Repetition and learning

Repetition is essential for vocabulary learning because there is so much to know about each word that one meeting of it is not sufficient to gain this information, and because vocabulary items must not only be known they must be known well so that they can be fluently accessed. Repetition thus adds to the quality of knowledge and also to the quantity or strength of this knowledge.

There has been a great deal of research on how items should be repeated and much of this is relevant to learning vocabulary in another language.

The spacing of repetitions

A very robust finding in memory research in general (Baddeley, 1990) and second language vocabulary learning research in particular (Bloom and Shuell, 1981; Dempster, 1987) is that spaced repetition results in more secure learning than massed repetition. Massed repetition involves spending a continuous period of time, say fifteen minutes, giving repeated attention to a word. Spaced repetition involves spreading the repetitions across a long period of time, but not spending more time in total on the study of the words. For example, the words might be studied for three minutes now, another three minutes a few hours later, three minutes a day later, three minutes two days later and finally three minutes a week later. The total study time is fifteen minutes, but it is spread across ten or more days. This spaced repetition results in learning that will be remembered for a long period of time. The repetitions should be spaced at increasingly larger intervals.

Seibert (1927), Anderson and Jordan (1928) and Seibert (1930) investigated retention over periods of up to eight weeks. Their findings are all in agreement with Pimsleur’s (1967) memory schedule. Most forgetting occurs immediately after initial learning and then, as time passes, the rate of forgetting becomes slower. For example, Anderson and Jordan (1928) measured recall immediately after learning, after one week, after three weeks and after eight weeks. The percentages of material retained were 66%, 48%, 39% and 37% respectively. This indicates that the repetition of new items should occur very soon after they are first studied, before too much forgetting occurs. After this the repetitions can be spaced further apart. Griffin (1992) also found that most forgetting seems to occur soon after learning.

Bahrick (1984) and Bahrick and Phelps (1987) examined the recall of second language vocabulary items after very long periods of non-use, from 8 to 50 years. They found that the nature of the original learning influenced recall. Items which were initially easy to learn and which were given widely spaced practice (intervals of 30 days) were most likely to be retained over many years. The memory curves showed a decelerating drop for the first 3 to 6 years and then little change up to 25 to 30 years after
which there was further decline. Bahrick and Phelps' research supports the well established finding of the superiority of spaced over massed practice.

Pimsleur (1967), in a very clear and useful article, proposes a memory schedule to act as a guide for the size of the spaces between the repetitions. Pimsleur's suggestion, based on research evidence, is that the space between each repetition should become larger, with the initial repetitions being closer together and the later repetitions much further apart. There is no particular reason why the spacing between the repetitions is a matter of precise measurement, but it is interesting to look at Pimsleur's scale as a rough guide for the type of spacing suggested. The scale is exponential, so if the first interval was five seconds, then the next interval should be $5^2 = 25$ seconds, the next $5^3 = 125$ seconds, and the next $5^4 = 625$ seconds (about 10 minutes) and so on. Table 4.4 applies to the calculation across 11 repetitions.

### Table 4.4: Pimsleur's memory schedule

<table>
<thead>
<tr>
<th>Repetition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time spacing before the next repetition</td>
<td>5 secs</td>
<td>25 secs</td>
<td>2 mins</td>
<td>10 mins</td>
<td>1 hour</td>
<td>3 hours</td>
<td>1 day</td>
<td>3 days</td>
<td>25 days</td>
<td>4 months</td>
<td>2 years</td>
</tr>
</tbody>
</table>

The general principle which lies behind the spacing is that the older a piece of learning is, the slower the forgetting. This means two things. Firstly, after a piece of learning, the forgetting is initially very fast and then slows down. Figure 4.1 shows this. Notice that the curve drops steeply (a lot of forgetting) and then starts to level off (less forgetting).

![Figure 4.1: Amount of forgetting over a period of time](image)

Secondly, on the second repetition a piece of learning is older than it was on the first repetition and so the forgetting on the second repetition will be slower than it was. On the third repetition the forgetting will be even slower. Figure 4.2 illustrates this.

![Figure 4.2: Speed of forgetting during a series of repetitions](image)
Figure 4.2 shows that if we set an 80% probability of recall as the point at which a repetition is needed, then the time between the repetitions will need to become longer and longer. The right probability of recall level is one where the learner has forgotten enough to feel that the repetition is worthwhile attending to and yet not forgotten too much so that there is still a good chance of recalling and thereby strengthening the form-meaning connection.

Baddeley (1990: 154-155) speculates that, because long-term learning depends on physical changes in the brain, spacing repetitions allows time for the regeneration of neuro-chemical substances that make these changes. Massed learning does not allow enough time for these substances to regenerate and thus they cannot continue to make the physical changes needed for learning. This explanation is still a matter for debate and investigation.

Mondria and Mondria-de-Vries (1994) describe the “hand computer” as a way of organizing and focusing repetition. The “hand computer” is simply a box divided into five sections, with the second section larger than the first, the third larger than the second and so on. The words to be learned are put on cards and initially go into Section 1. When a word is known it is put into Section 2. When Section 2 fills up the words in Section 2 are reviewed and those that are still known go into Section 3 and those not recalled go back to Section 1. The same procedure continues for Sections 3, 4 and 5, with words not recalled going back to Section 1. This procedure can be easily computerized.

**Types of repetition**

Learning from repetition not only depends on the spacing of the repetitions but also on the nature of the repetition.

If there is a delay between the presentation of a word form and its meaning, learners have an opportunity to make an effort to guess or recall the meaning, and presumably this extra effort will result in faster and longer retained learning. However, the guessing can only be successful if the foreign word form gives a good clue to its meaning, either because the foreign and native words are cognates, or because the word form and its translation have previously been seen together. Experimental evidence shows that simultaneous presentation of a word form and its meaning is best for the first encounter and, thereafter, delayed presentation is best because there is then the possibility of effort leading to successful recall.

In an experiment by Royer (1973) the learners saw each foreign word and its English translation simultaneously on the first trial and guessed by attempting to recall on subsequent trials. The group who were studying under the recalling procedure learned significantly more correct responses on a test given immediately after the learning sessions.

... the act of successfully recalling an item increases the chance that that item will be remembered. This is not simply because it acts as another learning trial, since recalling the item leads to better retention than presenting it again; it appears that the retrieval route to that item is in some way strengthened by being successfully used.

(Baddeley, 1990: 156)
Landauer and Bjork (1978: 631) suggest that retrieval may be more effective than simultaneously seeing the word and its meaning because retrieval involves greater effort, or because retrieval is more similar to the performance required during normal use.

The use of retrieval is a very important part of the strategy of using word cards (see Chapter 19). It is the main justification for using cards instead of lists or notebooks. Learners need to know the importance of retrieval and how to make it a part of the whole range of their learning activities. Meeting words in listening and reading texts provides an opportunity for retrieval as does using words in speaking and writing. Teachers should tolerate and allow for delays in retrieving vocabulary in the strands of meaning-focused input and meaning-focused output because the retrievals are contributing to learning.

The combination of spaced repetition with retrieval is the basis of “a very powerful strategy that is easy to use and widely applicable ... Indeed, I would regard it as probably more broadly useful than any of the more traditional visual imagery mnemonics” (Baddeley, 1990: 158).

So far, we have been looking at repetition as being repetition of the same material. That is, the repetition contributes mainly to strength of knowledge. However, repetition can extend and enrich previous meetings. Table 4.5 outlines some of the possibilities.

Table 4.5: Types of repetition of word meaning

<table>
<thead>
<tr>
<th>Type of processing</th>
<th>Type of repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noticing</td>
<td>Seeing the same word form and simultaneously presented meaning again</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Recalling the same meaning several times</td>
</tr>
<tr>
<td>Generation</td>
<td>Recalling the meaning in different contexts requiring a different instantiation of the meaning</td>
</tr>
</tbody>
</table>

There are many degrees of generation depending on the closeness of the relationship of the meaning to be instantiated to the previously met concept or instantiation.

Learning from word cards will usually involve repetition of the same material because the cards themselves do not change from one repetition to another. Learners can change the way they process the cards, by thinking of new sentences containing the word, applying new mnemonic techniques, thinking of new instantiations of the word, and imagining contexts of use.

When words are met in reading and listening or used in speaking and writing, the generativeness of the context will influence learning. That is, if the words occur in new sentence contexts in the reading text, learning will be helped. Similarly, having to use the word to say new things will add to learning (Joe, 1995).

The types of repetition are related to the goal of learning. McKeown, Beck, Omanson and Pople (1985: 533) found that if simple definitional learning was the goal then more repetitions were better than fewer, but the fewer repetitions (four encounters) achieved respectable results. If dealing with the word in
context was needed, then the repetitions needed to enrich the knowledge of words. This enrichment was even more critical when fluency of access was required.

Stahl and Fairbanks (1986: 97) in a meta-analysis of vocabulary studies found somewhat similar results with repetition of the same forms, meanings and contexts having strong effects on measures of meaning recall. It seemed however that more elaborative repetition had stronger effects on passage comprehension measures than repetition of the same information.

The number of repetitions

Repetition is only one of a number of factors affecting vocabulary learning and the correlations between repetitions and learning generally are only moderate. For example, Saragi, Nation and Meister (1978) found a correlation of about .45 indicating that repetition accounted for around 20% of the factors involved in learning. It is thus not easy to fix on a particular number of repetitions needed for learning to occur.

Kachroo (1962) found that words repeated seven times or more in his coursebook were known by most learners. Crothers and Suppes (1967) found that most items in their vocabulary learning experiments were learned after six or seven repetitions. Tinkham (1993), like many other researchers, found that learners differed greatly in the time and number of repetitions required for learning. Most learners required five to seven repetitions for the learning of a group of six paired associates. A few required over twenty repetitions.

This chapter has looked at some of the important conditions that can lead to vocabulary learning. Teachers should develop the skill of examining the activities they use to determine why they are using the activity (the goals), if it is doing what it is supposed to be doing (the conditions and signs), and how it can be improved to better reach its goal (the features).
Vocabulary learning strategies are a part of language learning strategies which in turn are a part of general learning strategies. Schmitt (1997) provides a very useful overview of the rise in importance of strategy use in second language learning, noting that it grew out of an interest in learners' active role in the learning process.

It is not easy to arrive at a definition of what a strategy is, but to deserve attention from a teacher, a strategy would need to

1. involve choice. That is, there are several strategies to choose from.
2. be complex. That is, there are several steps to learn.
3. require knowledge and training.
4. increase the efficiency of vocabulary learning and vocabulary use.

There are numerous strategies which have these features. Learners not only need to know about these strategies, but need to have skill in using them.

This chapter does not look at communication strategies which compensate for vocabulary gaps in spoken language production (Poulisse, 1993). These are dealt with in the chapter on speaking.

A taxonomy of vocabulary learning strategies

There have been a few attempts to develop a taxonomy of vocabulary learning strategies, usually as a part of a piece of research into learners' strategy use. Schmitt (1997) developed an extensive taxonomy organised around Oxford's (1990) social, memory, cognitive, and metacognitive categories. Gu and Johnson (1996) also developed a substantial list divided into beliefs about vocabulary learning, metacognitive regulation, guessing strategies, dictionary strategies, note-taking strategies, memory strategies (rehearsal), memory strategies (encoding), and activation strategies. Williams (1985) identifies five potentially trainable strategies for working out the meaning of unfamiliar words in written text. These include inferring from context, identifying lexical familiarisation, unchaining nominal compounds, synonym search, and word analysis. Williams suggests that these become the focus of deliberate, intensive teaching. What is interesting in several of these, particularly lexical familiarisation and unchaining nominal compounds is how they involve reinterpretting known words. That is, a known word like *snap* (to break) may be used in the phrase *snap election*. Thus, they offer a different kind of challenge to a second language learner who might not know any meaning for the words than for a native speaker who has to extend the reference of known words.

The following taxonomy tries to separate aspects of vocabulary knowledge (what is involved in knowing a word) from sources of vocabulary knowledge, and learning processes. The taxonomy is best viewed as a matrix with the aspects of what is involved in knowing a word listed along one side, and the sources and processes along the other. Let us look at a few examples to make this clear. One of the sources of information about a word is the contexts in which it occurs, for example in a reading text. The context can be a source of information for the various aspects of what is involved in knowing a word - its written form, its spoken form, its word parts, its meaning, what it refers to, its grammar, its collocations, and constraints on its use. Similarly, the learning process of retrieval can be used to establish the written form of the word, its spoken form, its word parts and so on.
Table 15.1 lists the major divisions of the taxonomy.

**Table 15.1: A taxonomy of kinds of vocabulary learning strategies**

<table>
<thead>
<tr>
<th>General class of strategies</th>
<th>Types of strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning: Choosing what to focus on and when to focus on it</td>
<td>Choosing words</td>
</tr>
<tr>
<td></td>
<td>Choosing the aspects of word knowledge</td>
</tr>
<tr>
<td></td>
<td>Choosing strategies</td>
</tr>
<tr>
<td></td>
<td>Planning repetition</td>
</tr>
<tr>
<td>Sources: Finding information about words</td>
<td>Analysing the word</td>
</tr>
<tr>
<td></td>
<td>Using context</td>
</tr>
<tr>
<td></td>
<td>Consulting a reference source in L1 or L2</td>
</tr>
<tr>
<td></td>
<td>Using parallels in L1 and L2</td>
</tr>
<tr>
<td>Processes: Establishing knowledge</td>
<td>Noticing</td>
</tr>
<tr>
<td></td>
<td>Retrieving</td>
</tr>
<tr>
<td></td>
<td>Generating</td>
</tr>
</tbody>
</table>

Let us now look at each of these types of strategies in turn.

**Planning vocabulary learning**

The strategies in this category involve deciding on where to focus attention, how to focus the attention, and how often to give attention to the item.

**Choosing words**

In Chapter 2 we looked at the various levels of vocabulary (high frequency, academic, technical, low frequency) and the different returns for learning effort. Learners should know what their vocabulary goals are and should choose what vocabulary to focus on in terms of these goals. Gu and Johnson's (1996) study noted that this evaluative selective attention was a noted characteristic of the successful learners. In Chapter 14 on learner autonomy we noted the importance of learners having access to lists of high frequency and academic words and being able to obtain frequency information from dictionaries. Learners should have a clear strategy for deciding what vocabulary to focus on and where to find this vocabulary. The chapter on autonomy looks at this in more detail.

**Choosing aspects of word knowledge to focus on**

In Chapter 3 and Chapter 14 we looked at what is involved in knowing a word. Learners need to be aware of these aspects of word knowledge. Most often the main concern will be knowing the meaning of the word, but the need to use a word in speaking or writing will require attention to other aspects of knowing a word.

**Choosing strategies**

One of Gu and Johnson’s (1996) most successful groups of learners were those who actively drew on a wide range of vocabulary learning strategies. Their least successful group used a much more limited range. Successful strategy users need a strategy for controlling their strategy use. This involves choosing the most appropriate strategy from a range of known options and deciding how to pursue the strategy...
and when to switch to another strategy. For example, consulting a dictionary could be followed by the use of word cards to establish knowledge of the word.

Planning repetition

Most vocabulary learning requires repeated attention to the item. One of the most important strategies to encourage remembering is the use of increasingly spaced retrieval (Baddeley, 1990; Pimsleur, 1967). This can involve an informal schedule for returning to previously studied items on word cards and the recycling of old material, or it can involve a more organized review system using a computer or a filing system (Mondria and Mondria-de Vries, 1994). The role of repetition is looked at in more detail in the Chapter 4.

Procedures that integrate strategies

Several writers (Kramsch, 1979; Mhone, 1988; McComish, 1990; Sanaoui and Mot, xx) describe procedures for getting learners to select their own vocabulary for learning, record it, learn it, share it with others, and be monitored and assessed on their learning. These procedures also relate to the use of vocabulary notebooks (Schmitt and Schmitt, 1995) in that they aim at learners taking responsibility for their own learning and developing the necessary skills to do this.

Let us look closely at Kramsch’s (1979) procedure and consider and expand on the options available at each point.

1 Selecting the words: The learners are told that they need to learn five words a day - three chosen by them and two chosen by the teacher. Kramsch suggests that the learners look for vocabulary that they can readily use in talking or writing, and words that are easily adaptable to any context. The learners need to develop a feeling for which words are low frequency and which are more useful. Now learners can more readily gain information about which words are particularly useful by consulting the frequency markings in the later editions of the COBUILD Dictionary, by looking at the defining vocabulary in the back of the Longman Dictionary of Contemporary English, or by consulting word lists such as West (1953) or Hindmarsh (1980). Kramsch points out that sometimes a word is chosen for aesthetic reasons, because it sounds nice, because it represents an unusual concept, or because it has personal associations. McKenzie (1990) and Carroll and Mordaunt (1991) suggest that learners should choose words that are “semi-familiar” to them, that is, words that are partly known and that they can imagine themselves using in that future. This is largely to help receptive vocabulary become productive.

Robinson (1989) argues that more attention should be given to getting learners to use the high frequency, non-context dependent vocabulary that can be used to paraphrase and define. This allows learners to cope with breakdowns in communication and to more effectively engage in the negotiation of meaning.
2. **Recording the words and monitoring the recording:** Kramsch (1979) suggests writing the words on index cards along with a synonym, antonym or translation and an example sentence. The way the word is recorded will have a strong effect on how it is learned. The teacher can check the cards to ensure that the words are useful and that the information recorded such as the context sentence is correct. Carroll and Mordaunt (1991) also suggest noting definitions, etymology, the sentence the word occurred in, a sentence created by the student, and synonyms and antonyms. Schmitt and Schmitt (1995) suggest elaborating the information over a period of time by listing derivatives, collocates, mnemonic cues, and stylistic information. McComish’s (1990) word spider is a way of helping remember the various types of information to look for, and largely corresponds to the various aspects of what is involved in knowing a word.

3. **Learning the words:** If words are recorded on small cards with the word on one side and its translation on the back, then learners can be instructed in the best ways to apply rote learning procedures. Similarly, generative procedures like the keyword technique and mental elaboration through self-created contexts, cause-effect chains (Sokmen, 1992) and situational links can be used. Learners need to be aware of the ways they can enhance learning and the principles which lie behind the techniques.

4. **Sharing with others:** Learners regularly should present a word or a few words to others by writing it on the board, defining it and saying where they met it, why it’s worth learning, how they remembered it, and giving some example sentences containing it. Learners get a boost from this presentation when they find that others add the word to their own store of items to learn. The class can question the presenter about the word and make suggestions for learning. This is also a useful opportunity for the teacher to provide comments. These presentations could only deal with a very small number of words but they can be very useful in reinforcing what can be known about a word and how it can be learned, and in developing an enthusiasm for vocabulary.

5. **Assessing and monitoring learning:** Kramsch (1979) suggests the following procedure for testing learning. The learners work in pairs or small groups and exchange sets of cards. Each learner is tested on five words by her partner. The learner has to define the tested word and give a sample sentence containing it. Points are awarded. This testing provides another opportunity for sharing words. Another way is for learners to supply the teacher with a list of, say 20, words each week. The teacher makes a brief note next to ten of the words. If the teacher writes *der.*, after a word, the learner has to write three derived forms of the word. If the teacher writes *coll.*, the learner has to provide three collocates. If the teacher writes *sent.*, the learner has to write a sentence using the word.

6. **Recycling the vocabulary:** The learners are encouraged to indicate in their writing, by using an asterisk, the words that they have used which were on their cards. They are also encouraged to make conscious and deliberate efforts to use what they have learned. To a very small degree this can be done through classroom games and activities, but primarily it depends on each learner’s initiative.

Schmitt and Schmitt’s (1995) description of the principles lying behind vocabulary notebooks and the ways in which they can be used is an excellent guide for teachers wishing to develop a strategy programme.
Experimental studies

Pre-tests

Experimental studies of second language vocabulary learning typically involve a pre-test, some treatments, a post-test and possibly a delayed post-test. For example, learners' knowledge of particular words could be tested in a pre-test. The learners then read a text containing the words, and then the words are tested again in the post-test to see if reading the text resulted in vocabulary learning. We will look at three major vocabulary issues in this kind of design - the use of pre-tests, the range of variables affecting the difficulty of vocabulary learning, and the need to use tests which can measure an appropriate amount of learning.

In Chapter 27 on vocabulary testing, we saw that different testing formats testing the same vocabulary have only moderate correlations of around .7. This means that differing formats to some degree are testing different kinds of knowledge. If the pre-test and the post-test use different formats, for example the pre-test may be a yes/no test and the post-test a translation test, then the difference between the pre- and post-tests will be a result of the different test formats as well as a result of the experimental treatment. Treatment and test format will thus be confounded.

Pre-tests can cause other problems. They may make the learners aware that the experiment is focusing on vocabulary and thus the learners give attention to vocabulary that they would not usually give. A pre-test, particularly a multiple-choice test, may result in some learning.

There are several ways of dealing with these problems. In order not to alert the learners to the vocabulary focus of the research, the vocabulary test could be given a day or two before the treatment. Another option is to have no pre-test but to rely on random assignment to the groups to equate vocabulary knowledge. Some studies have not used a pre-test but have asked learners after the treatment and post-test to indicate which words they already knew before they experienced the treatment (Laufer and Hulstijn, in press). In order to control the learning effect of a pre-test, sometimes a control group is used which sits the pre- and post-tests but which does not have the treatment. Each of these ways of dealing with the problems of pre-tests have their strengths and weaknesses and the researcher needs to choose the option that causes the least reduction of the validity of the study.

Factors affecting vocabulary difficulty

There are numerous factors that affect the difficulty of vocabulary learning. Most of these factors have been discussed in Chapter 3 on what is involved in knowing a word. Higa (1965) provides a very useful survey of factors and Tinkham's (1993) experiment is an excellent model of control of many of these factors. Tinkham wanted to investigate whether learning words together in a lexical set, like the names of fruit, or clothing, would be easier or more difficult than learning unrelated words together. Tinkham decided to use artificial words because he wanted to control for the factors of word length (number of syllables), pronounceability and meaningfulness. All of these factors have been shown to have effects on vocabulary learning. Tinkham had some learners work with associating the artificial words with one set of meanings and had other learners work with associating the same forms with a different arrangement of meanings. If Tinkham had used real words, he would not have been able to control for
the important factors of word length, pronounceability and meaningfulness. The use of artificial words is thus a strength of the study. All the meanings associated with the artificial words were the same part of speech. Because all learners learned both the semantically related and the unrelated words in one of his studies, the order in which the sets were learned was balanced across the learners. That is, some learned the unrelated words first and others learned the related words first. Tinkham (1993) found that the related words were more difficult to learn. In a later study, Tinkham (1997) compared lexical sets with thematically related words. Using thematically related words, like frog, green, jump, meant that part of speech was not controlled. The semantically related words were all nouns, and previous research has shown that nouns are generally easier to learn than other parts of speech (Rodgers, 1969) (see Laufer, 1998: 148-149, questioning the effect of part of speech). In spite of this, the experiment showed that thematically related words were easier to learn. The effect would have been even greater if part of speech could have been controlled by choosing only nouns for the thematically related words. Laufer (1998) and Higa (1965) provide excellent surveys of features affecting vocabulary learning.

Table 28.2 lists features that can affect the difficulty of vocabulary learning and that would need to be controlled in an experiment comparing the learning of different words.

Table 28.2: Factors affecting vocabulary difficulty

<table>
<thead>
<tr>
<th>Type</th>
<th>Factor</th>
<th>Finding</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>part of speech</td>
<td>Nouns are easier to learn than other parts of speech.</td>
<td>Rodgers (1969) But see Laufer (1998) Ellis and Beaton (1993)</td>
</tr>
<tr>
<td></td>
<td>imageability</td>
<td>Words whose meaning is easy to visualize are easier to learn.</td>
<td>Ellis and Beaton (1993)</td>
</tr>
<tr>
<td></td>
<td>receptive/ productive</td>
<td>Receptive learning is easier than productive learning.</td>
<td>Griffin (1992) Stoddard (1929)</td>
</tr>
<tr>
<td>Relationship with L1 or previous L2 knowledge</td>
<td>pronounceability</td>
<td>The easier a word is to pronounce, the easier it is to learn.</td>
<td>Rodgers (1969) Ellis and Beaton (1993)</td>
</tr>
<tr>
<td></td>
<td>codability</td>
<td>A foreign word with a first language equivalent is easier to learn than a word representing a new concept.</td>
<td>Higa (1965)</td>
</tr>
<tr>
<td></td>
<td>meaningfulness</td>
<td>A word form which resembles a known word is easier to learn. But confusion of similar forms may occur with meaning connections.</td>
<td>Higa (1965) Ellis and Beaton (1993) Laufer (1998)</td>
</tr>
<tr>
<td>Relationship with other words learned at the same time</td>
<td>serial position</td>
<td>Words at the beginning and end of a list are easier to learn than those in the middle.</td>
<td>Higa (1965)</td>
</tr>
<tr>
<td></td>
<td>number of words</td>
<td>Difficult words are easier to learn in small groups of words.</td>
<td>Crothers and Suppes (1967)</td>
</tr>
</tbody>
</table>
Table 28.2 is organized using some of the categories used by Higa (1965), namely features intrinsic to the word itself, features relating the word to previous knowledge of the first or second language, and relationships between words learned together at the same time. These features relate to controlling word difficulty between different words and groups of words. There are additional features that could cause differences between the same words involved in different treatments. For example, Ellis and He (in press) looked at the same vocabulary being learned through premodified input, negotiated interaction and modified output. Features affecting vocabulary difficulty in these circumstances include the number of repetitions, degree of generative use (Joe, 1998), salience, and time spent on learning.

So far we have looked at the issues of pre-tests and controlling for vocabulary difficulty on experimental design. The third issue to consider is the kinds of tests used to measure vocabulary knowledge.

Types of tests

Tests of vocabulary may differ in their sensitivity, that is, the degree to which they give the learners the chance to make use of partial knowledge of a word. If a researcher uses a less demanding test then learners can be given credit for partial knowledge where this is appropriate. If a researcher uses several tests of differing sensitivity involving the same words, then the strength of knowledge of these words can be measured and related to what happened to them in the previous experimental treatment. In this section we will look at vocabulary measures which place differing knowledge demands on the learners focusing on knowledge of the word form, its meaning and the connection between the form and its meaning. The three sets of factors affecting item difficulty are reception\production, recognition\recall, and imprecise\precise.

A study by Nagy, Herman and Anderson (1985) was among the first to point out the need for tests which give credit for partial knowledge. Although common-sense suggested that the major way in which native speakers increase their vocabulary is through meeting words in context, experimental studies had great difficulty in showing that such learning occurred. Nagy, Herman and Anderson's explanation of this difficulty was that some learning occurred, but because vocabulary learning is largely a cumulative process, very easy tests would need to be used to pick up the small amount of learning that may result from a single encounter with a word in context. The measure that they chose to use was a multiple choice format where the distractors were not related in form or meaning to the correct answer. When learners worked on such an item, if they knew something about the tested word, they would have a good chance of making the correct choice. Here is a sample item.

gendarme means a) to trick or trap someone
  b) policeman
  c) spoken as if one was out of breath or having trouble breathing
  d) the secret collection of information about another country
  e) the illegal transportation of goods across a border
  f) don't know

Note that some of the distractors are a different part of speech and that the meanings of the distractors have little connection with the meaning of the tested item. If the learner knows that gendarme is a kind of job or involves the police, then the correct answer can be chosen. The knowledge needed to answer this item correctly is less than the knowledge needed to answer a more demanding item like the following.
By using a less demanding measure of vocabulary knowledge, Nagy, Herman and Anderson (1985) were able to show that some learning does occur through meeting words in context. Without an appropriate measure, learners would not have been given credit for the knowledge that they had gained. In Chapter 27 we looked at how tests can be used to measure words that learners do not know well.

Similar thinking lies behind the format of the Vocabulary Levels Test (Nation, 1983, 1990).

A range of vocabulary measures: There have been several attempts to create scales of vocabulary knowledge. It is important in such scales to distinguish different aspects of vocabulary knowledge from different strengths of knowledge of the various aspects. For example, knowing the inflected and derived forms of a word is better placed on a separate scale from being able to relate a word form to its meaning. This is the approach taken by Ringbom (1987: 37) who has separate scales for accessibility, morphophonology, syntax, semantics, collocation, and association, and by Henriksen and Hastrup (1998) who distinguish three scales of partial and precise comprehension, reception and production, and depth of knowledge. Although different aspects of vocabulary knowledge are related to each other, these relationships are not sufficiently predictable or consistent to allow them to be placed on a general scale of test item difficulty. Because the focus in this section is on the differing degrees of sensitivity of vocabulary test item formats, it is thus simpler and clearer to distinguish aspects of knowledge from strength of knowledge, where strength refers to how well a particular aspect is known.

The aspect of vocabulary knowledge examined here is one that is central to vocabulary knowledge and involves three kinds of knowledge - knowledge of the written form of a word, knowledge of a meaning, and the knowledge that connects a particular form with a particular meaning. For example, a learner of English might be aware of the form brunch. The learner might also know that there is a concept for a single meal which takes the place of breakfast and lunch. The learner might also know that the form brunch is the appropriate form to communicate the concept of a meal combining breakfast and lunch. It is possible to know the form brunch and have no concept of its meaning. It is also possible to be familiar with the form, to have the appropriate concept but not to connect the two.
Three factors affecting test item difficulty: The grading of vocabulary test formats suggested here is based on three major factors which affect the difficulty of test items - the receptive\productive distinction, the recognition\recall distinction, and the imprecise\precise distinction.

1 \textit{Receptive}\textit{productive}

The receptive\productive distinction is well recognised in second language teaching and is sometimes called passive\active (Stoddard, 1929; Morgan and Oberdeck, 1930). Receptive knowledge is that used in listening and reading, and involves going from the form of a word to its meaning, for example

\textit{Translate the underlined word into your first language.} He is a \underline{bold} writer.

Productive knowledge is that used in speaking and writing, and involves going from the meaning to the form, for example

\textit{Translate this word into English.} \underline{biji}.

Where the test item format is controlled for (Stoddard, 1929) receptive recall is easier than productive recall, irrespective of whether the tested items were learned receptively or productively.

2 \textit{Recognition}\textit{recall}

A recognition vocabulary item format involves the use of choices.

\begin{itemize}
  \item \underline{biji} a) seed \\
  \item b) path \\
  \item c) finger \\
  \item d) chair
\end{itemize}

A recall item requires the test-taker to provide the required form or meaning.

\textit{Translate this word into English.} \underline{biji}.

Recognition items are easier because even with partial knowledge a test-taker may be able to make the right choice. The recognition\recall distinction has been a matter of some debate in memory research (Baddeley, 1990: 271-275), but when the distractors are not very close in form or meaning to the target word, then recognition tests are easier than recall tests.

3 \textit{Imprecise}\textit{precise}

The imprecise\precise distinction relates to the degree of accuracy required in the answer. This can be reflected in the similarity of the choices provided, the degree of prompting, and degree of acceptance of an approximate answer. Items allowing for imprecise knowledge are easier because credit is given for partial knowledge.

\textbf{Table 28.3:} Eight test formats ranked according to three factors affecting difficulty
Let us now look at several different vocabulary test item formats, starting with the least demanding measure, that is the one requiring the smallest amount of knowledge to give a correct answer, and moving to the most demanding, that is the one requiring the greatest amount and strength of knowledge.

In the measures described below there is a deliberate attempt to test only the decontextualised word form in a written form, a single meaning of the word, and the connection between the two. Testing other aspects like the spoken form, knowledge of the linguistic contexts of the word, and multiple meanings would involve looking at other scales of vocabulary knowledge, and in combination with the measures to be described below would tend to decrease the difficulty of a measure.

Teachers and researchers may wish to mix items from other scales when devising tests, but for the purpose of clarity in this discussion of a range of measures, the discussion is deliberately limited to one important aspect of vocabulary knowledge - the form/meaning connection. It is likely that measurement is less demanding when the method of testing most closely resembles the method of learning (Stoddard, 1929).

**Receptive measures:**

*Sensitive multiple-choice (Imprecise receptive recognition)*

In their study of learning words from context, Nagy, Herman and Anderson (1985) developed three multiple-choice vocabulary tests for the same words at three levels of sensitivity.

Levels of difficulty were based on the similarity in meaning between the target word and the concepts represented by the distractors. At the highest level of difficulty, distractors represented concepts similar to or closely associated with the meaning of the target word. At the lowest level of difficulty, distractors were chosen to be as dissimilar from the target word meaning as possible, even in terms of the implied part of speech. At the intermediate level of difficulty, distractors were chosen to be mostly in the same part of speech, but otherwise fairly diverse semantically (Nagy, Herman and Anderson, 1985: 239).

The distractors for the imprecise items were made up of the correct answers for other items, short definitions of other concepts in the experimental passages, and other definitions, all of which were not closely related in meaning to the correct answer. At the easiest level where choices may describe different parts of speech, it is necessary to make sure that the form of the tested word, namely its suffix, does not signal the correct answer.
The Vocabulary Levels Test (Nation, 1983) uses a matching item type at Nagy, Herman and Anderson's intermediate level of difficulty, in that the words in a block are all the same part of speech but are unrelated in meaning.

In their study of the effect of distractors when testing recognition memory, Dale and Baddeley (1962: 93) claim that "subjects' incorrect responses are non-random".

Our results indicate that when memory for an item is imperfect the subject is able to identify certain of its characteristics. If a number of items in the set of alternatives from which he has to choose have similar characteristics, then he will be unable to tell which item is new and which has been seen before. If only one has them, then he can correctly identify it (Dale and Baddeley, 1962: 94).

Imprecise items try to avoid choices with significant similar characteristics.

Another way of making an easy "imprecise" multiple choice measure is to use choices where one of the choices has a clear connection with the target word but it is not a definition.

e.g. fertilizer
   a. growing plants
   b. medicine
   c. history
   d. don't know

Learners are asked to make a choice and then explain the connection between the target word and their choice. The correct choice (e.g. 'growing plants') is designed to be as helpful as possible for someone who has only a vague idea about the meaning of a word. This choice acts as a cue which triggers recall of learners' existing ideas about the word. A key feature of this format is that it provides only minimal information about the target word. This is important when the test is used as a pre- and post-measure in which every effort must be made to eliminate learning from the test itself. The provision of a Don't know option and the requirement to explain the connection are to discourage wild guessing. The difficulty in designing such items lies in providing the most obvious semantic association as the correct choice. The distractors must be feasible within the context of the test but clearly different in meaning. When the test is used as a dependent measure, the distractors in some items may be the correct answers in other items. Another way (Newton, in progress) is to make each distractor a semantic associate for another of the target words. That is, instead of using other answers as distractors, words of related meaning to other target answers are used.

Demanding multiple choice (Precise receptive recognition)

These have already been described in the discussion of sensitive multiple-choice items. They differ from the sensitive multiple-choice items in that the demanding items have distractors which are "closely related to or similar to the target word meaning" (Nagy, Herman and Anderson, 1985: 240).

This kind of item is typical of measures of receptive vocabulary knowledge. While demanding multiple-choice items are very useful in determining how well a learner knows a word, if these were the only items, the learners would not receive credit for a considerable amount of vocabulary knowledge.

The major issue with multiple-choice items is the number and nature of the choices. It is worthwhile having clear criteria for the construction of distractors and for trialing the items before they are used.
Recalling a related meaning (Imprecise receptive recall)

Interview procedures designed to examine vocabulary knowledge typically involve a question like “Does this word remind you of anything?” (Nagy, Herman and Anderson, 1985). The learner may be given credit for responding by suggesting something which shows partial knowledge of that word, such as a superordinate, one aspect of its meaning, or a part of an example. As vague ideas are usually involved because of the learner’s partial knowledge, the interviewer may need to seek some explanation of the connection. Training and inter-rater reliability checks may need to be used to ensure consistent presentation and scoring.

Meaning recall (Precise receptive recall)

This is the last and most demanding of the receptive measures described here. The learners see a word and are asked to define it, translate it into their first language, or produce a synonym, or indicate if they could provide a meaning for the word (the yes/no test). It may be used as part of an interview procedure “Can you tell me what this word means?” or, particularly for translation, as a pencil-and-paper test - “Translate the following words into your first language.” Usually a scoring scale is used when marking the answers. For example, in their interview procedure Nagy, Herman and Anderson (1985) used a four-point scale.

<table>
<thead>
<tr>
<th>Score</th>
<th>Complete understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No correct knowledge</td>
</tr>
<tr>
<td>1</td>
<td>A vague and related answer</td>
</tr>
<tr>
<td>2</td>
<td>A key part of the concept missing</td>
</tr>
<tr>
<td>3</td>
<td>Complete understanding</td>
</tr>
</tbody>
</table>

To make scoring more consistent, the scale may need to be rewritten for each item. Here is an example for the word disillusioned (Nagy, Herman and Anderson, 1985: 243).

<table>
<thead>
<tr>
<th>Student answer</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Not illustrated correctly”</td>
<td>0: no correct knowledge</td>
</tr>
<tr>
<td>“I think it is something imagined ... a picture of something in your mind”</td>
<td>1: answer shows mental activity and is vague</td>
</tr>
<tr>
<td>“If you’re like led astray. If you’re made to believe something that’s not really true.”</td>
<td>2: answer does not convey that the person must realize the deception and consequently feel let down and disappointed</td>
</tr>
<tr>
<td>“If you have ideas about something and you find out it’s the opposite, you’re disillusioned. Your beliefs are shattered.”</td>
<td>3: answer conveys a complete understanding</td>
</tr>
</tbody>
</table>

Translations could be placed on a similar scale (Nurweni and Read, in press).

This kind of item requires retrieval of the meaning of a word form. Being able to recall the meaning of a form without choices provided requires a substantial degree of receptive knowledge. The yes/no test is a precise receptive recall measure. It was initially developed by Anderson and Freebody (1983) and
subsequently refined and applied by Meara (Meara and Buxton, 1987) in his Eurocentres Vocabulary Size Test. Although the learners do not overtly produce a meaning for the listed words, the instructions and design requires them to do this covertly.

It may be possible to measure learning from an experimental treatment after the treatment by comparing the experimental and control groups on the number of trials needed to reach a set criterion of knowledge, for example, completely accurate recall of the meanings of a set of tested words. That is, the learners would take part in an experiment, say, involving incidental learning of vocabulary. After the experimental treatment ends, each learner is then tested by being shown a word, asked to recall its meaning, and then being shown its meaning. This continues for the group of tested words until the learner can recall the meaning of every word in the group. If learning occurred during the experimental treatment it should take the experimental group less repetitions to learn the words than it would take the control group who did not go through the experimental treatment.

Trials to criterion has been used as a combined treatment and testing measure of vocabulary learning by Tinkham (1993) and Waring (1997). A single case study by Beaton, Gruneberg and Ellis (1995) also used relearning vocabulary as a test of previous learning.

productive measures:

All the measures looked at so far have provided the form of the word which the test-taker has to respond to by recognizing or recalling some aspect of its meaning. The productive measures require the learner to demonstrate knowledge of the word form, a meaning, and the connection between a form and its meaning, by having to recognize or recall the word form when the meaning is provided. This usually requires greater knowledge of a word, and learners' productive vocabulary is normally considered to be much smaller than their receptive vocabulary (Channell, 1988: 85; Clark, 1993: 245; Aitchison, 1994: 224). Measures of productive knowledge are more demanding than receptive measures.

Sensitive multiple choice (Imprecise, productive recognition)

The stem of the multiple-choice item is a definition, translation, or picture and the choices are the word forms.

end or highest point (a) event
(b) profit
(c) tip
(d) copy

The degree of difficulty of productive multiple choice is affected by the formal similarity between the correct answer and the distractors, and by the resemblance of the distractors to words with a similar meaning to the correct answer. The greater the formal difference, the easier the item.

Demanding multiple choice (Precise productive recognition)

The stem is the meaning expressed as a picture, an L1 word, or an L2 definition. The choices are L2 words with similar forms or related meanings.

jas a) shirt
b) coat
c) jersey
d) jacket

**Cued recall (Imprecise productive recall)**

The test-taker sees a definition, a translation, or a picture and has to complete a word where some of the first letters are provided. The number of letters needed to complete the word is not indicated.

```
an additional part suppl___________
```

The number of letters provided is the minimum necessary to exclude other possible words. For example, if only suppl was provided in the above example, then some learners might write support. Laufer and Nation (1995) used a similar kind of measure in a sentence context in their validation of the Lexical Frequency Profile.

```
The money from fruit-picking was a suppl___________ to their regular income.
```

Alternatively, tip of the tongue type recall (Brown and McNeill, 1966) can be scored on a scale of nearness to the correct form.

**Productive retrieval (Precise productive recall)**

A productive retrieval test involves the learners seeing a definition, object, picture or first-language word, and having to provide the second-language word form.

A difficulty with such tests is that without cuing of the word form the learners may provide synonyms or hyponyms instead of the target word. The most realistic kind of retrieval is within the context of normal language use, such as when doing a piece of writing, but it is difficult to know if the non-occurrence of a particular word is the result of lack of knowledge of the target item or lack of opportunity to use the target item. The Lexical Frequency Profile (Laufer and Nation, 1995) was developed as a way of investigating learners' productive vocabulary, but it measures word frequency levels rather than specific words.

**Vocabulary development and the sensitivity scale:** There are numerous problems with this speculative classification. Firstly, it has been assumed that the receptive/productive distinction is the primary one (see Meara (1990) for an argument that might support this). But there is in fact almost no research on this in studies of human memory that is directly applicable to vocabulary learning. There is however considerable research on the recognition/recall distinction (Baddeley, 1990: 271-291). Recognition involves selecting from choices while recall involves having to retrieve the word form or meaning. It may be that the recognition/recall distinction is the more basic one. Secondly, while it is generally accepted that recognition is easier than recall, this is not always so. Some non-sensitive measures of recognition, that is those that have distractors that share many elements of form or meaning with the
correct answer, may be more difficult than recall, perhaps because they set up interference conditions
between the choices which make the test-taker confused. Thirdly, there are many factors affecting the
difficulty of particular item types. For example the difficulty of a productive recall item will be strongly
affected by the clarity of the translation, definition, or picture provided, and will also be affected by the
relationship between the nature of the test item and the manner in which the learner originally met or
studied the word. Fourthly, the items looked at here probably test declarative knowledge rather than
procedural knowledge.

For most purposes however, these problems are of little significance. The main issue is that studies of
vocabulary learning should choose an item type of appropriate difficulty, and where strength of learning
needs to be measured, should choose a range of suitably differing measures. The precise size of the
differences between the items in this range of measures is not as important as that the items should be
clearly different from each other.

Because different vocabulary tests measure vocabulary knowledge with different degrees of sensitivity,
it is important in experimental studies with pre- and post-tests that the pre-test is the same test or test
format as one of the post-tests. If this is not done, the difference in the results of the pre- and post-tests
will be partly the result of the different degrees of sensitivity of the tests. It may be possible to use a pre-
test which is more sensitive than the post-tests but, if the pre-test and post-test are compared, the gains
will be underestimated.

As more attention is given to vocabulary in second language learning, the investigation of vocabulary
knowledge needs to become more sophisticated. The investigative tools need to take account of what
we know about the nature of vocabulary knowledge, and what we know about the difficulties involved
in measuring it.

Observational studies

There is a very wide range of studies which look at the conditions under which vocabulary learning and
vocabulary strategy use occur. These studies do not usually involve the use of a control group and pre-
and post-tests, but look at how certain tasks influence vocabulary learning, the strategies that learners
use or say they use when they are involved in vocabulary learning, and how vocabulary knowledge
changes as learners’ proficiency changes.

Tests of vocabulary size

We have looked at the difficulties involved in measuring vocabulary size in Chapter 25. Most of the
research on vocabulary size has taken poor account of the issues of deciding what will be counted as
words, sampling of items for the test, and deciding on a test format (Nation, 1993). The result has been
that most studies of native speakers’ and non-native speakers’ vocabulary size provide little reliable
information.

Reporting vocabulary research

Research not only needs to be done well, it needs to be reported well with enough detail to allow
evaluation of the quality of the research and for the research to be replicated.
From a vocabulary perspective, it is very important that the vocabulary tests used are clearly described and examples of items provided. Different test formats measure different kinds and strengths of knowledge and it should be made clear what formats were used and how they were scored.

A very common finding in vocabulary research is that there is usually a very wide range of vocabulary knowledge and of skill in learning vocabulary. Where it is possible, the range of individual performance should be noted. Tinkham (1993), for example, noted that some learners required only five or fewer trials to learn a set of items, others required over twenty trials. This information is useful for teachers and researchers when planning vocabulary learning.

The research on the factors affecting vocabulary difficulty includes a set of factors where the first language and other previously learned languages play an influential role. These include pronounceability, known concepts, and similar word forms. Cognates can play a very significant role in learning. It is therefore important that the first language background of learners is reported, so that readers can gauge the influence of this important factor on vocabulary learning.

Research on learning vocabulary in another language has to observe the same controls, principles and checks as research in other fields. This chapter has attempted to point out some of the issues that are particularly important for vocabulary research. These issues however are only part of the picture and good research needs to follow the more widely used principles of design.

References from Nation (1999) reading.


NB 1: If you are quoting from this article, please include the page number at the top of the page. (Please refer to APA referencing style).

NB 2: I have not included any references over 10 years old. If you wish to refer to any of these references, please refer to them as ‘in Nation, 1999’. (Please refer to APA referencing style).