Picture Vocabulary Size Test (Web version)

Version 1.00

Released January 5, 2025

Help Guide

Paul Nation and Laurence Anthony

OVERVIEW

The Picture Vocabulary Size Test (PVST) is a test of receptive vocabulary size. The test measures whether the test-taker can choose a suitable meaning (a picture) for a given partly contextualised word form. It is a recognition test. It is primarily intended for young pre-literate native speakers up to eight years old and young non-native speakers of English, but it can be used with adults. The test was designed by Paul Nation and programmed by Laurence Anthony. Jannie van Hees played an important role in trialling the test.

There are currently two items banks available for the PVST (Web version), each following the same design procedures.

A desktop version of the PVST is also available. If the PVST is to be used for serious research and with young learners, the desktop version is recommended, because it provides additional presentation features (along with a 6,000 word item bank) as well as recording of the data. It is available from Laurence Anthony's web site (https://laurenceanthony.net/software). The desktop version has its own specifications document in the Help section of the program, and that should be read when working with young learners.

USING THE TEST

Purpose and limitations

The PVST is intended to be used for formative assessment, particularly diagnostic assessment, where performance on a task such as reading or participating in content-focused oral discussion may be affected by vocabulary size. Because the PVST is a largely decontextualised measure of receptive vocabulary knowledge to some degree isolating vocabulary knowledge from other skills, learners' performance on the test can be used as an indication of whether vocabulary knowledge is likely to be a major factor accounting for poor performance on the task. Thus, the test could be used along with other diagnostic procedures for examining learners' reading, listening, speaking, and writing skills.

The test is based on the most frequent 20,000 word families of English.

Test administration and instructions

The test can be found at https://laurenceanthony.net/software/pvstweb/. If you wish to have the written version of each sentence appear, click "Show audio transcript" at the beginning of the test. The learner hears a sentence and then has to click on one picture

from four pictures which best matches the spoken sentence. A red line appears around the picture. The learner then clicks next to move to the next screen.

The learner can listen to a sentence again by clicking on *Play audio* at the top of the screen. The learner can click on a different picture to change a choice. Clicking on *Back* takes the learner back to the previous test item.

Using the test with native speakers of English

The rule of thumb for estimating the vocabulary size of native speakers of English up to the age of around fourteen or fifteen years is to use the formula of age minus 2 times 1000. So, a six-year-old native speaker is likely to have a vocabulary size of around 4000 words (6 minus 2 times 1000 = 4,000). A thirteen-year-old native speaker is likely to have a vocabulary of 11,000 words. These estimates can vary up or down by 1000 to 2000 words. See Nation & Coxhead (2021) *Measuring Native-Speaker Vocabulary Size* for much more information on this.

The rule of thumb can be a rough guide to determine when to finish the test when using it with young native speakers. Adult native speakers can try sitting the whole test which tests knowledge of the most frequent 20,000 words.

Using the test with non-native speakers of English

The test can be used with older learners of English as a foreign language or second language, especially those with literacy problems, but the test should be stopped by clicking *Finish* when learners are obviously dealing with words they do not know. For intermediate learners this is likely to occur around the two to four thousand word levels (around items 10 to 20). The PVST tests five items at each 1000 word level up to the 20th 1000 word level, making a total of 100 items (5 times 20) in the test.

Increasing validity through test administration

One-on-one administration of the test makes sure that learners remain engaged and do not employ random guessing. When administering the test, it is useful to make sure that learners are all aware of the following, so it is worth having a little checklist to make sure that each learner knows what to do.

1   They should look at all four pictures when considering choices.

2   They know they can push the *Play audio* button to hear a sentence again.

3   They can change their mind about a choice, by clicking on a different picture.

Each of the three procedures should be pointed out to each learner with a demonstration, and the administrator should be watching each learner to ensure they perform each procedure where needed.

Coxhead et al (2014, 2015, 2018) carried out several studies using the Vocabulary Size Test (not the PVST) at a New Zealand university and at New Zealand secondary schools, and their findings should roughly match the results of the PVST. These articles are available on Paul Nation's resources pages under Publications. Coxhead, Nation, Woods and Sim (2018) found a substantial increase in scores on the Vocabulary Size Test for around 25% of secondary school students when a supervisor sat next to them and kept them on task as they sat the test.

Where learners respond too quickly, the administrator should ask them if they looked at all four pictures, encourage them to do that, and praise them for doing it.

Interpreting test results

Each test word represents 200 words in the source lists (the BNC/COCA lists on Paul Nation's resources pages) See Nation (2016) for more information on the lists. So, a learner's score on the test is multiplied by 200 to get their total vocabulary size. Thus, a learner with a score of 20 has a receptive vocabulary size of 4000 word families. The program makes this calculation in the results it produces.

For non-native speakers of English, vocabulary size can be matched with the Council of Europe CEFR guidelines (see Paul Nation's adaptation of the CEFR guidelines under Vocabulary Lists on his web resources pages). Elementary learners know a maximum of 1000 word families (a score of 5 or lower on the PVST). Intermediate learners range from 1000 to 4000 word families (a score of 20 on the PVST). Advanced learners range from 5000 word families up to 9000 or more word families. Learners with a vocabulary size of around 5000 words can successfully do undergraduate study at an English speaking university, but they will find it hard. Non-native PhD students with a vocabulary size of 9000 words can successfully complete their courses and theses. A larger vocabulary size however is clearly an advantage.

Because the test is a sensitive receptive recognition test, it gives credit for partial knowledge through the use of choices. This means that getting a word right on the test does not necessarily mean that the word is well known.

Test construct

Analysed using Read and Chapelle's (2001) framework, the PVST is a discrete, selective, relatively context-independent vocabulary test presented in a multiple-choice format. Test-takers are required to select the best picture for each word from four choices.

The PVST is designed to measure both first language and second language learners' spoken receptive vocabulary size in English. The test measures knowledge of the spoken word form, the form-meaning connection, and to a smaller degree concept knowledge. The four choices are not closely related in meaning to each other. This makes it possible for a learner to answer items in the test with a partial knowledge of the words involved.

Inferences: Although the tested words are presented in simple non-defining contexts, the test is essentially following a trait-definition of vocabulary which means that vocabulary knowledge is tested independently from contexts of use. At the item level, the test measures receptive knowledge of a spoken word form. At the test level it provides an estimate of total vocabulary size where vocabulary knowledge is considered as including only single words (not multiword units) and vocabulary size does not include proper nouns, transparent compounds, marginal words like *um, er, gee gosh,* and acronyms. It does not measure the ability to distinguish homonyms and homographs.

The test does not measure skill in vocabulary use, such as productive spoken use or receptive use in reading. That is, learners may know particular words enough to indicate what they mean, but may not know them well enough to use them either

receptively or productively. If the test shows that learners know the words in the test, then their learning focus needs to be on using the words rather than on focusing on the words themselves.

Uses: For instructional purposes the results can be used to guide syllabus design, choosing texts for extensive reading, and vocabulary instruction. For research purposes, the test can be used as a measure of total receptive spoken vocabulary size. Low proficiency learners of English as a foreign language will be better off taking one the recent Vocabulary Levels Tests (available on Paul Nation's resources pages.

Impacts: If it is used as intended, it is a relatively low stakes test for learners. One consequence may be that it underestimates the vocabulary size of learners who are not motivated to perform to the best of their ability, especially if they are judged to be low achievers within their education system. This could result in faulty instructional decisions being made about their vocabulary learning needs, and thus the test should be administered to such students on a one-to-one basis with the administrator present to keep the learner on task.

Test format

The test contains 100 items, each presented in a four-choice multiple-choice format.

Each test word appears in a non-defining sentence where the context uses only words from the first 500 word list. The contexts provide a small degree of orientation towards the meaning of the word for the learners, in that they indicate a particular sense of the word and its part of speech. The sentences are single clause sentences less than six words long. Although the first 500 words are tested, it is assumed that the learners sitting the test know the words which make up the context sentences.

There is growing criticism of the multiple-choice format for testing vocabulary knowledge for two major reasons. Firstly, by providing choices rather than requiring the learner to supply a response, it does not closely represent the kind of vocabulary knowledge needed for listening or reading. This disadvantage of multiple-choice was weighed against the practicality of multiple-choice when designing the test, but deserves further research. Secondly, the use of choices allows the possibility of getting items correct through guessing which may inflate scores. This effect can be reduced by finishing the test when learners get several consecutive items wrong.

Like other test items, multiple-choice items can be answered using a range of strengths of knowledge of a word. This range goes from only being able to select a meaning from a small number of choices not closely related to each other to having a strong productive knowledge of a word. This means that for some words answered correctly, the learners are at an early stage of developing knowledge of these words. Unfortunately, the data from one test item format will not be enough to show which particular words are well known, which are moderately well known, and which are only partly known. Test users need to be aware that although every correct answer on the test is given a 1 point score, a learner's knowledge of each word is not necessarily of equal strength.

MAKING THE TEST

The 20,000 PVST was built on the original 6,000 PVST (see Laurence Anthony's web site for the 6,000 versions). Both function words and content words were chosen. The

distractors were chosen from the same frequency level as the test words. The BNC/COCA lists were used for the 20,000 version.

The pictures for the tests were bought from the website www.shutterstock.com. A very small number of pictures were bought from www.dreamtime.com.

The pictures chosen are largely photos. Where items were homographs, the most frequent meaning was chosen, guided by Kevin Parent's (2012) research.

The criteria used were in order of importance:

1   The photo most clearly represented the meaning of the word
2   Pictures containing young children and people from a range of racial groups were given preference
3   Photos were preferred over diagrams or cartoons.

TRIALING THE TEST

The two early 6000 versions of the tests were initially trialled with adults to make sure they could get all the items correct. Then a ten year old and a seven year old sat the tests, experiencing no difficulty with the test format and having difficulty with only a few of the photos (*shabby, conquest, battered*) which were then changed for clearer photos. They were asked about several of the items as they sat the test to check that right and wrong answers reflected their knowledge and not any misleading aspects of the photos or sentences.

The changes made to the test as a part of its development are detailed here as one aspect of the tests' validity. Some context sentences were changed that might contain a give-away word which is not the test word. For example, "He is a good chap" testing "chap" was the only picture of a person in the set of four, so "He" gives away the answer. The change was "This is a good chap". "It's a large flock" was replaced by "It's a flock", as "large" was helping to eliminate possible choices. Similarly, "He's working independently" was replaced by "He's doing it independently", as "working" helped eliminate choices. We changed distractors that could have been correct answers. For example, one of the distractors for *crimson* had a part of the picture coloured crimson. The distractor was replaced.

We replaced one or two items that might upset young children - *nightmare, demon*. We replaced a few pictures to get closer to what children knew of the word - *limb* (tree to leg), *function* (vending machine to computer), *coordinate* (from a meeting to two children jumping), *shabby* (from a young man dishevelled with a tie to an old man with rough clothes).

The following words could not be satisfactorily pictured in PVST 1 and were replaced (replacement in brackets): *selection* (*electric*), *resume* (*thistle*), *mature* (*objected*).

WEAKNESSES OF THE TEST

The format of the test means that it has several unavoidable weaknesses.

1   Some choices do not require detailed knowledge of the word, that is, they require just enough knowledge to distinguish it from the distractors. For example, *savage*.

2 Some items require a greater degree of world knowledge than other items. For example, the item testing *flit* requires the learner to work out that a butterfly moves in this way, or that a meeting may involve an *agenda*. Items like *table* or *grass*, require no such inferencing. Surprisingly even the five year olds seemed to have no great difficulty performing this inferencing. Some children spontaneously thought aloud while they sat the test. One of the distractors in the set containing *earn* was a picture of a knight. One child said as he choose that for *earn* said, "He must have done a lot of work". He changed his answer when he looked more closely at the other pictures. Another said regarding *hobby*, "Playing the guitar is a hobby". Yet another said for *tour*, "Buses take you on tours".

3 A small number of the contexts contain other content words and so understanding the item at least partly depends on understanding the words in the context (for example, *see* is a context word in *She saw the horizon*, *full* is a context word in *full of anguish*). To reduce this problem, all context words are in the first 500 words of English.

4 Learners sitting multiple-choice tests can benefit from applying test-taking strategies. One of the most common was to eliminate the alternatives. This then resulted in fewer choices to focus on. Such a test-taking strategy does not undermine the validity of the test, because it indicates close engagement with the test and a thoughtful search for the answer.

5 The multiple-choice item format allows guessing. This could inflate scores.

6 The unit of counting when making the word lists was Level 6 word families. This could result in the test underestimating vocabulary size because high frequency derived forms which young learners might treat as different words from their stem form were not tested.

Adults sitting the test inevitably criticise some of the choices of pictures even though they can see which answer is correct. However, it is worth checking to see if young learners have a problem as the test has been primarily designed for them and was trialled with them.

CITING THE TEST

Use the following method to cite/reference the software according to the APA style guide:

Anthony, L. & Nation, I.S.P (2025). Picture Vocabulary Size Test - Web (Version 1.00) [Computer Software]. https://www.laurenceanthony.net/software/pvstweb/

REFERENCES

Coxhead, A., Nation, P., & Sim, D. (2014). Creating and trialling six versions of the Vocabulary Size Test. *The TESOLANZ Journal*, *22*, 13-27.

Coxhead, A., Nation, P., & Sim, D. (2015). The vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, *50*(1), 121-135.

Coxhead, A., Nation, P., Woods, L., & Sim, D. (2018). Group vs. individual: How

administration procedure can affect vocabulary test scores. *New Zealand Studies in Applied Linguistics*, *24*(2), 24-41.

Nation, I. S. P. (2016). *Making and Using Word Lists for Language Learning and Teaching*. Amsterdam; John Benjamins.

Nation, I. S. P. & Coxhead, A. (2021). *Measuring Native-Speaker Vocabulary Size*. Amsterdam: John Benjamins.

Nation, I.S.P. & Anthony, L. (2016). Measuring vocabulary size. In *Handbook of Research in Second Language Teaching and Learning*, Volume III, E. Hinkel (Ed.) New York: Routledge.

Parent, K. (2012). The most frequent English homonyms. *RELC Journal, 43*(1), 69-81.

Read, J., & Chapelle, C. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 3-32.