

New General Service List Test: Updated Documentation

Tim Stoeckel stoeckel@unii.ac.jp

Phil Bennett philbennett59@gmail.com

Tomoko Ishii ishii@gen.meijigakuin.ac.jp

November, 2016

This paper provides a brief description of several versions of the New General Service List Test (NGSLT), an instrument designed to assess written receptive knowledge of words on the New General Service List (NGSL) (Browne, 2013). It is an updated version of earlier documentation.

Test Description

Three- and Five-Level Formats

Each version of the test is a “levels test,” meaning that it is designed to determine whether examinees have mastered frequency-based levels of the NGSL. There are both five- and three-level variants of the test.

In the five-level format, each level consists of 20 items and assesses a separate 560-word frequency-based band of the NGSL. Though it is common for second language (L2) vocabulary tests to be divided into 1,000-word frequency levels, these smaller 560-word bands are used for two reasons. First, it is not uncommon for learners in EFL contexts to have large gaps in knowledge of high frequency words, and by testing smaller groups of words, the point at which learners no longer have mastery of the majority of words can be identified with greater precision. This enables teachers and students to design plans of study which address individual student needs with more accuracy. Second, one complete 560-word band is a reasonable goal for a semester of study, particularly considering that most non-beginners will have knowledge of some words in each level, meaning the learning burden would be somewhat less than 560 words. Among the five-level versions of the test, there are both monolingual and Japanese-English bilingual variants, each of which is described in Item Format below.

In the three-level format, each level consists of 30 items and assesses a separate frequency-based band of approximately 1,000 words. This format corresponds with the way the NGSL has been divided on the Lextutor website (<http://www.lex tutor.ca/>) and would therefore be useful for teachers who use Lextutor and want to make sure learning materials are suitable for their students’ level of vocabulary knowledge.

Item Format

Test items utilize a multiple-choice format similar to that used in the well-known Vocabulary Size Test (VST) (Nation & Beglar, 2007). Because word forms in the NGSL itself are grouped into modified lemmas (for details see Browne, 2013), each test item represents one modified lemma.

Item stems. Each item stem consists of a target word followed by a sentence which uses the word in a non-defining context. To reduce the likelihood of examinees incorrectly answering items testing words that they do in fact have some knowledge of, three steps were taken, each based on frequency counts from the Corpus of Contemporary American English (COCA)

(<http://corpus.byu.edu/coca/>). First, for modified lemmas consisting of more than one part of speech (e.g., *approach* as a noun and as a verb), the more frequently occurring part of speech is tested (the noun form, in the case of *approach*). Second, the form of the word used in the example sentence in the item stem is the most frequently occurring form for the tested part of speech. Third, when this word form has more than one meaning or use, the sample sentence and correct answer are based on the most common meaning or use (e.g., *approach* is more frequently used to mean 'method' than 'coming closer'). Regarding language, English is used in the item stems for both monolingual and bilingual formats.

Answer choices. Each item has four answer choices which include three distractors and the correct answer. In the monolingual versions of the test, answer choices are written with simplified syntax and are almost always expressed in vocabulary that is of higher frequency than the tested word. These answers offer a simple definition of the tested word as well as definitions of three other words that are of similar frequency to the tested word.

In the bilingual versions of the test, answer choices are given in examinees' first language (L1). These variants are intended to minimize the risk of test results conflating knowledge of tested words with understanding of answer choices written in examinees' L2. In most cases, the correct answer is a direct translation of the tested word, and the distractors are three other L1 words. However, when the tested word appears in the L1 as a phonologically similar cognate or loanword from English, short definitions are used instead of direct translations. For example, the correct answer for the item testing *approach* in the Japanese bilingual format is やり方, rather than アプローチ. This is to prevent correct answers on unknown words from phonological matching (for details see Stoeckel, Ishii, & Bennett, 2016).

Unique Test Forms

The six available variants of the NGSLT are listed below. Because these instruments occasionally undergo minor revisions as a result of ongoing item analysis, the date of the most recent revision is given in parentheses. This is the version currently available here on Academia (and older versions have been removed to avoid confusion).

Five-Level Monolingual:

- Form A (March, 2016)
- Form B (March, 2016)
- Form C (July, 2016)

Five-Level Japanese-English Bilingual:

- Form A (March, 2016)
- Form B (May, 2016)

Three-Level Monolingual:

- Form A (June, 2014)

Five-level monolingual forms. The five-level monolingual Forms A, B, and C each assess a unique set of 100 words. These words, 300 in total, were randomly chosen from the NGSL, 60

from each 560-word frequency band. For Form A, all items at each test level were selected to represent the range and average item difficulty of items in a larger item bank for each test level. This process was followed for most items in Forms B and C while other items (for which difficulty values had not yet been established) were distributed randomly between the two test forms. Equivalency of Forms A, B, and C has yet to be established. However, the process of test development described here suggests that the three forms can be considered of similar difficulty for low stakes diagnostic purposes.

Five-level bilingual forms. The five-level bilingual Forms A and B assess knowledge of the same words that appear on the corresponding five-level monolingual test forms. Equivalency of these forms has yet to be established.

Three-level monolingual form. The items used on this test form belong to the same bank of items that is used for the five-level monolingual forms, with considerable overlap between this form and the five-level monolingual Form A.

Piloting

The Table below shows descriptive statistics from piloting for five of the test forms.¹ Though Cronbach alpha values display considerable range ($\alpha = .77$ to $.97$), estimates of reliability are partially dependent upon the range of scores in any given sample. We suspect this, rather than any important qualitative differences among the test forms, explains the difference in estimates of reliability reported here. For the variant with the highest alpha value (Japanese-English Bilingual Form A, $\alpha = .97$), for instance, the sample included learners in all three years of high school as well as university students. On the other hand, the version with the lowest alpha value (Japanese-English Bilingual Form B, $\alpha = .77$) was piloted with a sample consisting only of university students, many of whom achieved scores near the test’s ceiling. (For more information regarding the psychometric properties of the tests see Stoeckel & Bennett, 2015 [monolingual] and Stoeckel, Ishii, & Bennett, 2016 [bilingual].)

Table: Descriptive Statistics for Different Forms of the New General Service List Test

	Items	<i>n</i>	<i>M</i>	<i>SD</i>	High	Low	α
Five Levels							
Monolingual							
Form A (March, 2016)	100	54	67.4	11.1	97	46	.87
Form B (March, 2016)	100	57	70.0	9.3	92	48	.84
Form C (July, 2016)	100	50	83.8	7.3	94	67	.83
Japanese-English Bilingual							
Form A (March, 2016)	100	382	80.1	18.1	99	25	.97
Form B (May, 2016)	100	304	89.2	5.6	99	70	.77
Three Levels							
Monolingual							
Form A (June, 2014)	90		<i>(no available data)</i>				

¹ The term “piloting” here refers to the initial administration of a test *form* in its current configuration. Many of the individual *items* in the monolingual test forms have already been through several cycles of use, careful analysis, and – when warranted – revision.

Test Administration

The NGSLT should be administered without any time restrictions. (For purposes of planning, most students are able to complete the test in 20 to 30 minutes.) Though some researchers (e.g., Zhang, 2013) have explored the use of an “I don’t know” (IDK) answer choice in multiple choice tests of vocabulary knowledge to reduce the use of test strategies or blind guessing, it is best not to use IDK because there is evidence that it introduces a “willingness to use I don’t know” as a non-relevant construct affecting test scores (Stoeckel & Stewart, 2016, September). The use of test strategies or blind guessing can certainly increase scores (Gyllstad, Vilkaitė, & Schmitt, 2015), but this is less problematic with levels tests in which the purpose is not to estimate vocabulary size and where the threshold for successful mastery of a given level is set sufficiently high.

Score Interpretation and Practical Use

The NGSLT assesses written receptive vocabulary knowledge, the kind of lexical knowledge needed for reading. It does not provide information regarding examinees’ understanding of vocabulary in listening or of their ability to use English vocabulary productively. Furthermore, the primary aspect of lexical knowledge that is assessed with the NGSLT is whether examinees have made a basic form-meaning link for the tested words. For diagnostic purposes, a good way to use the NGSLT is to examine each learners' scoring profile across all levels of the test in order to identify the point at which they no longer have mastery of around 80 to 85% of the words in a test level¹. Because of the importance of high frequency vocabulary, this is a good place for learners to target their intentional vocabulary study. To that end, there are a several spaced-repetition flashcard programs and other self-study materials available in the “Tools” section of the NGSL website (<http://www.newgeneralservicelist.org/>).

Further Developments

In addition to initial diagnostic testing, repeated assessment is useful for seeing whether students have met their learning goals and whether new goals should be established. It is for this purpose that parallel forms of the five-level monolingual and Japanese-English bilingual versions of the test have been developed. We are in the process of completing a large enough item bank for four complete monolingual test forms and four complete Japanese-English bilingual test forms. We will then establish difficulty values for all test items and reassign them as necessary to create four forms of equivalent difficulty. To help with this and with ongoing item refinement, we are always grateful to teachers who are willing to share anonymous test results with us. Kindly contact any of us at the email addresses above.

References

- Browne, C. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 37(4), 13-16.
- Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal for Applied Linguistics*, 166(2), 278-306. doi:10.1075/itl.166.2.04gyl

¹ This threshold is based on Milton (2009), who found that it is common even for highly proficient learners to achieve average scores of about 85%, and not 100%, in tests of high frequency words.

- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Stoeckel, T., Ishii, T., & Bennett, P. (2016). *A Japanese-English Bilingual Version of the New General Service List Test*. Manuscript submitted for publication.
- Stoeckel, T., & Bennett, P. (2015). A test of the New General Service List. *Vocabulary Learning and Instruction*, 4(1), 1-8. doi: 10.7820/vli.v04.1.stoeckel.bennett
- Stoeckel, T., Stewart, J. (2016, September). *The "I don't know" option and L1 answer choices: A comparison of four variants of the Vocabulary Size Test*. Paper presented at Vocab@Tokyo 2016.
- Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly*, 47(4), 790-811. doi: 10.1002/tesq.98