapy has of___________.
   a. being learned
   b. being studied
   c. being controlled
   d. being adjusted

9. According to the text, nature has a(n) ________ effect on human beings.
   a. controlling
   b. healing
   c. mesmerizing
   d. undetermined

10. In the sentence People respond to watching things grow, the verb respond to indicates that human beings are _________ ________ plants.
    a. not affected by
    b. indifferent to
    c. sensitive to
    d. knowledgeable about
### APPENDIX C
### TEST RESULTS

Scores and Percentages

<table>
<thead>
<tr>
<th>Participants</th>
<th>TOEFL</th>
<th>M.C. test</th>
<th>Passages test</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>543</td>
<td>24/30 - 80%</td>
<td>23/31 - 74%</td>
</tr>
<tr>
<td>B</td>
<td>500</td>
<td>23/30 - 76%</td>
<td>16/31 - 52%</td>
</tr>
<tr>
<td>C</td>
<td>572</td>
<td>26/30 - 87%</td>
<td>24/31 - 78%</td>
</tr>
</tbody>
</table>