Testing the reliability of the New General Service List Test (NGSLT) in order to better evaluate Japanese university students' written receptive vocabulary levels

Paul Nadasdy[1], Kazumi Aizawa[2], and Tatsuo Iso[3]

Abstract. The New General Service List Test (NGSLT) (Stoeckel & Bennett, 2015) was designed as a diagnostic test to measure students' written receptive vocabulary knowledge. This test battery was developed based upon the New General Service List (NGSL) (Browne, 2013), which makes it appealing to teachers in Japan, and especially those who see vocabulary as key to English as a foreign or second language learning. The research focused on finding out whether and to what degree the test accurately and reliably measures students' vocabulary knowledge, and to find if there are any incongruences with the scores on this test and those on extraneous standards. Three versions of the NGSLT were distributed and a triangulation method was used to analyze the data, with the findings suggesting that the NGSLT may be less a measure of students' knowledge of the target words than a measure of how well they can understand the answer choices.

Keywords: new general service list test, reliability, vocabulary testing.

1. Introduction

The NGSL (Browne, 2013), an upgrade on West's (1953) General Service List (GSL), is a list of high-frequency English words that has been compiled as an educational resource. As it has been well established through corpora studies, only a small number of words from the large amounts of vocabulary available cover the running words in a wide range of texts – 4,000 word families provide around 95%

1. Tokyo Denki University, Tokyo, Japan; nadasdy@cck.dendai.ac.jp
2. Tokyo Denki University, Tokyo, Japan; aizawa@cck.dendai.ac.jp
3. Tokyo Denki University, Tokyo, Japan; tiso@mail.dendai.ac.jp

coverage (Nation, 2006) – and the NGSL has contributed to the available tools by adding more contemporaneous language.

Researchers have been encouraged to use the list as a resource to devise pedagogical tools. One such instrument includes Stoeckel and Bennett's (2015) NGSLT, designed to measure L2 learners' written receptive knowledge of the NGSL. Though a bilingual form has now been published (Stoeckel, Ishii, & Bennett, 2018), at the time of this study, the original monolingual NGSLT was being used as a measure. At this time, the authors were considering whether the original NGSLT could be used to influence design of a placement test at a technical university in northeast Tokyo. What follows is a test of the reliability of the original NGSLT (Stoeckel & Bennett, 2015) using a triangulation method to answer the following questions:

- Does the NGSLT accurately and reliably measure students' vocabulary knowledge?

- Are there discrepancies between the scores of the NGSLT and those on extraneous standards?

2.   Method

2.1.   Participants

A total of 98 Japanese first-year university students completed all the necessary processes in their regular classes. The students' English level was around A2, making use of the NGSLT a suitable means of testing of their vocabulary knowledge.

2.2.   Procedure

Three versions of the test were distributed: The original NGSLT (a multiple-choice, monolingual version, EE), a version where students had to translate target words (TR), and a multiple-choice version with Japanese translations added for target words (EJ). The participants also took the TOEIC[4] as an English proficiency measure (scores range from 195 to 595). Online versions of the test were administered via Google Docs to five intact classes during seven lessons

---

4. Test of English for International Communication®

over a period of two months to avoid practice effects. At first, the EE, comprised of 100 items in English, was distributed during week one. Over the subsequent five weeks, the test was split into five sections, reflecting the five, 20-item bands of the NGSLT and students translated the English target words into Japanese (TR). A list of possible answers was created to identify correct and incorrect answers and the actual rating was done by a computer to eliminate any rater variables. Lastly, in week seven, the EJ version was distributed. Due to occasional absences and lack of TOEIC scores, some data had to be eliminated from the final set. The population of participants who qualified for the final count reduced to 98 from 105 at completion. For the analysis, scores from EE and EJ versions, as well as students' current TOEIC scores, were compared to see if a gap exists in the comparative data.

3.   Results

As a measure of test performance, a Cronbach alpha showed a reliability for the three tests at .90 (EE), .83 (EJ), and .89 (TR). Results in Table 1 show there was a discrepancy between levels one and two in the EJ version of the test and a considerable drop between levels four and five on the EE. Results also showed that mean scores of the EJ were statistically significantly higher than the EE. Furthermore, the differences in average scores between EE and EJ across five frequency bands were also significant, with the participants performing better on the EJ. It was also discovered that correlations between TOEIC scores and the EE or EJ version of the NGSLT were rather weak (r=.31 and .37, respectively).

Table 1.   Descriptive statistics of tests

	Overall	1	2	3	4	5
EE	72.06 (12.02)	16.26 (3.18)	15.61 (3.02)	15.4 (2.21)	13.69 (3.02)	11.1 (3.17)
EJ	90.8 (6.28)	18.54 (1.87)	18.8 (1.29)	18.55 (1.2)	18.22 (1.47)	16.68 (2.31)
TR	50.57 (12.14)	12.15 (4.35)	10.76 (2.72)	10.01 (2.84)	9.74 (3.23)	7.91 (3.27)

As can be seen in Figure 1, there is a gap between the two lines between levels four and five, and there is a considerable drop on the EE side at this level. However, with some help in the Japanese version via translations, there is a distinct difference at levels four and five. As expected, the participants did not perform on the TR as well as on the EE or EJ, confirming that it is more difficult to translate the examples than choose the right answers. There was no interaction observed among

the three versions of the test, indicating no NGSL levels acted irregularly in any of the versions of the test.

Figure 1.   Comparison of EE, EJ, and TR
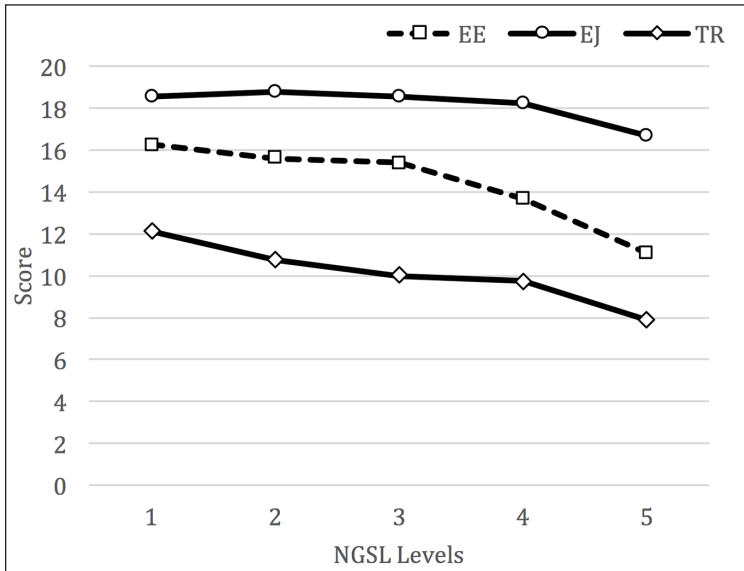
4.   Discussion

It appeared that students did not understand the answer choices in the EE version, suggesting that this may have interfered with how well the test was determining vocabulary level. Results showed that students all did well on the Japanese version of the test, but that did not reflect the range of TOEIC scores. Taken together, these results suggested that participants had difficulty figuring out the meanings of the language used in the possible answers in the EE. Given the fact that the participants were given the translation of the target words, it was apparent that they did not understand in what sense the target words were described in the sample sentences.

The lack of stronger correlation between both versions of the NGSLT and the TOEIC test implied a threat to the validity of the NGSLT, at least when the test takers are similar to the participants in the current study. There also appears to be a ceiling effect in the Japanese version as average scores were high across the board, which may also apply to issues in the English version. As the original validation

tests of the NGSLT mainly checked to see whether the test accurately reflected knowledge of the NGSL, there may have been a discrepancy in how it reflected students' actual levels of English, as the triangulation of results in our study proved that there was a discrepancy between what the NGSLT measured and the students' TOEIC scores. Some help given via the translations in the Japanese version resulted in there being a distinct difference at levels 4 and 5. That gap suggests test takers did not completely understand the answer choices. Therefore, we conclude that the original NGSLT appears to be less a test of vocabulary knowledge of target words and more a test of understanding the possible answer choices.

5.   Conclusions

Analyzing the reliability and validity of vocabulary tests is vital if we want to be more informed of the potential gaps in our students' knowledge. From this, we can make better pedagogical choices based on their needs. This study set out to use a triangulation method to see if the original NGSLT withstood scrutiny. The NGSLT is no doubt a good measure of the knowledge of the NGSL, and the more recent bilingual version has gone some way to rectifying potential issues with the test. We showed, however, that there may be a slight weakness in the levels of the original test. Results suggest the effectiveness of our triangulation method in identifying potential weaknesses in tests of written receptive knowledge. Therefore, we will continue to analyze the accuracy of tests like these in order to use them to help us better understand our students' levels of vocabulary knowledge.

References

Browne, C. (2013). The new general service list: celebrating 60 years of vocabulary learning. *The Language Teacher*, *37*(4), 13-16. https://jalt-publications.org/files/pdf-article/37.4tlt_featureds.pdf

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Stoeckel, T., & Bennett, P. (2015). A test of the new general service list. *Vocabulary Learning and Instruction*, *4*(1), 1-8.

Stoeckel, T., Ishii, T., & Bennett, P. (2018). A Japanese-English bilingual version of the new general service list test. *JALT Journal, 40(*1), 5-21. https://www.jalt-publications.org/node/2/articles/24281-japanese-english-bilingual-version-new-general-service-list-test

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

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.