CHAPTER ONE

The place of vocabulary in language assessment

Introduction

At first glance, it may seem that assessing the vocabulary knowledge of second language learners is both necessary and reasonably straightforward. It is necessary in the sense that words are the basic building blocks of language, the units of meaning from which larger structures such as sentences, paragraphs and whole texts are formed. For native speakers, although the most rapid growth occurs in childhood, vocabulary knowledge continues to develop naturally in adult life in response to new experiences, inventions, concepts, social trends and opportunities for learning. For learners, on the other hand, acquisition of vocabulary is typically a more conscious and demanding process. Even at an advanced level, learners are aware of limitations in their knowledge of second language (or L2) words. They experience lexical gaps, that is words they read which they simply do not understand, or concepts that they cannot express as adequately as they could in their first language (or L1). Many learners see second language acquisition as essentially a matter of learning vocabulary, so they devote a great deal of time to memorising lists of L2 words and rely on their bilingual dictionary as a basic communicative resource. Moreover, after a lengthy period of being preoccupied with the development of grammatical competence, language teachers and applied linguistic researchers now generally recognise the importance of vocabulary learning and are exploring ways of promoting it more effectively. Thus, from various points of view, vocabulary can be seen as a priority area in language teaching, requiring tests to monitor the

learners' progress in vocabulary learning and to assess how adequate their vocabulary knowledge is to meet their communication needs.

Vocabulary assessment seems straightforward in the sense that word lists are readily available to provide a basis for selecting a set of words to be tested. In addition, there is a range of well-known item types that are convenient to use for vocabulary testing. Here are some examples:

Multiple-choice (*Choose the correct answer*)

The principal was irate when she heard what the students had done.

a. surprised
b. interested
c. proud
d. angry

Completion (*Write in the missing word*)
At last the climbers reached the s_____ of the mountain.

Translation (*Give the L1 equivalent of the underlined word*)
They worked at the mill.

Matching (*Match each word with its meaning*)

1	accurate	____	a. not changing
2	transparent	____	b. not friendly
3	constant	____	c. related to seeing things
4	visual	____	d. greater in size
5	hostile	____	e. careful and exact
			f. allowing light to go through
			g. in the city

These test items are easy to write and to score, and they make efficient use of testing time. Multiple-choice items in particular have been commonly used in standardised tests. A professionally produced multiple-choice vocabulary test is highly reliable and distinguishes learners effectively according to their level of vocabulary knowledge. Furthermore, it will usually be strongly related to measures of the learners' reading comprehension ability. Handbooks on language testing published in the 1960s and 1970s (for example Lado, 1961; Harris, 1969; Heaton, 1975) devote a considerable amount of space to vocabulary testing, with a lot of advice on how to write good items and avoid various pitfalls.

Tests containing items such as those illustrated above continue to be written and used by language teachers to assess students' progress in vocabulary learning and to diagnose areas of weakness in their knowledge of **target-language** words, i.e. the language which they are learning. Similarly, scholars with a specialist interest in the learning and teaching of vocabulary (see, for example, McKeown and Curtis, 1987; Nation, 1990; Coady and Huckin, 1997; Schmitt and McCarthy, 1997) generally take it for granted that it is meaningful to treat words as independent units and to devise tests that measure whether – and how well – learners know the meanings of particular words.

Recent trends in language testing

However, scholars in the field of language testing have a rather different perspective on vocabulary-test items of the conventional kind. Such items fit neatly into what language testers call **the discrete-point approach** to testing. This involves designing tests to assess whether learners have knowledge of particular structural elements of the language: word meanings, word forms, sentence patterns, sound contrasts and so on. In the last thirty years of the twentieth century, language testers progressively moved away from this approach, to the extent that such tests are now quite out of step with current thinking about how to design language tests, especially for proficiency assessment.

A number of criticisms can be made of discrete-point vocabulary tests.

- It is difficult to make any general statement about a learner's vocabulary on the basis of scores in such a test. If someone gets 20 items correct out of 30, what does that say about the adequacy of the learner's vocabulary knowledge?
- Being proficient in a second language is not just a matter of knowing a lot of words – or grammar rules, for that matter – but being able to exploit that knowledge effectively for various communicative purposes. Learners can build up an impressive knowledge of vocabulary (as reflected in high test scores) and yet be incapable of understanding a radio news broadcast or asking for assistance at an enquiry counter.
- Learners need to show that they can use words appropriately in

their own speech and writing, rather than just demonstrating that they understand what a word can mean. To put it another way, the standard discrete-point items test receptive but not productive competence.

- In normal language use, words do not occur by themselves or in isolated sentences but as integrated elements of whole texts and discourse. They belong in specific conversations, jokes, stories, letters, textbooks, legal proceedings, newspaper advertisements and so on. And the way that we interpret a word is significantly influenced by the context in which it occurs.
- In communication situations, it is quite possible to compensate for lack of knowledge of particular words. We all know learners who are remarkably adept at getting their message across by making the best use of limited lexical resources. Readers do not have to understand every word in order to extract meaning from a text satisfactorily. Some words can be ignored, while the meaning of others can be guessed by using contextual clues, background knowledge of the subject matter and so on. Listeners can use similar strategies, as well as seeking clarification, asking for a repetition and checking that they have interpreted the message correctly.

The widespread acceptance of the validity of these criticisms has led to the adoption – particularly in the major English-speaking countries – of **the communicative approach** to language testing. Today's language proficiency tests do not set out to determine whether learners know the meaning of *magazine* or *put on* or *approximate*; whether they can get the sequence of tenses right in conditional sentences; or whether they can distinguish *ship* and *sheep*. Instead, the tests are based on **tasks** simulating communication activities that the learners are likely to be engaged in outside of the classroom. Learners may be asked to write a letter of complaint to a hotel manager, to show that they understand the main ideas of a university lecture or to discuss in an interview how they hope to achieve their career ambitions. Presumably good vocabulary knowledge and skills will help test-takers to perform these tasks better than if they lack such competence, but neither vocabulary nor any other structural component of the language is the primary focus of the assessment. The test-takers are judged on how adequately they meet the overall language demands of the task.

Recent books on language testing by leading scholars such as

Bachman and Palmer (1996) and McNamara (1996) demonstrate how the task has become the basic element in contemporary test design. This is consistent with broader trends in Western education systems away from formal standardised tests made up of multiple items to measure students' knowledge of a content area, towards what is variously known as alternative, performance-based or standards-based assessment (see, for example, Baker, O'Neil and Linn, 1993; Taylor, 1994; O'Malley and Valdez Pierce, 1996), which includes judging students' ability to perform more open-ended, holistic and 'real-world' tasks within their normal learning environment.

Is there a place, then, for vocabulary assessment within task-based language testing? To look for an answer to this question, we can turn to Bachman and Palmer's (1996) book *Language Testing in Practice*, which is a comprehensive and influential volume on language-test design and development. Following Bachman's (1990) earlier work, the authors see the purpose of language testing as being to allow us to make inferences about learners' language ability, which consists of two components. One is **language knowledge** and the other is **strategic competence**. That is to say, learners need to know a lot about the vocabulary, grammar, sound system and spelling of the target language, but they also need to be able to draw on that knowledge effectively for communicative purposes under normal time constraints. As I noted above, one of the main criticisms of discrete-point vocabulary items is that they focus entirely on the knowledge component of language ability.

Within the Bachman and Palmer framework, language knowledge is classified into numerous areas, as presented in Table 1.1. The table shows that language knowledge covers more areas than I indicated in the previous paragraph, but at the same time knowledge of vocabulary appears to be just a minor component of the overall system, a sub-sub-category of organisational knowledge. It is classified as part of Grammatical knowledge, which suggests a very narrow view of vocabulary as a stock of meaningful word forms that fit into slots in sentence frames. I will have a great deal more to say about the nature of vocabulary in Chapter 2, but for now let me point out that vocabulary knowledge is a significant element in several other categories of the table. The most obvious area is Sociolinguistic knowledge, which includes 'natural or idiomatic expressions', 'cultural references' and 'figures of speech'. Most people would regard these as belonging to the vocabulary of the language. In addition, the sociolinguistic

Table 1.1 *Areas of language knowledge (Bachman and Palmer, 1996: 68)*

Organisational knowledge (how utterances or sentences and texts are organised)
Grammatical knowledge (how individual utterances or sentences are organised)
Knowledge of vocabulary
Knowledge of syntax
Knowledge of phonology/graphology
Textual knowledge (how utterances or sentences are organised to form texts)
Knowledge of cohesion
Knowledge of rhetorical or conversational organisation
Pragmatic knowledge (how utterances or sentences and texts are related to the communicative goals of the language user and to the features of the language use setting)
Functional knowledge (how utterances or sentences and texts are related to the communicative goals of language users)
Knowledge of ideational functions
Knowledge of manipulative functions
Knowledge of heuristic functions
Knowledge of imaginative functions
Sociolinguistic knowledge (how utterances or sentences and texts are related to features of the language use setting)
Knowledge of dialects/varieties
Knowledge of registers
Knowledge of natural or idiomatic expressions
Knowledge of cultural references and figures of speech

category includes knowledge of registers, which are varieties of language associated with particular users, uses and contexts. One of the primary features of a register is the distinctive words and phrases used in it (McCarthy, 1990: 61–64). Thus, in these and other ways, Table 1.1 understates the contribution of vocabulary to language knowledge.

Bachman and Palmer (1996: 67) acknowledge that many language tests focus on just one of the areas of language knowledge, such as vocabulary. They give as an example a test for primary school children learning English as a foreign language in an Asian country. In the context of a teaching unit on 'Going to the zoo', the students are tested on their knowledge of the names of zoo animals (Bachman and Palmer, 1996: 354–365). The authors argue that, even at this elementary level of language learning, vocabulary testing should relate to some meaningful use of language outside the classroom.

However, their main concern is with the development of test tasks that not only draw on various areas of language knowledge but also require learners to show that they can activate that knowledge effectively in communication. An illustration of the latter kind of task is found in an academic writing test for non-native speakers of English entering a writing programme in an English-medium university (Bachman and Palmer, 1996: 253–284). The test-takers are required to write a proposal for improving the institution's admissions procedures. Rather than the single global scale that is often employed to rate performance on such a task, Bachman and Palmer advocate the use of several analytic scales, which provide separate ratings for different components of the language ability to be tested. In the case of the academic writing test, they developed five scales, for knowledge of syntax, vocabulary, rhetorical organisation, cohesion and register. Thus, vocabulary is certainly being assessed here, but not separately; it is part of a larger procedure for measuring the students' academic-writing ability.

Three dimensions of vocabulary assessment

Up to this point, I have outlined two contrasting perspectives on the role of vocabulary in language assessment. One point of view is that it is perfectly sensible to write tests that measure whether learners know the meaning and usage of a set of words, taken as independent semantic units. The other view is that vocabulary must always be assessed in the context of a language-use task, where it interacts in a natural way with other components of language knowledge. To some extent, the two views are complementary in that they relate to different purposes of assessment. Conventional vocabulary tests are most likely to be used by classroom teachers for assessing progress in

vocabulary learning and diagnosing areas of weakness. Other users of these tests are researchers in second language acquisition with a special interest in how learners develop their knowledge of, and ability to use, target-language words. On the other hand, researchers in language testing and those who undertake large testing projects tend to be more concerned with the design of tests that assess learners' achievement or proficiency on a broader scale. For such purposes, vocabulary knowledge has a lower profile, except to the extent that it contributes to, or detracts from, the performance of communicative tasks.

As with most dichotomies, the distinction I have made between the two perspectives on vocabulary assessment oversimplifies the matter. There is a whole range of reasons for assessing vocabulary knowledge and use, with a corresponding variety of testing procedures. In order to map out the scope of the subject, I propose three dimensions, as presented in Figure 1.1.

The dimensions represent ways in which we can expand our conventional ideas about what a vocabulary test is in order to include a wider range of lexical assessment procedures. I introduce the dimensions here, then illustrate and discuss them at various points in the following chapters. Let us look at each one in turn.

Discrete – embedded

The first dimension focuses on the **construct** which underlies the assessment instrument. In language testing, the term construct refers to the mental attribute or ability that a test is designed to measure. In the case of a traditional vocabulary test, the construct can usually be labelled as 'vocabulary knowledge' of some kind. The practical significance of defining the construct is that it allows us to clarify the meaning of the test results. Normally we want to interpret the scores on a vocabulary test as a measure of some aspect of the learners' vocabulary knowledge, such as their progress in learning words from the last several units in the course book, their ability to supply derived forms of base words (like *scientist* and *scientific*, from *science*), or their skill at inferring the meaning of unknown words in a reading passage. Thus, a **discrete** test takes vocabulary knowledge as a distinct construct, separated from other components of language competence. Whether it is valid to do so is a matter for debate and an issue that I

Discrete A measure of vocabulary knowledge or use as an independent construct	<————>	**Embedded** A measure of vocabulary which forms part of the assessment of some other, larger construct
Selective A measure in which specific vocabulary items are the focus of the assessment	<————>	**Comprehensive** A measure which takes account of the whole vocabulary content of the input material (reading/listening tasks) or the test-taker's response (writing/ speaking tasks)
Context-independent A vocabulary measure in which the test-taker can produce the expected response without referring to any context	<————>	**Context-dependent** A vocabulary measure which assesses the test-taker's ability to take account of contextual information in order to produce the expected response

Figure 1.1 Dimensions of Vocabulary Assessment

return to in Chapter 4. However, most existing vocabulary tests are designed on the assumption that it is meaningful to treat them as an independent construct for assessment purposes and can thus be classified as discrete measures in the sense that I am defining it here.

In contrast, an **embedded** vocabulary measure is one that contributes to the assessment of a larger construct. I have already given an example of such a measure, when I referred to Bachman and Palmer's task of writing a proposal for the improvement of university admissions procedures. In this case, the construct can be labelled 'academic writing ability', and the vocabulary scale is one of five ratings which form a composite measure of the construct. Another example of an embedded measure is found in reading tasks consisting of a written text followed by a set of comprehension questions. It is common practice to include in such tests a number of items assessing the learners' understanding of particular words or phrases in the

text. Usually the vocabulary item scores are not separately counted; they simply form part of the measure of the learners' 'reading-comprehension ability'. In that sense, vocabulary assessment is more embedded here than in the academic-writing test, where the vocabulary rating may well be included in a profile report of each learner's writing ability.

It is important to understand that the discrete–embedded distinction does not refer primarily to the way that vocabulary is presented to the test-takers. Many discrete vocabulary tests do require the learners to respond to words which are presented in isolation or in a short sentence, but this is not what makes the test discrete. Rather, it is the fact that the test is focusing purely on the construct of vocabulary knowledge. A test can present words in quite a large amount of context and still be a discrete measure in my sense. For instance, I can take a suitable reading passage, select a number of content words or phrases in it and write a multiple-choice item for each one, designed to assess whether learners can understand what the vocabulary item means as it is used in the text. This may appear to be very much the same kind of test as the one I described in the last paragraph to illustrate what an embedded measure is, but the crucial difference is that in this case all the items are based on vocabulary in the passage and I interpret the test score as measuring how well the learners can understand what those words and phrases mean. I do not see it as assessing their reading comprehension ability or any other broader construct. Thus, to determine whether a particular vocabulary measure is discrete or embedded, you need to consider its purpose and the way the results are to be interpreted.

Selective – comprehensive

The second dimension concerns the range of vocabulary to be included in the assessment. A conventional vocabulary test is based on a set of target words selected by the test-writer, and the test-takers are assessed according to how well they demonstrate their knowledge of the meaning or use of those words. This is what I call a **selective** vocabulary measure. The target words may either be selected as individual words and then incorporated into separate test items, or alternatively the test-writer first chooses a suitable text and then uses certain words from it as the basis for the vocabulary assessment.