Developing a vocabulary size test measuring two aspects of receptive vocabulary knowledge: visual versus aural

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

Abstract. Testing learners' English proficiency is central to university English classes in Japan. This study developed and implemented a set of parallel online receptive aural and visual vocabulary tests that would predict learners' English proficiency. The tests shared the same target words and choices – the main difference was the presentation of the target words. Test results showed that for both the Aural Test (AT) and the Visual Test (VT), average scores from 1,000 to 4,000 of the JACET8000 word list's frequency bands decreased in proportion to word frequency. More detailed analyses highlighted gaps between visual and aural aspects of receptive vocabulary knowledge of frequency of 4,000 and beyond. It further indicated that the aural version of the test was a better predictor of English proficiency. For future directions, we need to develop other sets of the VT and the AT to improve the reliability of the tests, and this is especially important in order to satisfy the needs of educational institutes that demand high accuracy in vocabulary testing.

Keywords: vocabulary knowledge, computerized tests, test format, English proficiency.

1. Introduction

The popularity of English as a broad, global subject has been growing in recent times. University students rely on clear measuring of their English ability through systematic testing, and this is a factor that contributes to their final grade upon

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

How to cite this article: Aizawa, K., Iso, T., & Nadasdy, P. (2017). Developing a vocabulary size test measuring two aspects of receptive vocabulary knowledge: visual versus aural. In K. Borthwick, L. Bradley & S. Thouësny (Eds), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 1-6). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.679

graduation. As this is the case, it is fundamental that Japanese students' English ability is tested as accurately as possible. Therefore, universities usually give placement tests to freshmen. According to the questionnaire survey on university placement tests (Aizawa, Iso, & Sasao, 2015), 72% of universities give placement tests to new students upon entering university. Among them, about 80% of those universities purchase placement tests provided by reliable outside sources – *Test of English for International Communication* (TOEIC) 40%, *TOEIC Bridge* 20%, *Test of English as a Foreign Language* 10%, *Assessment of Communicative English* 10%. Moreover, 48% out of the 80% responded that they would be interested in quick and easy-to-use placement tests if they are available.

In the area of vocabulary testing, it has been traditionally the case that only the visual aspects of vocabulary have been measured. A test like TOEIC, for example, is 50% listening questions and therefore relies on students recognizing words within multiple-choice questions. As the demands increase for more accurate measurement of student vocabularies, new innovations in testing need to be developed and considered. Current research suggests that the ways to evaluate the level of an individual's vocabulary is still relatively limited (Pignot-Shahov, 2012). Various tests, however, have been popular for some time, and several vocabulary tests have been developed for research and educational purposes. Among them the *English as a Foreign Language Vocabulary Test* (Meara, 1992) and the *Vocabulary Levels Test* (Nation, 1990) have been widely used. As for these tests, Milton (2009) outlines the positive aspects of the former: speed and ease of use, a large number of items, reliability, multiple versions, less boredom, and better student concentration; and the negative aspects: increased guessing, faced with complex decisions, and overestimation. However, regarding aural vocabulary tests, there is only a limited amount available: the dictation form of Nation's (1990) *Vocabulary Levels Test* (Fountain & Nation, 2000) and *AuralLex* (Milton & Hopkins, 2005). As tests like these are limited in number, the current study aims to provide an alternative option for those wishing to concurrently test vocabulary both aurally and visually.

2. Methods

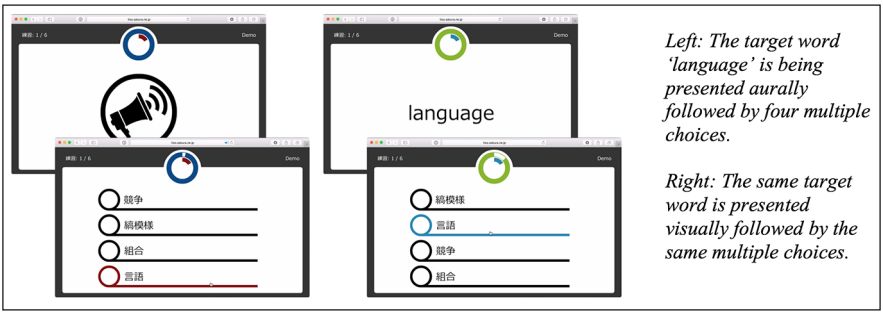
2.1. Test development

In the development of vocabulary tests, it is important to first decide what the exact definition of knowing a word is. There have been several attempts at this, most notably by Richards (1976), Palmberg (1987), and Nation (1990). However, there

is yet to be a conclusive answer to what clearly constitutes vocabulary knowledge. In this paper, we broadly define knowing a word as having an ability to choose, among the list, the corresponding word with the meaning assigned.

The 15 target words and 45 distractor words were randomly chosen from each frequency band of JACET8000 words to construct four multiple-choice questions. Each set of questions consisted of one key and three distractors (the same parts of speech) from eight different frequency bands of 1,000 words each. In total, there were 15 questions of four multiple choices for each eight bands, totaling 120 questions. The AT and the VT are parallel in terms of test format, target words, and choices: the only difference is the mode of target word presentation. In both tests, the target words are presented prior to the four choices (Figure 1). In the AT, each target word was presented aurally, while in the VT, the common target words are shown in a written form. The time length of presenting words (five seconds) and choices (ten seconds) were almost identical. If test takers did not choose an answer for each question, the answer became invalid and the test proceeded to the next question.

Figure 1. Sample screens of AT and VT



Left: The target word 'language' is being presented aurally followed by four multiple choices.

Right: The same target word is presented visually followed by the same multiple choices.

2.2. Purpose and procedure

The purpose of this study is to identify the gaps in average scores between aural and visual tests of receptive vocabulary knowledge. It is also designed to find out whether the test can be used to predict students' TOEIC score. The participants were 140 lower-intermediate engineering students who were taking English as part of their course study. The participants voluntarily took the tests: the AT in the first week, followed by the VT in the third week to avoid the priming effect, since both tests share the same target words as described above. It took about 15 minutes to complete each test.

3. Results and discussion

The descriptive statistics are shown in Table 1. The mean score of the VT was slightly higher than that of the AT (88.56 and 78.16, respectively). Figure 2 shows that the frequency band scores of the AT and the VT declined until frequency level 5,000, but from level six and upward the gaps did not follow this pattern. On every frequency band, the scores of the VT were statistically significantly higher than those of the AT. The gaps in scores between the VT and the AT became larger up until level six, but the differences were not straightforward on levels seven and eight. A further analysis of scores suggested that the AT better predicted TOEIC scores as seen in Figure 3 with a correlation coefficient of r=.57 as opposed to .44 of the VT. Furthermore, the multiple regression analysis revealed that the AT score of level three most reliably accounted for the TOEIC score, followed by level five of the AT. The VT scores were not deemed as estimating factors.

Table 1. Descriptive statistics of the tests

		Mean	SD	MAX	MIN
Vocabulary Test	VT	88.58	10.75	104	44
	AT	78.16	12.24	115	59
TOEIC	Listening	225.30	55.84	375	100
	Reading	165.73	52.53	295	45
	Total	**391.03**	**98.83**	**610**	**180**

Figure 2. Means of scores by JACET8000 levels

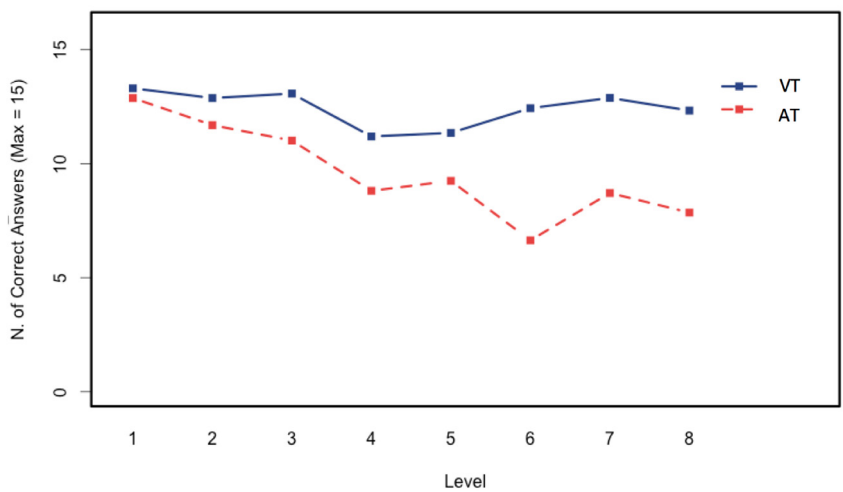
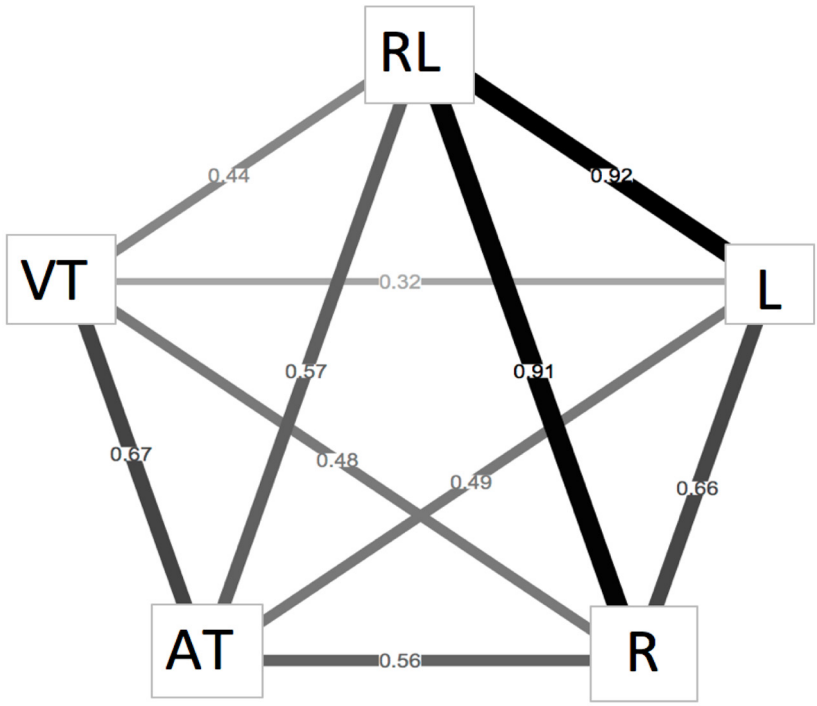


Figure 3. Correlations of TOEIC, AT, and VT



The declining pattern in the AT scores is not observed in the VT scores, which seem to plateau at level five. A possible explanation is the overall ceiling effect in the VT. Another is the inherent problem found in the word list itself – it has been reported elsewhere that levels five and after of the JACET8000 words contain a certain number of loan words and derivational forms of higher frequency words. Some of those could be inadvertently used as target words.

The current study showed that the AT was a better predictor of the TOEIC score. This is an important finding as no previous studies directly compared the results of visually and aurally tested vocabulary sizes.

4. Conclusion

Testing in Japanese educational institutions is fundamental for placing students in suitable levels and for evaluating ability and progress. We have strived to create

testing instruments that will predict most accurately the vocabulary sizes of a small group of Japanese students, and though there was a considerable amount of success in the data returned, more research is needed and more refinement of the instruments is necessary if we are to provide high levels of reliability in vocabulary testing. Further experimentation using refined versions of the same tools will be carried out in subsequent research projects, and we would invite others that are interested in measuring the accuracy of their students' vocabulary levels to do the same.

5. Acknowledgements

This work was supported by JSPS Research Grant (No. 26580114).

References

Aizawa, K., Iso, T., & Sasao, Y. (2015). *Developing an online placement test (interim report)*. Foreign Language Teaching Expo 2015, Tokyo.

Fountain, R., & Nation, I. S. P. (2000). A vocabulary-based graded dictation test. *RELC Journal 31*, 29-44. https://doi.org/10.1177/003368820003100202

Meara, P. (1992). *EFL vocabulary tests.* University College Swansea.

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

Milton, J., & Hopkins, N. (2005). *AuralLex*. Swansea University.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle and Heinle.

Palmberg, R. (1987). Patterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition, 9*, 201-220. https://doi.org/10.1017/S0272263100000474

Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers, 4*, 37-45.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*, 77-89. https://doi.org/10.2307/3585941

Research-publishing.net

Published by Research-publishing.net, not-for-profit association
Contact: info@research-publishing.net

CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017
Edited by Kate Borthwick, Linda Bradley, and Sylvie Thouësny

Typeset by Research-publishing.net

Cover design based on © Josef Brett's, Multimedia Developer, Digital Learning, http://www.eurocall2017.uk/, reproduced with kind permissions from the copyright holder.

Cover layout by © Raphaël Savina (raphael@savina.net)
Photo "frog" on cover by © Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-04-7 (Ebook, PDF, colour)
ISBN13: 978-2-490057-05-4 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-03-0 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit: Bibliothèque Nationale de France - Dépôt légal: décembre 2017.