Vocabulary Learning and Instruction
Volume 4, Issue 1, October 2015
http://vli-journal.org

Article

A Test of the New General Service List

Tim Stoeckel[a] and Phil Bennett[b]
[a]*University of Niigata Prefecture;* [b]*Miyazaki International College*

doi: http://dx.doi.org/10.7820/vli.v04.1.stoeckel.bennett

Abstract

This paper introduces the New General Service List Test (NGSLT), a diagnostic instrument designed to assess written receptive knowledge of the words on the New General Service List (NGSL) (Browne, 2014). The NGSL was introduced in 2013 as an updated version of West's (1953) original General Service List. It is comprised of 2,800 high frequency headwords plus their inflected forms and is designed to provide maximal coverage of modern English texts. The test introduced here is divided into five 20-item levels, each assessing a 560-word frequency band of the NGSL. Using a multiple choice format, the NGSLT is intended to assist teachers and learners in identifying gaps in knowledge of these high frequency words. Data from 238 Japanese university students indicate the NGSLT is reliable (α = .93) and that it measures a single construct. A comparison of NGSLT and Vocabulary Size Test (Nation & Beglar, 2007) scores for a small group of learners shows that the NGSLT provides more detailed diagnostic information for high frequency words and may therefore be of greater pedagogic use for low and intermediate level learners. Ongoing developments include parallel versions of the NGSLT as well as a separate instrument to assess knowledge of the New Academic Word List. Both the NGSLT and New Academic Word List Test are freely downloadable from the NGSL homepage (www.newgeneralservicelist.org).

1 Introduction

The impact of vocabulary knowledge on more general language proficiency is now widely acknowledged. Studies by Stæhr (2008) and Milton, Wade, and Hopkins (2010) have shown healthy correlations between written and aural receptive vocabulary size and tests of the four main language skills. Additionally, the findings of corpus-driven analyses (Hanks, 2013; Hoey, 2005) are now putting vocabulary knowledge at the heart of the language learning process by revealing the extent to which lexical choice influences and determines syntactic structure in the creation of meaning.

Accordingly, vocabulary assessment can fulfill important roles in language education for teachers, learners, and researchers. In classroom settings, having reliable estimates of vocabulary knowledge enables teachers to provide suitable materials for learners' needs, to judge the efficacy of a course of study, and to set appropriate goals for further development. Vocabulary goal-setting is particularly important, as successful learners will have a desire to build vocabulary, an

Please cite this article as: 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

awareness of the particular words that are most likely to benefit them, and the capability to achieve their goals (Dörnyei, 2005; Nation, 2001). For research purposes, vocabulary tests could be used to better understand the relationship between lexical knowledge and other skills, to assess the impact of learning experiences on lexical development, and to measure lexical growth.

Vocabulary tests are typically based on the frequency model – the notion that the more often a word is encountered, the more likely it is to be known. While other factors, such as polysemy, cognate status, and orthographic or phonological form can affect the difficulty of individual words, for 60–80% of learners, it seems that knowledge of bands of words grouped by frequency generally decreases as the bands become less frequent (Brown, 2012; Milton, 2009). Additionally, high-frequency words offer a disproportionally high percentage of text coverage, making them important to prioritize over mid- and low-frequency words. Thus, the frequency model remains a useful principle in developing vocabulary assessment tools and word lists.

2 The General Service List and New General Service List

Since frequency has such an impact on vocabulary assessment, it is important that frequency counts are accurate and reflect modern usage patterns. For many years, the General Service List (GSL; West, 1953) was used to provide this information. The GSL was not based purely on frequency, although this was one of its criteria. This list contains around 2,000 headwords, with related forms grouped underneath into word families. Although the GSL was originally intended to aid the development of reading materials, it later became the basis for test development (Schmitt, Schmitt, & Clapham, 2001) or a prerequisite for further word lists (Coxhead, 2000). However, the rapid growth of large computerized corpora and revised opinions over what should be considered a "word" have made the GSL appear somewhat dated.

The New General Service List (NGSL; http://www.newgeneralservicelist.org) is an attempt to address these concerns while preserving the goal of the GSL, which is to include those items that provide maximal coverage of texts with as few headwords as possible. The NGSL is based on frequency and dispersion data in a 273 million-word sample of the Cambridge English Corpus. The sample contains texts drawn from fiction, journals, magazines, non-fiction, radio, documents, TV, spoken interactions, and learner output.

The NGSL also differs from the GSL in its grouping of words. The original GSL was inconsistent in how it categorized derived forms under each headword. Although an updated version of the GSL (Bauman & Culligan, 1995) brought more consistency to the organization of the list, it has been recognized that, for learners below advanced levels, knowledge of a headword does not guarantee understanding of all of the derived forms in a word family (Milton, 2009; Vermeer, 2004). The NGSL resolves this issue by grouping words into "modified lemmas." A lemma is simply a headword and its inflected forms, with different parts of speech belonging in separate lemma groups. The NGSL's modified lemma approach varies this slightly by considering a headword as all parts of speech with the same written form, and then including all of the various inflected forms. For example, the modified lemma for *stage* would include *stages*, *staged*, and *staging* as verbal inflections and *stages*, *staging*, and *stagings* as nominal inflections. While the NGSL contains around 400 more word families than the GSL, it has approximately 800 fewer lemmas and provides 6% more coverage of a sub-section of the Cambridge English Corpus (Browne, 2014).

3 Currently Available Tests

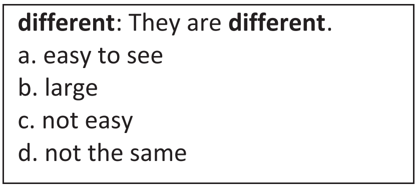
Two well-known vocabulary assessment instruments are the Vocabulary Levels Test (VLT; Nation, 1983; Schmitt, Schmitt & Clapham, 2001) and the Vocabulary Size Test (VST; Nation & Beglar, 2007), both of which assess written receptive vocabulary knowledge. The VLT provides diagnostic information for the first, second, fifth, and tenth 1,000-word frequency bands and has a separate section for the Academic Word List. The test uses a 6:3 matching format, with six answer choices provided from the same frequency band, three of which must be matched to given definitions.

The VST assesses learners' knowledge of the first 14 (or more recently, 20) 1,000-word frequency bands based on word families appearing in the British National Corpus. The test uses a typical four-option multiple-choice format and is intended to provide an estimate of overall vocabulary size, rather than a reliable indication of how well each frequency band is known.

4 The New General Service List Test

This paper introduces the New General Service List Test (NGSLT), a diagnostic instrument designed to assess written receptive knowledge of the words on the NGSL. This purpose differs from that of the VLT, which provides a broad estimate of learners' vocabulary profiles across selected, non-consecutive frequency bands, and from that of the VST, which estimates overall vocabulary size. The NGSLT also differs from these other tests in that it is based on an underlying list of modified lemmas rather than word families. As such, there are fewer assumptions associated with correctly answered items on the NGSLT. The NGSLT is intended to diagnose mastery of each of five 560-word levels of the NGSL and to identify gaps in knowledge of these high-frequency words. The size of each of these levels constitutes a manageable goal for one semester of intensive study, especially considering that non-beginners are likely to already know some words.

Test items follow the same specifications as those of the VST (Nation & Beglar, 2007). The stem consists of the target word followed by a sentence which uses this word in a non-defining context. Answer choices include three distractors



different: They are **different**.

a. easy to see

b. large

c. not easy

d. not the same

Figure 1. Example NGSLT item.

and the correct answer. When the target word has more than one possible meaning or use, the sample sentence and correct answer are based on the more common meaning or use as determined by consultation of concordance lines in the Corpus of Contemporary American English (http://corpus.byu.edu/coca/). Figure 1 depicts an example test item.

Items were written with high-frequency vocabulary. Whenever possible, items testing words in the first three bands were written only with words from the first two of these bands. Exceptions were the inclusion of *dirty*, *bottom*, *better*, and *repeat* in items which tested words in bands of the same or higher frequency as these words. Each item testing words in the fourth and fifth bands was written exclusively with words of higher frequency than the word being tested.

The test contains 100 items, 20 at each level. The items on the test were selected from a bank of over 200 items, each of which targets a randomly chosen member of the NGSL. These items have been piloted in Japanese colleges and analyzed to determine how likely learners are to know them (for details, see Bennett & Stoeckel, 2013). The items selected represent the range and average item difficulty at each level. As such, we are reasonably confident that the test is representative of the range of word difficulties at each level for Japanese learners.

We are currently collecting data for a validation study of the NGSLT, and preliminary results from 238 learners in four Japanese colleges are favorable. Taken as a whole, the test provides a reliable estimate of knowledge of the NGSL ($\alpha = .93$). The reliability coefficients for individual 560-word frequency bands are somewhat lower but acceptable, ranging from .70 to .80. Overall item quality also appears to be good. An inspection of point measure correlations as well as Rasch infit and outfit statistics has flagged just one item (Level 1: *teacher*) as misfitting. This misfit appears to have been caused by incorrect responses from three high-ability examinees. Data from a subset of our sample shows that NGSLT results also correlate with scores on the Test of English for International Communication (TOEIC), a test of general English proficiency: $r(33) = .72, p < .001$.

5 New General Service List Test Score Interpretation

For diagnostic purposes, one way to use the test is to examine learners' scoring profiles in order to identify the point at which they no longer have mastery of around 80–85% of the words in a band. This threshold is based on Milton (2009), who found that it is not unusual for proficient learners' scores to plateau at around this mark, rather than at 100% in tests of high-frequency words. Our own observations are consistent with this; students who have TOEIC scores of 800 or higher sometimes score as low as 75% on individual frequency bands on the NGSLT. Because high-frequency words are of such importance, it is probably beneficial for learners scoring less than 80% or so on a band to review a list of the complete word band (available at www.newgeneralservicelist.org) and to highlight unknown words in that band for further study. If teachers take this approach, they should exercise care in selecting a higher threshold for mastery, as learners could be presented with word lists that are not challenging enough to provide a suitable goal for a particular program of study (Dörnyei, 2005).

Table 1. A Comparison of TOEIC, VST, and NGSLT Results

			NGSLT (% score)				
			560-Word Level				
Examinee	TOEIC	VST[a]	1	2	3	4	5
1	345	2,400	**100**	**85**	80	70	55
2	265	2,400	75	65	40	45	30
3	485	3,400	**100**	**90**	**90**	75	55
4	405	3,400	**90**	80	**85**	75	70
5	310	3,400	**90**	75	70	50	55
6	240	3,400	80	70	65	45	40
7	950	6,200	**100**	**95**	**95**	**100**	**100**

Bold font indicates a score of 85% or higher on an NGSLT level
[a]Estimated vocabulary sizes based on the first eight levels of the VST (see McLean, Hogg, & Kramer, 2014, for reasons to administer abbreviated VST)

NGSLT results may be more informative pedagogically than VST scores, particularly for learners of low to intermediate proficiency. The VST supplies estimates of overall vocabulary size but not knowledge of individual frequency bands. This is illustrated in Table 1, which shows VST and NGSLT results for seven learners who are representative of a group of 33 students who took both tests. TOEIC scores are also shown in order to provide information regarding overall English proficiency. The first two learners obtained identical estimated vocabulary sizes of 2,400 word families with the VST, but NGSLT results show not only that each of these learners has gaps in knowledge of high-frequency words but also that their learning needs differ significantly. While learner 1 appears to have mastered the first two word bands, learner 2 could benefit from further study of the most frequent 560 words. Entries three through six show that even when VST results indicate knowledge of over 3,000 words, large gaps can persist among the most frequently occurring words. The last entry in the list illustrates the limitations of the NGSLT for diagnostic purposes with highly proficient learners.

The detailed profile of high-frequency vocabulary knowledge which the NGSLT provides can assist in pedagogical decisions such as appropriateness of extensive and intensive reading materials and vocabulary learning goals. Regarding the last of these, there is now a variety of useful study materials for the NGSL that can be accessed from the NGSL homepage (www.newgeneralservicelist.org/vocabulary-links/). Among these are free online flashcard programs which are divided into 50- or 100-word sets according to frequency level, and with results of the NGSLT, teachers and students can identify the most appropriate groups of words to study.

6 Future Research and Development

Though the NGSLT in its current state is capable of providing teachers and learners with valuable pedagogical information, there are ways to make it more useful. To minimize the risk of a testing effect when the instrument is repeatedly used in educational programs, parallel forms should be developed. To this end, we hope to release a second 100-item form later this year. It would also be beneficial to create bilingual versions of the test with answer choices in the L1. As a test of high-frequency vocabulary, the NGSLT is intended for learners of low to intermediate proficiency, whose answers under a monolingual format may at times reflect lack of understanding of lexical or syntactic elements of the answer choices rather than knowledge of the target word (Elgort, 2013; Karami, 2012; Nguyen & Nation, 2011).

For learners of higher proficiency levels in academic contexts, we have also recently completed and begun piloting a two-level, 40-item test to assess written receptive knowledge of the New Academic Word List (NAWL; available at www.newgeneralservicelist.org/vocabulary-links/). This list was developed by the authors of the NGSL and is comprised of words that frequently occur in academic texts and that are not included in the NGSL (www.newacademicwordlist.org). The NGSL and NAWL tests are identical in format, meaning they can be used together seamlessly. For learners who have previously demonstrated partial mastery of the NGSL, teachers can easily tailor diagnostic tests to include only the upper levels of the NGSL plus the NAWL.

Our use of a multiple-choice format identical to that of the VST was intentional. Though it is becoming clear that score interpretations may need to be revised to reflect the possibility of correctly guessing unknown words (Stewart, 2014), the multiple-choice format is familiar to students, requires more than simple recognition of word form, and is quick and easy to mark. More importantly, the creation of a large pool of items written to the same specifications opens up the possibility of first calibrating these items to a single scale under item response theory, and then using them flexibly to assess different areas of lexical knowledge as needed.

Finally, further research of the NGSL itself is also needed. As stated, when compared to the original GSL, the NGSL has been reported to provide approximately 6% better coverage (84.24% versus 90.34%) of a sub-section of the Cambridge English Corpus (Browne, 2014). Though this is an impressive advantage, it must be remembered that the criteria for selecting members of the NGSL was derived from this same corpus, so we might expect superior coverage. Though Browne (2014) has also reported notable improvements in coverage with corpora of texts from *Scientific American* and *The Economist*, coverage of more general sources needs to be more thoroughly investigated. In addition, the degree to which the construct *modified lemma* is useful pedagogically needs to be examined. There is evidence that when compared to the word family, it is a step in the right direction because fewer assumptions are made regarding knowledge of related word forms (Gardner, 2007). However, the degree to which knowledge of a modified lemma's headword can be associated with all of its constituents has not yet been established.

References

Bauman, J., & Culligan, B. (1995). *The general service list*. Retrieved from http://jbauman.com/aboutgsl.html

Bennett, P., & Stoeckel, T. (2013). Developing equivalent forms of a test of general and academic vocabulary. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings* (pp. 636–644). Tokyo: JALT.

Brown, D. (2012). The frequency model of vocabulary learning and Japanese learners. *Vocabulary Learning and Instruction*, *1*(1), 20–28. doi:10.7820/vli.v01.1.brown

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction*, *3*(1), 1–10. doi:10.7820/vli.v03.2.browne

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238. doi:10.2307/3587951

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the vocabulary size test. *Language Testing*, *30*, 253–272. doi:10.1177/0265532212459028

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*, 241–265. doi:10.1093/applin/amm010

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon: Routledge.

Karami, H. (2012). The development and validation of a bilingual version of the vocabulary size test. *RELC Journal*, *43*(1), 53–67. doi:10.1177/0033688212439359

McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction*, *3*(2), 47–55. doi:10.7820/vli.v03.2.mclean.et.al.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, MA: Multilingual Matters.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. del Mar Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol, MA: Multilingual Matters.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*(1), 12–25.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. Retrieved from http://jalt-publications.org/tlt/issues/2007-07_31.7

Nguyen, L. T. C., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, *42*(1), 86–99. doi:10.1177/0033688210390264

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, *18*(1), 55–88. doi:10.1177/026553220101800103

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*, 139–152. doi:10.1080/09571730802389975

Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST? *Language Assessment Quarterly*, *11*, 271–282. doi:10.1080/15434303.2014.922977

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 173–189). Amsterdam: John Benjamins Publishing Company.

West, M. (1953). *A general service list of English words.* London: Longman, Green.