DOCUMENT RESUME

ED 350 829 FL 020 528

AUTHOR Meara, Paul; Jones, Glyn
TITLE Vocabulary Size as a Placement Indicator.
PUB DATE 88
NOTE 9p.; In: Grunwell, Pamela, Ed. Applied Linguistics in Society. Papers from the Annual Meeting of the British Association for Applied Linguistics (20th, Nottingham, England, United Kingdom, September 1987. British Studies in Applied Linguistics, 3; see FL 020 520.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Computer Assisted Testing; *English (Second Language); Foreign Countries; Knowledge Level; Language Research; *Mathematical Formulas; *Student Placement; Test Use; *Vocabulary
IDENTIFIERS Eurocentres Schools (United Kingdom); *United Kingdom

ABSTRACT

The student placement test in English as a Second Language used for the Eurocentres schools in the United Kingdom differs from traditional placement tests in that it is a vocabulary test and does not attempt to measure other aspects of learner knowledge of English. The test attempts to measure absolute size of the learner's English vocabulary. The totally automated test is administered by displaying a large number of words on a computer screen, asking the learner which he knows, and using mathematical formulas to estimate vocabulary size. The examinee proceeds through a series of such screens. An example with French vocabulary illustrates the method. The mathematical model used for calculating vocabulary size evolved from military research. Results of administration of the test to about 250 students from a wide range of language backgrounds were compared with results of the previously-used placement test at the Eurocentres schools found relatively high correlations between the tests but some variation by language group. It is concluded that the test works well for placement but needs further refinement, and that visual recognition of words may not accurately reflect knowledge of them. Use of imaginary words in the test should also be reconsidered. (MSE)

# VOCABULARY SIZE AS A PLACEMENT INDICATOR

**Paul Meara**
Birkbeck College, London University
and
**Glyn Jones**
Eurocentres

## Background

This paper describes a placement test which was developed for the Eurocentres Group during 1986-87. The Eurocentres schools, like many other private sector language schools in the UK work on a cycle of short courses each lasting for four weeks. This means that every four weeks there is a huge turn-over of students, and a large number of new students have to be assessed and assigned to classes of an appropriate level. In most schools, this assessment is done by means of a complex battery of tests specially designed for this purpose, and generally referred to as placement tests. The tests currently used by Eurocentres, the *Joint Entrance Test* (JET), are fairly typical of this sort of test; they comprise a listening comprehension test, a grammar test and a reading test, supplemented by an oral interview.

The main problem with tests of this sort is that they take a long time to administer and mark. In a situation where time is at a premium because classes cannot be started until the placement procedure is completed this is obviously a serious shortcoming.

The tests that we have devised differ radically from traditional placement tests. They are very quick to administer (typically they need only 10-15 minutes to complete) and because the whole test is run by a small micro-computer the test is self-scoring and does not need to be checked by a teacher. This represents a large saving in teacher-hours, and greatly simplifies the placement procedure.

## The Test

The test we devised for Eurocentres is very different from a traditional placement test, in that it is basically a vocabulary test, and does not attempt to measure other aspects of the learner's knowledge of English. The justification for this approach is that there is a large body of evidence (for English as an L1) that vocabulary knowledge is heavily implicated in all practical language skills, and that in general, speakers with a large vocabulary perform better on a wide range of linguistic indicators than speakers with a more limited vocabulary (Anderson and Freebody 1981).

However, our test is not just a traditional vocabulary test of the type familiar from Cambridge Proficiency examinations. Instead of testing a small number of vocabulary items with complicated multiple-choice type tests, our test is

an attempt to measure the absolute size of a learner's vocabulary in English. We do this by simply displaying a large number of English words on the computer screen and asking the testee to decide whether he knows each of the words displayed or not. The computer program then uses some sophisticated mathematical techniques to estimate the testee's actual vocabulary size. The principal advantage of this methodology is that the test is totally automated. It takes less than 10 minutes to run, and scores itself without any manual intervention.

It is obviously not possible to demonstrate this technique in printed format, but you will get a rough idea of how the test works if you try the test in Table 1 before you go any further.

**Table 1**

Look through the French words listed below. Cross out words that you do not know well enough to say what they mean. Keep a record of how long it takes you to do the test.

| | | | |
|---|---|---|---|
| VIVANT | TROUVER | MAGIR | ROMPANT |
| MELANGE | LIVRER | IVRE | FOMBE |
| MOUP | VION | LAGUE | INONDATION |
| SOUTENIF. | SIECLE | TORVEAU | PRETRE |
| REPOS | GANAL | HARTON | TOULE |
| GOUTER | FOULARD | EXIGER | AVARE |
| ETOULAGE | ECARTER | MIGNETTE | JAMBONNANT |
| DEMENAGER | POIGNEE | EQUIPE | MISSONNEUR |
| AJURER | BARRON | CLAGE | TOUTEFOIS |
| LEUSSE | CRUYER | HESITER | SURPRENDRE |
| LAVIRE | SID | ROMAN | CHIC |
| ORNIR | CERISE | PAPIMENT | CONFITURE |
| GOTER | PONTE | | |

The test in Table 1 presents you with a list of French words and asks you to say which of these words you know. The words are actually a sample of words from the *deuxieme degré* of *Francais Fondamental*, which comprises a total of approximately 2000 high frequency French words, and if you have studied school French even to an elementary level, you should have been able to recognise at least some of these words. The test in Table 1 actually contains two types of item: real words (which you might have recognised) and imaginary, non-existent words (which you cannot possibly have recognised). This combination of real and imaginary words gives us four combinations of items and answers:

| type of item | real | imaginary |
|---|---|---|
| response YES | RY | IY |
| response NO | RN | IN |

Now suppose that you identified all the real words, and rejected all the imaginary words in the test. In this case we would want to say that you reliably recognised the real words, and, because these words are a sample from a set of 2000 words, we would probably want to say that you would be able to recognise reliably all 2000 words in the set.

Suppose, on the other hand, that you identified half the real words and rejected all the imaginary ones. In this case, we would want to say that you could probably recognise 50% of the 2000 word set, i.e. about 1000 words.

More interesting cases arise where people produce YES responses to imaginary words. Suppose, for example, you recognised all the real words, but also claimed to recognise half the imaginary words. In this case, we would want to argue that your score of 100% on the real words is too high; it needs to be reduced because your threshold for saying that you recognise a word is too low. The size of the adjustment depends on the number of IY responses you make - obviously, if you make lots of IYs, then your acceptance threshold is very low and you are likely to produce RY responses by chance.

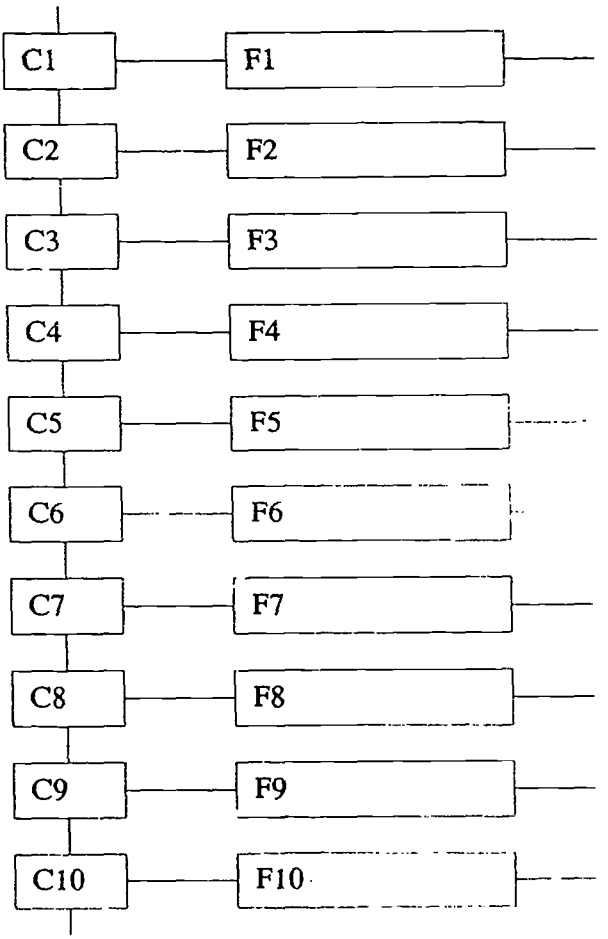
The mathematics of all this is not too difficult. In the 1950s, the Navy carried out a great deal of research on how well ASDIC operators could identify enemy submarines. They were interested in three types of behaviour: times when an operator correctly identified a submarine that was actually there; times when an operator failed to identify a submarine that was actually there; and times when an operator identified a submarine that didn't actually exist. You will see that there is an obvious parallel between these three situations and the RY IY and RN responses described above; all that is necessary is to replace "submarines" by "French words". The mathematical model devised to handle the submarine situation (signal detection theory) should also apply to our vocabulary recognition task.

The test which we devised for Eurocentres uses this basic principle, but is rather more complicated than the test outlined above. A schematic version of our test is shown in Figure 1. Basically, our test is divided up into a number of levels, each corresponding to a frequency band of 1000 words. The first part of the test starts off at the highest frequency band, and assesses how many of these words a testee can be deemed to know by sampling 10 real words and 10 imaginary words. If the testee scores highly on this band, s/he is tested on the next

**Figure 1 The structure of the test files**

Coarse files (C1-C10) contain 10 items from the bottom end of a specified frequency band. The testee moves through the coarse files in turn, until her performance is too poor to allow her to continue, or until she successfully completes the final file C10.

Fine files (F1-F10) contain 50 items from the specified frequency range, and thus allow us to test explicitly one word in twenty. Once the testee finishes a fine file, her total vocabulary score is calculated.



band, and this process continues until performance drops below a preset threshold. At this point, the test works out a rough estimate of how many words we think the testee knows, and tests a further fifty words from the appropriate frequency band. So, suppose our testee scores 100% on bands 1-4, but only 20% on band 5, the program reckons that the testee knows somewhere betwee 4 and 5 thousand words, and does its detailed testing on band 5. The detailed testing phase actually tests one word in twenty at the appropriate level.

**Assessment**

So far we have run three versions of this test with about 250 students from a wide range of language backgrounds, 109 at the Cambridge Eurocentre School, and two groups totalling 158 in London. For practical reasons, we have mainly been interested in correlating the results of our test with the results of the Eurocentres JET test - i.e. we are interested in establishing how far our Vocabulary test can be used as an alternative to JET. The results of this work are summarised in Table 2.

**Table 2**
Correlations between the Vocabulary Test and JET

| | | | | |
|---|---|---|---|---|
| 1: CAMBRIDGE | 109 testees | overall correlation | | .664 |
| | | subgroups: | French | .549 |
| | | | German | .807 |
| | adjustments: | 4 out of 5 | | |
| 1: LONDON | 159 testees | overall correlation | | .717 |
| | | subgroups: | French | .556 |
| | | | Italian | .792 |
| | | | Spanish | .723 |
| | | | Portuguese | .756 |
| | | | German | .790 |
| | | | Non-IE | .735 |
| | adjustments: | 9 out of 14 | | |

There are a number of interesting points to note here. Firstly, the correlations between JET and VOC (the vocabulary test) are generally high: in fact, given the diverse nature of the tests, the results are surprisingly high. Obviously, the correlations are not perfect, but given that JET is itself unsatisfactory in some ways, this is only to be expected. More interesting is the fact that the correlations vary slightly for different language groups. In general, correlations for homogeneous language groups are better than correlations for mixed groups, and

some linguistic groups produce very high levels of correlation indeed. This is not always the case, however. With the French speakers studied here, the correlations between VOC and JET are consistently low. At the moment, we don't really know how to interpret these differences. One possible explanation is that the VOC test in its present format is systematically biassed against speakers of particular languages, but it is equally possible that the VOC test is accurate, and that the JET test is biassed in the same way. Some evidence for this latter view comes from another study (Meara and Buxton 1988) in which very high levels of correlation between a VOC test and a more traditional M/C test were found with French speakers.

A further check on the effectiveness of the VOC test as a placement indicator comes from adjustments made to class registers one week after the original placement by JET. In the Cambridge study (109 Ss) five students were reallocated to a different group on the basis of their actual performance in class. Four of these cases were moved up to a higher level than their original placement, and in every case this move was in line with the placements produced by VOC. In the London trials (159 Ss), a questionaire was used to assess major discrepancies in the placements produced by JET. This trawl produced 14 cases; in nine of these cases, teachers' assessments agreed with the VOC score rather than with the JET score. Not surprisingly, if these cases are excluded from the data, the overall correlation between JET and VOC increases.

**Conclusion**

This paper has described a relatively small-scale study which uses a measure of vocabulary size as a way of placing students at the start of their course. The data that we have presented suggests that the test works well, though obviously a great deal more work will be needed before we can claim it is thoroughly reliable. The test in its present format, for example, is basically a test of visual familiarity, and it assumes that recognition of a word form is an adequate test of word knowledge. This assumption is clearly one that needs to be probed carefully. Obviously, formal recognition is necessary but not sufficient for word knowledge, but by relying on recognition, the test probably over-estimates true vocabulary knowledge. Whether this really matters or not is anybody's guess: it could be, for example, that passive recognition vocabulary is generally closely related to the size of a learner's active vocabulary, and that a more accurate estimate of vocabulary size could be obtained by suitably adjusting the raw scores found on the VOC test. Another problem arises from the use of imaginary words. The current version of the test uses imaginary words which are very carefully constructed so that they share the physical characteristics of the real words in the same set. However, it is clear to us that some of the imaginary words are easier to handle than others: some can be rejected instantaneously,

while others cause even native speakers of English to puzzle for a long time. We also think that some imaginary words cause difficulty to speakers from particular language backgrounds. Again, we don't know why this should be, but the problem is one that can easily be solved by further work.

At the moment, then, the best we can say is that the work we have done looks very promising, and if further developments live up to these promises, then it looks as though the tedious and time-consuming task of placing students at the start of a course could be greatly simplified and stream-lined. A small contribution to "applied linguistics in society" perhaps, but one that will be welcomed by many teachers.

However, the VOC test has other advantages, besides these practical ones. One major advantage from the research point of view is the speed with which the VOC test can be administered. Since it only takes ten minutes, there is no reason why it should not become a standard tool for assessing subjects in empirical research. At the moment, the research literature uses only vague labels for describing people who take part in research: "50 first certificate students", "25 students following a pre- university course at Stanford", or "150 air-force pilots" are typical examples of this sort of labelling. Clearly they are not very informative; it would be much more helpful to be told that we were dealing with, say 150 air-force pilots who scored a mean of 4500 on the VOC test with a standard deviation of 50 words. The fact that the VOC test is so quick to administer makes this kind of standardisation a real possibility.

The VOC test is also interesting because it opens up areas of research which have not been accessible before. If the VOC test really does measure vocabulary size, then we can begin to ask questions like these:

how fast do people learn new words?
how much individual variation is there in this skill?
is it affected by other variables, such as L1, or L1 vocabulary size?
how effective are different types of teaching program? e.g. do intensive courses produce more vocabulary learning than less intensive eight-week ones?
how quickly do learners who don't practice lose their vocabulary?
is the fall-out rate such that it reaches a stable assymptote? i.e.
is there a residue of words that you never really forget, no matter how little you practice?

These are questions, that we hope to address in the near future.

To sum up, then, the VOC started out as a practical research problem aimed at providing a solution to an organisational problem. In R and D circles, it is common to hear people talking about the practical spin-offs from theoretical re-

search: the VOC test seems to be a clear case of theoretical spin-offs from the practical research. Maybe the real future of Applied Linguistics lies down this road?

## BIBLIOGRAPHY

Anderson, R and Freebody, P (1981) Vocabulary Knowledge. In J Guthrie (Ed), *Comprehension and teaching: research reviews.* Newarke, DE.: International Reading Association.

Meara, P and Buxton, B (in press) An alternative to multiple choice vocabulary testing, *Language Testing.*