



## Paul Raine

In this column, we discuss the latest developments in ed-tech, as well as tried and tested apps and platforms, and the integration between teaching and technology. We invite readers to submit articles on their areas of interest. Please contact the editor before submitting.

Email: [jaltpubs.tlt.wired@jalt.org](mailto:jaltpubs.tlt.wired@jalt.org)

Web: <https://jalt-publications.org/tlt/departments/tlt-wired>

Paul Raine has been a Japan-based teacher and coder since 2006. He has developed the web-based language teaching and learning platform *TeacherTools.Digital*, and many other web-based tools.

# Create Your Own Vocabulary Levels Tests with VocabLevelTest.org

Brett Milliner

[milliner@lit.tamagawa.ac.jp](mailto:milliner@lit.tamagawa.ac.jp)

Vocabulary Levels Tests (VLTs) represent one of the most practical test instruments for any language teacher's toolbox. VLT results can inform a teacher's selection of classroom materials, help track vocabulary growth, and identify gaps in high-frequency vocabulary knowledge. At the program level, VLTs can be used for class placement and program evaluation. For classroom research, VLTs can help researchers to group participants in terms of lexical knowledge. This article will introduce VocabLevelTest.org: a reliable, customizable, and free VLT system which teachers can use to create self-marking VLTs in less than five minutes.

### What Do VLTs Measure?

VLTs are test instruments that target vocabulary breadth, or how many words learners know. Nevertheless, it is perhaps more informative to conceptualize VLTs in terms of the area of vocabulary knowledge that they are designed to evaluate, the form-meaning link. The form-meaning link concerns whether a language learner knows the form of a word (i.e., what the word looks or sounds like) and its meaning, and whether the learner can connect these two parts of knowledge. Because VLTs provide evidence that target words can be comprehended while listening or reading, VLT results can inform a range of decisions for the foreign language classroom.

### VLT Design

Target words for VLTs are sampled from a word frequency lists such as the JACET 8000 (Mochizuki,

2016), the NGSL (Browne et al., 2013), and the BNC/COCA (Nation, 2017). Target words are then selected to represent specific word frequency bands.

Generally, between 10 and 30 questions will represent a 1000-word band, and a 90% or above score on a band indicates mastery. Outside of variations in word frequency lists, another big difference between VLTs is question format. Table 1 outlines two popular formats used in VLTs: meaning recognition (multiple-choice or matching) and meaning recall. Recent research in language testing has, however, started to question the reliability of the meaning recognition format because it is susceptible to guessing, and it lacks the power to measure the type of vocabulary knowledge suitable for reading practice (see McLean et al., 2015; McLean et al., 2020; Stoeckel et al., 2021). The meaning recall format, on the other hand, appears to more reliably measure written receptive vocabulary knowledge (the vocabulary knowledge required for reading) than the meaning recognition format (McLean et al., 2020).

Table 1  
VLT item formats

	Recognition	Recall (VocabLevelTest.org)
<b>Meaning</b>	Q. House. It is a house. (a) 本 (b) 果物 (c) 車 (d) 家	Q. House. It is a house. 家 Questions delivered via written or spoken modalities
<b>Form</b>	Q. I like that 家. (a) Book (b) Fruit (c) Car (d) House	Q. I like that 家. <u>House</u>

Note. The VocabLevelTest.org website offers only the meaning-recall and form-recall formats.

## VocabLevelTest.org Design

With the assistance of a Grant-in-Aid for Scientific Research (KAKENHI; 20K00792), Stuart McLean designed and I developed VocabLevelTest.org (McLean & Raine, 2018). On this free testing platform, users can create online, self-marking meaning recall (reading or listening) and form recall (typing) vocabulary tests. Many of the design features of this site address limitations of existing VLTs. Test-taker responses are checked against a bank of possible answer choices so students and teachers receive immediate feedback on their performance. The testing site also allows test administrators to check which response types are treated as correct or incorrect, and then override how the answer bank has marked response types. Please see McLean et al. (in press) for details on the accuracy of the automatic marking. Test administrators can also download typed responses for manual marking.

As shown in Figure 1, teachers have a wide array of options to customize tests for specific contexts. Learners can answer items in a variety of first languages, including Japanese, Arabic, French, Dutch, Vietnamese, and Chinese. There is a range of frequency lists to design tests (e.g., JACET 8000, NGSL, BNC/COCA). The length and focus of test items can be controlled with the *Band Size*, *Starting and Ending Band*, and *Items per Band* options. Test creators can allow the system to choose items automatically, or select target words by themselves. There are also three test formats to choose from—*Receptive Reading* (meaning recall), *Receptive Listening* (meaning recall), and *Productive Typing* (form recall). Taking the settings used in Figure 1 as an example, VocabLevelTest.org would create a 25-item *Receptive Reading* (meaning recall) test. The target words would come from the 250–1500 frequency bands of the NGSL. After designing a test, the test creator is provided with a URL and QR code to share with students, who then can complete the test on any internet-connected device.

After finishing a test, learners receive immediate feedback, which they can review in order to determine any gaps in their vocabulary knowledge. Teachers and test creators have the ability to review individual student scores or summaries of class results (Figure 2), and they can choose to override responses (i.e., change responses marked incorrect by the system). Test creators can download a complete dataset (Excel file) including scores, learner responses, and response times.

Figure 1

Customization options for creating a test in VocabLevelTest.org

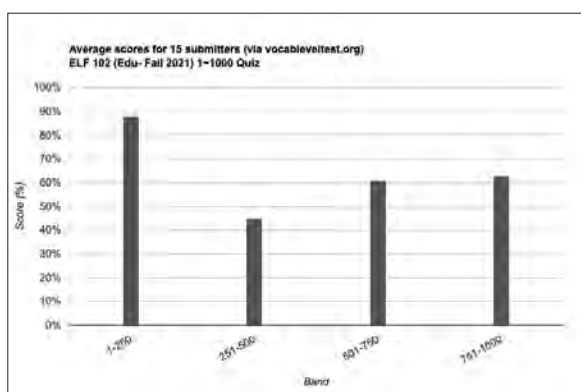
The screenshot shows the 'Create Vocabulary Test' interface. Key settings include:
 

- Test Title:** TLT Wired Exam
- Native Language:** Japanese
- Wordlist:** NGSL (General)
- Band Size:** 250
- Starting Band:** 251-500
- Ending Band:** 1251-1500
- Items Per Band:** 5
- Question Type:** Receptive Reading (Meaning Recall)
- Show Feedback:** Show feedback on every question
- Test Takers (0):** 0

 A 'Create' button is located at the bottom right of the form.

Figure 2

Class summary report in VocabLevelTest.org



## How I Have Used VocabLevelTest.org

At the start of one of my university classes, I asked students to take a 50-item *Receptive Reading* (meaning recall) VLT which was based on the NGSL

word list. The band size was 250 words (10 items per band) and it targeted words ranked between the 250–1500 bands. The average class score for the 250–500 and 500–750 bands was over 85%. For the 750–1000 bands, the average class score was 70%, and for the 1000–1250 and 1250–1500 bands the average was close to 60%. During the language-focused learning component of my class, we worked on learning unknown vocabulary from the 750–1500 bands of the NGSL. I shared spreadsheets of the NGSL list, and each week we focused on studying a 100-word band of words. Then, I used VocabLevelTest.org to create short, formative assessment tasks to follow up on my students' vocabulary study.

A second application of the VLT for me was selecting materials for fluency development. I run timed reading practice every class. For fluency development activities it is essential that learners know close to 100% of the words in a text (Nation, 2007). Therefore, for my timed reading component (see Milliner, 2021 for a more detailed description), I selected Millett's (2017) BNC 500 text because my class's average scores were only close to mastery in the 250–500 and 500–750 bands. Albeit brief, these two examples show how I use VocabLevelTest.org to select level-appropriate materials and make decisions about where language-focused learning needs to occur.

### Final Thoughts

VocabLevelTest.org addresses many of the methodological deficiencies of previously published VLTs. The *Receptive Reading* (meaning recall) format has been shown to reliably appraise written receptive vocabulary knowledge (McLean et al., 2020). The automatic marking system for *Receptive Reading* (meaning recall) is reported to be very reliable for Japanese L1 test-takers (see McLean et al., in press), and work is being done to validate the *Receptive Listening* (meaning recall) format. For a more detailed description of the VocabLevelTest.org website, see McLean et al. (in press).

### References

Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. <http://www.newgeneralservicelist.org>

McLean, S., Kramer, B., & Stewart, J. (2015). An empirical examination of the effect of guessing on vocabulary size test scores. *Vocabulary Learning and Instruction*, 4(1), 26–35. <http://vli-journal.org/wp/wp-content/uploads/2015/10/vli.v04.1.2187-2759.pdf#page=31>

McLean, S., & Raine, P. (2018). *VocabLevelTest.org*. [Online program]. <https://www.vocableveltest.org/>

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, 37(3), 389–411. <https://doi.org/10.1177/0265532219898380>

McLean, S., Raine, P., Pinchbeck, G., Huston, L., Kim, Y., Nishiyama, S., & Ueno, S. (In press). The internal consistency and accuracy of automatically scored written receptive meaning-recall data: A preliminary study. *Vocabulary Learning and Instruction*.

Millett, S. (2017). Speed readings for ESL learners, 500 BNC. *ELI Occasional Publication No. 28*. [https://www.victoria.ac.nz/\\_data/assets/pdf\\_file/0011/1068077/SRs-for-ESL-Learners-500-BNC-April-2017.pdf](https://www.victoria.ac.nz/_data/assets/pdf_file/0011/1068077/SRs-for-ESL-Learners-500-BNC-April-2017.pdf)

Milliner, B. (2021). The effects of combining timed reading, repeated oral reading, and extensive reading. *Reading in a Foreign Language*, 33(2), 191–211. <http://hdl.handle.net/10125/67400>

Mochizuki, M. (2016). *JACET 8000: The new JACET list of 8000 basic words*. Kirihara.

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2–13. <https://doi.org/10.2167/illt039.0>

Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>

## The Japan Association for Language Teaching (JALT)

### A nonprofit organization

The Japan Association for Language Teaching (JALT) is a nonprofit professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and a means of keeping informed about developments in the rapidly changing field of second and foreign language education.

### JALT Board of Directors, 2021

- President . . . . . Dawn Lucovich
- Vice President . . . . . Robert Chartrand
- Director of Membership . . . . . Melodie Cook
- Director of Program . . . . . Wayne Malcolm
- Director of Public Relations . . . . . William Pellowe
- Director of Records . . . . . Kenn Gale
- Director of Treasury . . . . . Michael Mielke
- Auditor . . . . . Steve Brown