

Test Taking and DK Use on the Vocabulary Size Test

Dawn Lucovich

Teachers College Columbia University, The Writing Center

doi: <http://dx.doi.org/10.7820/vli.v03.2.lucovich>

Abstract

This study continued an investigation of test strategy and usage of an “I don’t know” (DK) option on the 20,000-word family, 100-item Vocabulary Size Test (VST; Nation & Beglar, 2007). Unlike previous studies, which reported on an L1 American English user (Lucovich, 2014), the participants in this study were two L1 Greek–L2 American English users. Each participant took the original and DK-added versions of the VST. A structured qualitative interview, based on Nagy, Herman, and Anderson’s (1985) word interview protocol and piloted in previous studies, asked the participants about their answers on the VST. This study aimed to discover how non-Japanese L1/non-English L1 test takers determined their answers, how they qualitatively perceived and used the DK option, and whether this differed from the L1 American English and L1 Japanese users of English. As in previous studies, the participants used the DK option only on unknown items where partial knowledge or test taking strategies were unusable.

1 Introduction

Although vocabulary learning and instruction is considered integral to academic curricula and to reading comprehension, knowledge of learners’ vocabulary size tends to be overlooked. Having an estimate of vocabulary size is a crucial part of determining reading level, which in turn affects curriculum development, textbook selection, learning objectives, and classroom goals. Pedagogically, vocabulary size can be used to gauge prior knowledge, learner progress, effectiveness of instruction, and to guide vocabulary instruction and assessment.

The Vocabulary Size Test (VST; Nation & Beglar, 2007) has been established as a valid and reliable measure of written receptive vocabulary size (Beglar, 2010). In addition, it is easy to administer, take, score, and interpret, and is useful to both test takers and administrators. However, guessing on multiple-choice VSTs and potential vocabulary size inflation is a concern (Stewart, 2014; Stewart & White, 2011). Although the VST typically lacks an “opt-out” response, one method that has been shown to counteract guessing on multiple-choice vocabulary tests is the addition of an “I don’t know” (DK) option (Bennett & Stoeckel, 2012; Zhang, 2013). However, as the collection of both quantitative and qualitative data on the process of vocabulary test taking and on the VST has been encouraged (Nation, 2012a, 2013, personal communication), more intensive qualitative research is

needed to complement the existing quantitative data on vocabulary testing. As in this study, Nation (2012a) has also employed a “labour-intensive method of testing” that used one-on-one interviews in order to “make sure that the results meant something” (p. 10).

This study aimed to discover: (1) how L1 Greek–L2 American English (i.e. non-Japanese L1/non-English L1) test takers determined their answers on the VST; (2) how they qualitatively perceived and used the DK option; (3) whether this differed from an L1 American English and L1 Japanese users of English.

1.1 The Issue of Guessing in Vocabulary Testing

As with most multiple-choice vocabulary tests, the VST lacks an “opt-out” response. Since the VST is a written multiple-choice test of receptive vocabulary size, partial knowledge that leads to a correct answer is encouraged; hence, the test does not include an opt-out response (Nation, 2012b). Instead of answering items, however, test takers might leave items unanswered (no response or NR) or guess randomly. For the test administrator, items coded as NR do not provide information about an item or the test taker’s vocabulary knowledge. NR cannot differentiate between human error and (lack of) knowledge: The item could have been accidentally missed (human error); intentionally skipped (possibly due to lack of vocabulary knowledge) and then forgotten (human error); or intentionally skipped (lack of vocabulary knowledge).

From a test taker’s point of view, there is little advantage to leaving an item unanswered as it will automatically be marked wrong. Without an opt-out response, penalty directions, or with unknown consequences, there is a much greater advantage, score-wise, to guessing, as test takers have a chance of getting the answer correct. Without an opt-out response, a test administrator cannot differentiate between a correct answer (vocabulary knowledge), an informed guess (partial vocabulary knowledge or test taking strategy), and an uninformed guess (lack of vocabulary knowledge or test taking strategy).

1.2 The DK Option in Vocabulary Testing

One method that could potentially deter guessing is the inclusion of a DK option among the choices. The DK option on its own has been found to slightly discourage guessing and to significantly discourage guessing when coupled with a scoring penalty (Zhang, 2013). However, whether a DK option discourages informed guessing or only uninformed guessing, and how and when, is something that can be better explored through qualitative means.

The addition of a DK option could have a variety of effects. Test takers could potentially be using a DK option when: they have never studied or heard a word before, they cannot guess based on partial knowledge, they cannot apply a test strategy like process of elimination, they cannot definitively decide between two choices, they feel test fatigue, they do not know or are not sure of one or more answers sequentially, they do not care about the outcome of a test, or when the test results do not matter or the effort of test taking outweighs the impact of the results.

How test takers perceive and use the DK option is important to answer as the inclusion of a DK option could lead to score overestimation or underestimation. Similarly, these perceptions and usage may differ between different test takers or types of test takers for reasons such as education level, L1, or English proficiency, which are explored in this study.

2 Methodology

2.1 Instruments

The VST (Nation & Beglar, 2007) can test up to the 20,000-word family level (Nation, 2012b) for users of English, whether they are L1, EFL, or ESL users. The 100-item monolingual (i.e. English) VST was used as an instrument for this study. Since two versions of the VST have been tested and deemed equivalent versions (Beglar, 2010), one version was utilized in its original format, while a fifth choice (*I don't know*) was added to the second, equivalent VST. A sample item from each of the VSTs follows:

Original VST:

1. drive: He <drives> fast.
 - a swims
 - b learns
 - c throws balls
 - d uses a car

DK-added VST:

1. see: They <saw it>.
 - a closed it tightly
 - b waited for it
 - c looked at it
 - d started it up
 - e I don't know

Directions are not provided on the VST, and should be added by the test administrator according to the level of test strictness required. Unlike Zhang (2013), which included three testing conditions (including one with a guessing penalty), this study included test directions with high ecological and face validity; that is, they were most likely to resemble program or classroom-level administrations of the VST. The directions used in this study follow:

Original VST:

For each item, read the word and the example sentence. Circle the letter of the answer (a, b, c, or d) that most closely matches the word.

DK-added VST:

For each item, read the word and the example sentence. Circle the letter of the answer (a, b, c, or d) that most closely matches the word. If you do not know the answer, circle e.

2.2 Procedure

The participants were two Greek American males in their mid-thirties. Participant A holds a B.A. Economics and Participant B holds a B.S. Psychology, respectively. Both participants spoke Greek at home until approximately kindergarten; both attended elementary school in Greece for 2–3 years before returning to the USA. Participant A returned to Greece in his early thirties for several years. However, Participant A volunteered that “my English is better than my Greek now” and Participant B agreed that “now I definitely speak English a lot more fluently”. Compared to previous research, which reported the results of a high-proficiency L1 English user and intermediate-proficiency L1 Japanese users of English, the L1 Greek users can be said to occupy an intervening position in English-language proficiency: they are L2 users of English, but proficient enough they that now consider their L2 to be their dominant language.

The participants were met on two separate occasions and individually briefed about the study. Both the original and DK-added VSTs were taken by each participant. Each test taking period was between 20 and 60 minutes long. Each interview was conducted one-on-one and lasted an average of 20 minutes. A short debriefing was then conducted. Each session lasted no longer than 2 hours in total.

The qualitative interview utilized a structured item interview based on the word interview protocol by Nagy, Herman, and Anderson (1985) with elements of a retrospective think-aloud protocol. Each interview was conducted immediately after the test was administered. The interviews were transcribed and coded by the interviewer using provisional coding (Miles & Huberman, 1994) developed in previous research (Lucovich, 2014). The questions included:

1. Can you tell me what the word means?
2. Why did you choose that answer?
3. How did you arrive at that answer?

3 Results

Below are the results of the dichotomously scored (correct–incorrect) answers on the two versions of the VST (Figures 1 and 2), including the frequency range of the incorrect and DK answers.

Research Question 1. How do L1 Greek–L2 American English test takers determine their answers on the VST?

First, the results of the participants’ structured interviews fell into several broad categories. When asked how they knew a word and if they could demonstrate

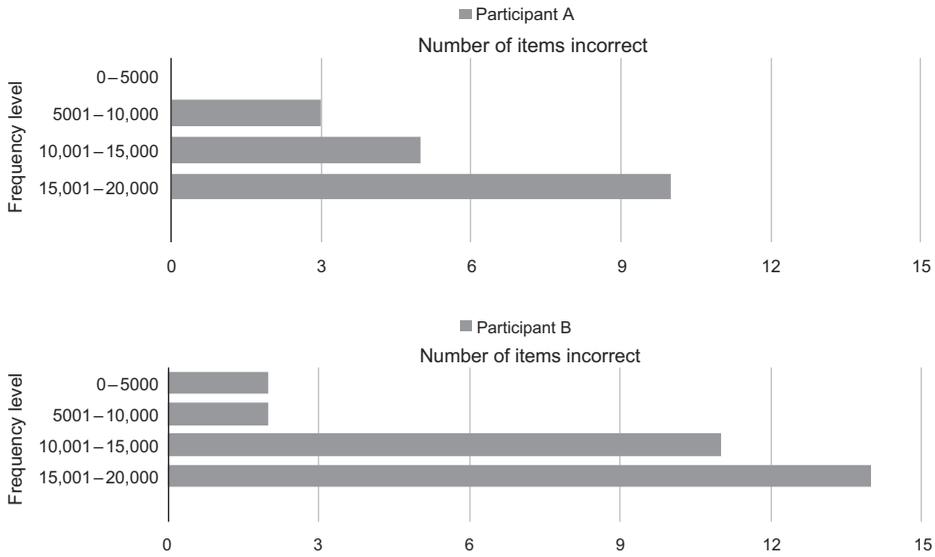


Figure 1. Number of incorrect items on the original VST according to frequency level by participant.

knowledge, they consistently identified or were classified into the following conditions:

1. *Knowledge-informed answer*: They were familiar with the target word.
2. *Partial knowledge-informed answer*: They had some knowledge of the target word, but this knowledge was not sufficient enough to satisfy condition #1.

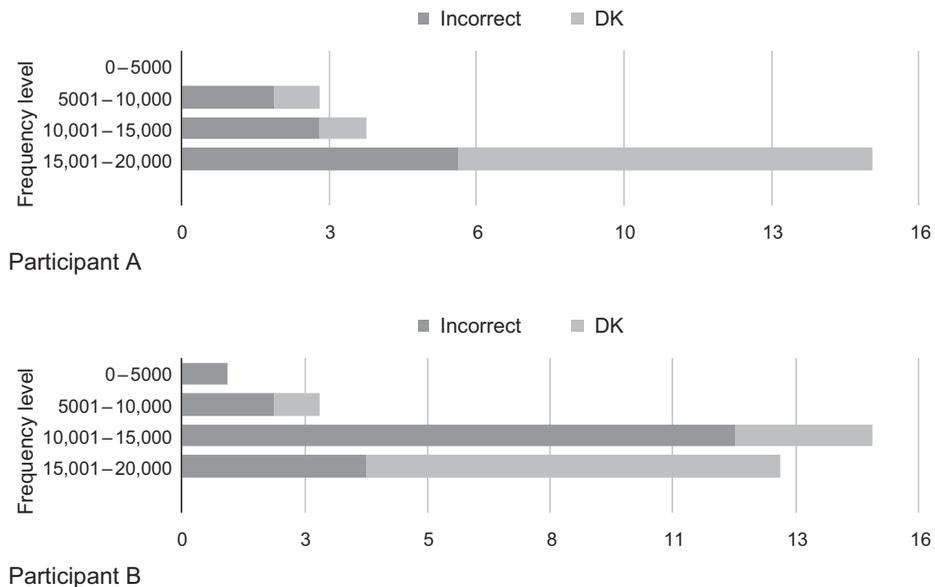


Figure 2. Number of incorrect items vs. DK usage on the DK-added VST according to frequency level by participant.

3. *Test strategy-informed answer*: They had some knowledge of a non-target word, a distractor, non-distractor, or item stem.
4. *Informed guess*: They had some knowledge or strategy, but it was insufficient to satisfy conditions #1, 2, or 3.
5. *Uninformed guess*: They lacked any knowledge or were unable to provide any additional information on the item or target word.

The participants' coded answers confirmed the provisional coding developed in previous research (Lucovich, 2014). In pilot studies, an L1 American English user and L1 Japanese users of English identified or were classified into identical conditions.

The participants cited the following reasons for their answers on the 100-item VST (in decreasing order of use, not differentiating between correct/incorrect answers):

Original VST:

- Knowledge-informed answer (70%)
- Partial knowledge-informed answer (16%)
- Test strategy-informed answer (6%)
- Uninformed guess (5%)
- Informed guess (4%)

DK-added VST:

- Knowledge-informed answer (61%)
- Partial knowledge-informed answer (17%)
- DK usage (12%)
- Test strategy-informed answer (7%)
- Uninformed guess (2%)
- Informed guess (2%)

However, unlike previous studies, the participants also mentioned their use of cognates (or perceived cognates) as partial knowledge or as a test strategy. For #75 “cenotaph”, Participant B said: “I don’t know if there’s any relation, but there’s a Greek word which means ‘coffin’. So I think that’s why I picked [‘memorial for people buried elsewhere’ for] ‘cenotaph.’” Unlike the L1 English users or the L1 Japanese users of English, the L1 Greek test takers could rely on their L1 in order to inform guesses at lower frequency words – that is, words at a frequency level that they would not typically be able to answer correctly.

Research Question 2. How do test takers qualitatively perceive and use the DK option?

Both test takers used the DK option only when they had no other choice, either due to lack of knowledge, lack of partial knowledge, or inability to apply test strategy to the item. The most common reason cited for using the DK option was “never hearing the word before”.

Participants arrived at a DK option in one of two ways. First, participants used DK when they had no initial knowledge of a VST item. They cited reasons that illustrated a clear lack of knowledge: “I just didn’t know what the answer was” or “Because I really had no idea...”. Second, they used DK when they had cycled through their partial knowledge and potential test strategies, but still lacked sufficient knowledge to make an informed guess. In this condition, they skipped making an uninformed guess and instead chose the DK option. Participant A stated, “If I saw two or three [choices] that I didn’t know or really had no idea, then I put ‘I don’t know.’” The following excerpt from Participant B’s retrospective think-aloud during the interview additionally illustrates this test taking process:

35. ubiquitous: Many unwanted plants <are ubiquitous >.

- a are difficult to get rid of
- b have long, strong roots
- c are found everywhere
- d die away in the winter
- e I don’t know

I was pretty sure it wasn’t B. But when you see many unwanted plants everywhere, it’s probably A or C. I think that would make sense ‘cause it says “unwanted plants” . . . if that was associated back to “unwanted”. So many unwanted plants are difficult to get rid of, that makes sense. Many unwanted plants are found everywhere. That also kind of makes sense, but I didn’t know, and I got frustrated so I just picked E ‘cause. . . it could be D, it could be C, it could be A.

Here, Participant B is clearly using partial knowledge of the target word “ubiquitous”, followed by the test strategy of elimination and stem usage. Since he was unable to narrow the choices any further, he chose the DK option.

At the end of the testing period, both participants were asked if they would have used an “I don’t know” option on the Original VST. Participant B volunteered: “Definitely . . . there was definitely many times when. . . it could be A, B, C, or even D.”

The DK option is working as intended: it is not being overutilized, as it is chosen only when the test taker lacks sufficient knowledge; nor is it being underutilized, as it is chosen (in preference to make an uninformed guess) when the test taker lacks any knowledge of the correct answer.

Research Question 3. Do these perceptions and DK usage differ from the L1 American English and L1 Japanese users of English?

Finally, compared to the results of the pilot studies with an L1 American English user and intermediate-proficiency Japanese users of English, both L1 Greek–L2 English participants guessed more often; accordingly, compared to the L1 Japanese user of English, they used DK less often.

Participants A and B were similar to the L1 American English user (Lucovich, in press), who only used the DK options under the following conditions: They “did not know the meaning at all”, had “heard the word before but [didn’t] know the meaning”, or it “[was] a word [they’d] never heard before”.

4 Discussion

No penalty for guessing was included in the test taking directions, nor applied when scoring the test. As such, the directions replicate typical test taking at the program or classroom level. Since the directions do not disclose the scoring procedure, they neither overly encourage guessing, nor discourage it (as the penalty for guessing incorrectly is the same as an incorrect answer or NR).

Future administration of the VST at the program or classroom level should consider how much test strictness is necessary and how accurate of a vocabulary size estimate is needed in order to ascertain whether a penalty should be included. The testing conditions for this study result in a more accurate vocabulary size estimate as participants neither overused the DK option (i.e. using it when they could actually demonstrate knowledge of the word), nor underused it (i.e. they used it when they encountered an unknown word and could not demonstrate knowledge or apply a test strategy).

As this study investigated only two L1 Greek–L2 American English users, the findings cannot be broadly generalized to other populations or contexts. However, the findings are important as they are aligned with previous findings on the same topic and lend credence that test taking and DK usage on the VST are consistent among English users with higher education, regardless of L1 and despite varying English proficiency levels.

5 Conclusion

Based on the findings of this study in conjunction with previous studies, an interview code is being developed for use in future VST and multiple-choice test research. Utilizing this interview code, future research will examine whether these findings are consistent across larger sample sizes of populations and with lower-proficiency English users. Future work on the VST itself and other vocabulary tests could also focus on increasing their appeal, accessibility, and meaningfulness to language programs, classroom instructors, and language learners. Finally, further research should examine the effects of adding a DK option in other language learning and testing contexts, and its subsequent effect on guessing, test strategy, and score.

References

- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101–118. doi:10.1177/0265532209340194
- Bennett, P., & Stoeckel, T. (2012). Variations in format and willingness to skip items in a multiple-choice vocabulary test. *VERB*, 1(2), 2–3.
- Lucovich, D. (2014). Adding “I don’t know” to the Vocabulary Size Test. In H. Brown & N. Sonda (Eds.), *JALT 2013 Conference Proceedings*. Tokyo: JALT.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

- Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233–253.
- Nation, I. S. P. (2012a, August). Measuring vocabulary size in an uncommonly taught language. Paper presented at the International Conference on Language Proficiency Testing in the Less Commonly Taught Languages, Bangkok, Thailand. Retrieved from <http://www.sti.chula.ac.th/files/conference%20file/doc/paul%20nation.pdf>
- Nation, I. S. P. (2012b, October 23). *The Vocabulary Size Test*. Retrieved from: <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>
- Nation, I. S. P., & Beglar, D. (2007). A Vocabulary Size Test. *The Language Teacher*, 31(7), 9–13.
- Stewart, J., & White, D. (2011). Estimating guessing effects on the Vocabulary Levels Test for differing degrees of word knowledge. *TESOL Quarterly*, 45, 370–380.
- Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST? *Language Assessment Quarterly*, 11(3), 271–282. doi:10.1080/15434303.2014.922977
- Zhang, X. (2013). The *I don't know* option in the Vocabulary Size Test. *TESOL Quarterly*, 47, 790–811. doi:10.1002/tesq.98