Validation of an online questionnaire of vocabulary learning strategies for ESL learners

Peter Yongqi Gu
Victoria University of Wellington, New Zealand
peter.gu@vuw.ac.nz

Abstract

This article reports on the updating and validation of a questionnaire for vocabulary learning strategies. An English as a second language (ESL) version of the *Vocabulary Learning Questionnaire* (VLQ), first published in Gu and Johnson (1996), was created based on its latest version. Two response formats were piloted, that is, a paper version with a 7-point Likert scale and an online version with a 100-point slider bar. A series of validation procedures resulted in a 62-item instrument which was then administered online among 682 English language learners at the tertiary level in China. The paper presents evidence of content validity, construct validity, internal consistency reliability, and predictive validity. It also provides suggestions for interpreting and using the questionnaire for both research and instructional purposes.

Keywords: vocabulary learning strategies; questionnaire validation

1. Introduction

Research on vocabulary learning strategies (VLS) frequently makes use of questionnaires to elicit data. Teachers and learners of vocabulary may also make use of a questionnaire to diagnose vocabulary learning problems. There is, however, an apparent lack of information on the validation of the most often used taxonomies of VLS. This study aims to create an updated English as a second language

(ESL) version of the VLQ (Gu & Johnson, 1996). It also aims to pioneer the use of an online 0-100 slider bar as a potential alternative to the Likert scale format. The main body of this article will document the updating and validation procedures, and provide various evidences of validity. I will finally offer suggestions for the interpretation and use of the questionnaire. A few remaining concerns for using the VLQ will be presented at the end.

1.1. Vocabulary learning strategies

Learning vocabulary in a second language is a conspicuously long and demanding task. We start the whole process of language learning by learning the most basic words and phrases, and we never stop developing our vocabulary even at the highest level. The demanding nature of the task makes strategic learning necessary, in the sense that the appropriate choice and deployment of strategies can make a big difference in determining if vocabulary learning becomes an efficient or inefficient, and even pleasant or frustrating experience.

Strategic learning is a deliberate, dynamic and iterative process for solving a learning problem, boosting the learning speed, or making the learning process efficient, effective, and pleasant. Although we can proactively plan for our learning and deploy pre-emptive strategies, a large part of strategic learning is essentially a problem-solving process. In other words, when a novel or demanding task comes into our attention, we quickly analyze the task, analyze our own resources for learning, and analyze the context of learning so as to come up with a plan to overcome the learning hurdle. As the plan is carried out, we monitor its effectiveness and adjust the plan accordingly. When the task reaches a completion stage, we evaluate if, when, and which strategic effort has succeeded or failed in helping us solve the problem. This process repeats itself in cycles every time a new or difficult task is being attended to. Repeated and successful deployment of the same strategy can turn the strategy from a declarative stage to a procedural and automatic stage (O'Malley & Chamot, 1990), with many of the conditions associated with each successful or unsuccessful application stored together as conditional knowledge. When a strategy reaches this level of expertise, future tasks and conditions that are similar enough can trigger an automatic application of the strategy with or without the learner's awareness.

There has been a sustained interest in VLS in second language acquisition. As in the research of language learning strategies in general, initial efforts concentrated on the identification, description and classification of VLS (e.g., Ahmed, 1989). Typically the research is exploratory and bottom-up. Researchers used interviews, observations and other qualitative measures to collect the naturally occurring strategies for vocabulary learning. The resulting list is then matched

against theory for classification and grouping. This research has resulted in a few taxonomies of VLS that have been widely used.

In the latest edition of Nation's (2013) comprehensive book on vocabulary learning and teaching, he retains his three-way classification of "vocabulary-learning strategies" in the 2001 edition of the book: *planning* (choosing what to focus on and when to focus on it), *sources* (finding information about words), and *processes* (establishing knowledge). In addition, Nation devotes five other chapters to vocabulary learning: "learning words from context," using "word parts," "using dictionaries," "deliberate learning from word cards," and "finding and learning multiword units." One of the most often used taxonomies is Schmitt's (1997) VLS taxonomy of 58 items. Schmitt groups VLS into two: *discovery strategies* and *consolidation strategies*. The former include *determination strategies* and *social strategies*. The latter cover *social strategies*, *memory strategies*, *cognitive strategies*, and *metacognitive strategies*. Gu's (2013) taxonomy includes two main components: *metacognitive* and *cognitive VLS*. The metacognitive component includes *beliefs about vocabulary learning* and *metacognitive regulation of vocabulary learning*. The cognitive component includes *initial handling strategies*, *reinforcement strategies*, and *activation strategies*.

The three taxonomies overlap considerably. For example, all three include selectively choosing words to learn; all three contain strategies for initial learning after a word is chosen (*sources for information* in Nation, or *discovery* in Schmitt), and further consolidation after initial handling. In fact, this process perspective is also reflected in another classification by Brown and Payne (1994) who describe "five essential steps in vocabulary learning" (Hatch & Brown, 1995, p. 372): *encountering new words*, *getting the word form*, *getting the word meaning*, *consolidating the word form and meaning in memory*, and *using the word*.

A latest addition is Tseng, Dörnyei, and Schmitt (2006), whose 20-item *Self-Regulating Capacity in Vocabulary Learning Scale* is meant to measure the "capacity of strategic learning." These 20 items are grouped into five subscales: *commitment control*, *metacognitive control*, *satiation control*, *emotion control*, and *environment control*. As a theory-driven taxonomy, these "action control strategies" represent holistic approaches in learning management. They are more related to generic self-regulation behaviors and not directly related to the various tasks for vocabulary learning (e.g., "When I feel stressed about vocabulary learning, I know how to reduce this stress," "When learning vocabulary, I have special techniques to achieve my learning goals). As a measure of a learner's capacity for self-management and control, the *Self-Regulating Capacity in Vocabulary Learning* scale does not serve the purpose of guiding learners and teachers as to how vocabulary should be learned.

In addition to discovering types of VLS, considerable efforts have been invested in mapping the relationship between VLS and learning results. In general,

qualitative explorations (e.g., Gu, 2003; Parry, 1993; Sanaoui, 1995) confirm correlational patterns (e.g., Gu & Johnson, 1996; Kojic-Sabo & Lightbown, 1999) that a large repertoire and frequent use of strategies are positively related to vocabulary size and even general proficiency. Teaching VLS (e.g., Mizumoto & Takeuchi, 2009) has resulted in improvement in measures of strategic learning as well as in vocabulary. Recent years have also seen research efforts concentrating on relating VLS to different aspects of vocabulary and language learning. For example, Zhang and Lu (2015) explored the relationship between VLS and both vocabulary breadth and vocabulary depth. Ranalli (2003, 2012) examined the incorporation of VLS in coursebooks and in web-based training. There is also effort in exploring the effect of VLS instruction for the learning of English for Specific Purposes (Little & Kobayashi, 2015).

1.2. Questionnaire validation

Systematic studies of VLS need valid, reliable and practical instruments. Among all the elicitation tools available, questionnaires stand out as a quick instrument that can easily collect data and map out the strategy pattern among a large group of students (Oxford, 2017). Unfortunately, the most popular taxonomies of VLS have never been properly validated; or the validation information has never been formally published. It is therefore the intention of the present article to provide not just an update, but also validation information for the *Vocabulary Learning Questionnaire* (VLQ) based on Gu's (2013) taxonomy.

A basic assumption underlying the use of a questionnaire as an elicitation tool for VLS is that strategies are latent sets and episodes of behavior that can be observed and described. For both research and pedagogical purposes, a convenient tool for "catching" learning strategies is the use of questionnaires. In these cases, sets of statements representative of the strategic learning behavior are presented to learners. Respondents are asked to rate how true of them each statement is or how often they perform the strategy represented by a particular statement. When the cumulative deployment of certain strategic learning behaviors reaches a certain level, we say that the learner uses this strategy. In this sense, like the items in a test, statements in a questionnaire are at best representative samples of strategic learning performance. They can approach a complete representation but will never completely represent the construct being studied. A good questionnaire instrument endeavors to validly elicit the intended construct by systematically making sure that the statements included in the questionnaire are relevant to and representative of the target construct, and that the overall structure of the construct being elicited should resemble as closely as possible the theoretical conceptualization of the construct. In addition, a good questionnaire instrument should make sure that the statements constituting the same scale should

be consistent among themselves. Finally, a good questionnaire should also make sure that the interpretation of the scores and sub-scores is appropriate, and that they are used for the right purposes and among the suitable population.

Validation is a process of making sure an instrument is good enough for its intended purpose. Since major threats to content validity include construct irrelevance and construct under- representation (Messick, 1989), content validation should have started from the moment the VLS construct is conceptualized and operationalized. Construct validation will require trialing among a large-enough sample so as to be sure that the theoretical construct of VLS matches the patterns that emerge from the data.

1.3. The *Vocabulary Learning Questionnaire*

The *Vocabulary Learning Questionnaire* was initially designed to be used among Chinese EFL learners at the tertiary level (Gu & Johnson, 1996). Although it was designed in English, only the translated Chinese language version was used and validated. An updated version (VLQ5) was published in 2003 (Gu & Hu, 2003) but, again only the Chinese version was validated and used. Twenty years after the initial publication, when teachers and learners are still very much interested in the strategic learning of vocabulary, it would be useful to design an ESL version that can be easily understood by English language learners.

The VLQ was originally designed based on both top-down (review of theories and empirical research on vocabulary and learning strategies) and bottom-up (observing and interviewing learners) procedures. It contained a metacognitive component (beliefs and self-regulation) and a cognitive component which followed the natural stages of word learning from initial handling of a new word (i.e., guessing, dictionary use, note-taking) to consolidation and reinforcement (i.e., rehearsing and encoding), and finally to the activation and use of the newly learned word.

The broad coverage resulted in a long questionnaire, with 21 strategies elicited through 108 items in Gu and Johnson (1996). In the early 1990s when the original study was conducted, the Chinese students were not exposed to many strategies for learning vocabulary, and they rarely completed questionnaires. As a result, many students loved the long version and thought that the questionnaire raised their awareness of a whole range of strategies for vocabulary learning. Over the years, when scholars began to adapt the questionnaire for learners in other parts of the world, I have heard many complaints about the questionnaire being too long, even after the updated 90-item version came out in 2003. Thus a major aim of the current update, in addition to creating an ESL version, is to reduce the VLQ to a shorter version while maintaining the same content coverage.

The VLQ makes use of a 7-point Likert scale, ranging from "extremely untrue of me" (1) to "extremely true of me" (7). Likert scales that specify a bipolar continuum of a few points (e.g., 3, 5, or 7 points) have become standard practice in survey research. However, the ordinal nature of the data thus obtained has received on-going debate in terms of statistical analysis. With the increasing ease of online research tools such as Qualtrics, it is felt that a slider bar between bipolar ends with an extended scale (0-100) may provide continuous data with finer grains than traditional Likert scales, which, as a result, may bring an alternative, if not a solution, to the problem presented by ordinal scales. Hence a second aim of the project is to create a slider bar version of the VLQ and see how it performs against the Likert scale version.

2. This study

This study aims to update and validate the VLQ. Specifically, the first aim is to create an ESL version that is easy to read for ESL learners of English. In addition, an online slider-bar format will be trialed. The second aim is to provide validation information detailing various sources of evidence for the validity and reliability of the online version. The research questions I will try to answer include the following:

Research Question 1: What is the evidence for content validity?

1. Is the ESL version easy enough for ESL students?
2. Are all variables and items relevant to and representative of vocabulary learning strategies?

Research Question 2: What is the evidence for construct validity?

1. Does the underlying factor structure match the theoretical structure of vocabulary learning strategies?
2. Do inter-factor correlations show meaningful relationships among the strategies?

Research Question 3: What is the evidence for reliability?
Research Question 4: Is there evidence for predictive validity?

3. Updating the VLQ

3.1. The questionnaire

VLQ5 is a 90-item questionnaire measuring 21 strategies under metacognitive and cognitive dimensions (Gu, 2013). Each item is rated on a 7-point Likert scale.

The section on metacognitive beliefs about vocabulary learning uses a scale from "1 = absolutely disagree" to "7 = absolutely agree." The rest of the questionnaire on strategies uses a scale ranging from "1 = extremely untrue of me" to "7 = extremely true of me."

Two versions of VLQ6 were designed based on VLQ5, a paper version using the same 7-point Likert scale, and an online slider-bar version using the Qualtrics platform. The slider bar version made use of a slider bar with 0 at one end and 100 at the other, as presented in Figure 1.
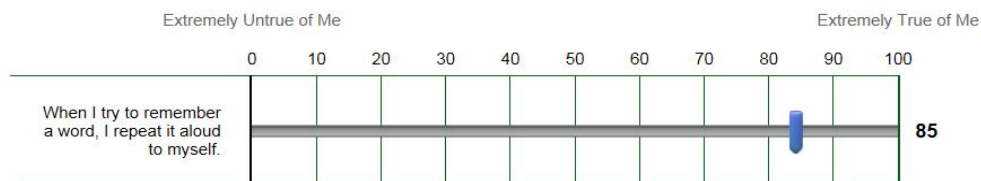

Figure 1 A slider bar item

3.2. Procedures

The updating started with the creation of an ESL version for the 90 Chinese language items in VLQ5 (Table 1). First, all previous English versions were combed through and each Chinese statement was matched with an English version. Then, a research assistant with a recent degree in applied linguistics went through the 90 items one by one and wrote a simplified version. Tom Cobb's Lextutor.ca was used to make sure that the vocabulary of the questionnaire was as simple as possible. The resulting version was labelled as version 6.0. The following criteria were applied to VLQ6.0:

- Include mainly the first 2000 most frequent bands.
- Include as few as possible academic words and off-list words.
- Include as few as possible meta-language jargons.
- Where jargons cannot be avoided, give simple examples.
- Use simple sentence structures.

Table 1 VLQ6 Updating procedures

Version	Purpose	Procedures
VLQ6.0	Create English version of VLQ5.	Find all English statements in previous versions and match them with the 90 statements in VLQ5.
	Make an ESL version.	Go over each statement and simplify them.
VLQ6.1	Trial and fine-tune for clarity.	• Get ESL teacher feedback re clarity of statements. • Trial 1, one advanced level ESL user, to ensure clarity of statements.

VLQ6.2	Create online slider-bar version. Trial among a small group of students.	• Trial 2: four ESL students complete both paper and online versions. Perform item-by-item think-aloud on both versions to: ○ catch understanding issues and simplify statements further, and ○ catch potential similarities and differences for Likert scale and slider bar versions. • Re-write problematic statements after Trial 2.
VLQ6.3	Formal piloting.	• Trial 3: Among a group of 105 ESL learners: ○ statistically compare the Likert-scale version with the slider bar version; ○ catch potentially problematic items; ○ catch problems with online administration.

VLQ6.4	Construct validation.	• Administer questionnaire among 699 participants.
		• Factor analysis.
		○ Do the data support the existing structure of the questionnaire?
		○ Can the questionnaire be shortened?
		  ▪ Should certain categories be deleted?
		  ▪ Can some categories be combined?
		  ▪ Can some items be deleted?
		○ Make sure resulting categories are meaningful.
		○ Make sure resulting categories are internally consistent (Cronbach's alpha).
		• Predictive validity: Relate VLS to learning outcomes.

VLQ6.0 was presented to a group of teachers teaching on the English language proficiency (EPP) program at Victoria University of Wellington, New Zealand. These teachers were asked to provide feedback on the clarity of the statements and the possible problems their students could have when reading the statements. At this consultation session, it was decided that "looking-up strategies" containing five items about how a word is looked up in the dictionary were removed. This was partly due to the complaint from these teachers that the questionnaire was too long, and the deleted items all contained jargons (e.g., *inflected form*, *affixes*, *collocation*) that were not necessarily familiar to the ESL students. The most important reason was that "looking up" a new word was not considered to be a challenging task. After teacher consultations, one advanced EFL learner from China trialed the 85-item version and provided feedback item-by-item about possible misunderstanding and potential difficult words. The resulting version was labelled as VLQ6.1.

At this point, an online version of VLQ6.1 was created on the Qualtrics platform with a slider bar scale ranging from 0 to 100. The default point was placed at 50 on the slider-bar. Respondents could drag the bar left or right to a point they felt best described their use of the strategy. If the bar was not moved by a respondent, the system would record a missing value. Four ESL students from the EPP program were asked to complete both the paper version and the online version, and to think-aloud item-by-item as they completed the questionnaires. These sessions were audio-recorded. The digital files were then transcribed. The

think-aloud procedure was meant to catch potential issues of understanding and simplify the statements further if problems were found. It was also meant to capture the students' perceptions of the similarities and differences between the Likert-scale and slider bar versions. A few issues were found and fixed, and the resulting version was labelled as VLQ6.2.

Next, one hundred and five (105) students in the EPP program were asked to answer the Likert-scale version in a lecture theatre. They were all able to complete the questionnaire within 15-25 minutes. They were also invited to complete the online version within one week. Twenty five of these students answered both the Likert-scale version and the slider bar version. Preliminary analysis of the data from this trial (Gu, 2015) suggested that the online version was comparable to the paper version, and that the students perceived the slider bar version as providing them with choices that were more fine-grained than the 7-point Likert scale version.

The final slider bar version was administered online over a period of three weeks, among 699 students from two Chinese universities. After initial cleaning which removed 17 cases that contained duplicate data or too little data, 682 cases remained to form the final data for analysis. Among them, 127 were male, and 474 female, with 81 missing data. The majority were non-English majors ($N = 423$, 62%). There were 179 English majors (26.2%). 80 students did not declare their majors. Most of the students were studying in Year 1 ($N = 354$, 51.9%), and Year 2 ($N = 164$, 24%). There were also 57 Year 3, 27 Year 4 students, and 80 missing values.

In order to answer Research Question 4, that is, to gauge the predictive validity of the VLQ, the participants in the final round of validation were asked to complete Paul Nation's online vocabulary size test at http://my.vocabulary-size.com/ and report their scores along with other items in the questionnaire.

4. Validating the VLQ

4.1. Content validity

4.1.1. Is the ESL version easy enough for ESL students?

The final version (VLQ6.4, online version) contains 62 items, and 925 tokens. The following is the vocabulary profile of this version.

- First 1000 most frequent words (K1): 90.59%.
- First 2000 most frequent words (K2): 5.30%.
- K1+K2 words: 95.89%.
- Words on the Academic Word List (AWL): 1.84%. The words included: *context*, *create*, *focus*, *items*, *link*, *logical*, *paragraph*, *similar*, *similarities*, *similarly*, *structure*, *text*, *topic*.

- Off-list words: 2.27%. The words included: *collocations*, *grammatical*, *phrase*, *prefixes*, *pronunciation*, *repetition*, *suffixes*, *textbooks*, *usage*, *usages*, *vocabulary*.

K1+K2 words account for almost 96%, nearing the 98% necessary for comprehension ease (Nation, 2013). A close look at the AWL words suggests that none of these words should be unknown for students preparing to study at university. There are still a few technical terms in the off-list category, such as *collocations*, *prefixes* and *suffixes*. Clear examples have been provided where these jargons appear. Other off-list words such as *grammatical* should be easy for ESL students as well.

The sentence structures of all statements are kept simple, and difficult words and jargons have been removed, replaced, or explained in examples. After the teacher consultations and trials among ESL students preparing for university entry in New Zealand, it can be claimed that the latest version is ESL-friendly. In fact, no student has ever voiced any concerns of difficulties in understanding after VLQ6.2, and this was corroborated by the think-aloud protocols (Gu, 2015).

4.1.2. Relevance and representativeness

The only issue of relevance comes when we look at the metacognitive component of the questionnaire. In other words, a question arises whether metacognitive beliefs about vocabulary learning and metacognitive self-regulation of vocabulary learning (such as selective attention and self-initiation) should be included in an instrument on VLS? Strictly speaking, beliefs are not strategies, but are directly linked to the choice and use of strategies. Selecting which word to learn, what information about the word to pay special attention to, and exercising human agency in the proactive learning of vocabulary are very much part of VLS, although they do not deal with the cognitive tasks of vocabulary learning. In other words, if we include only the strategies for cognitive tasks, we may well miss out on the learner self-management aspect of vocabulary learning, which are essentially a crucial part of language learning strategies. I have therefore decided to keep the metacognitive component in the questionnaire.

Coverage becomes a critical issue in maintaining construct representation while at the same time trying to shorten the questionnaire. The main guiding principle in this process was the principle of *parsimony*. Two other principles were *theoretical meaningfulness* and the *guidance of data*. Except for deleting the "look-up strategy," all other deletions were based on the structure that the data revealed. A rule of thumb which was also applied was that there had to be at least three items representing each strategy. Given that strategies are latent variables, there would be serious construct under-representation if the number of items eliciting each strategy drops down below three.

4.2. Construct validity

4.2.1. Factor structure

Exploratory factor analysis (EFA) was the major tool used to explore the underlying structure of the data and see if it matched the theoretical structure of the questionnaire. Two rounds of EFA were performed taking into account sampling concerns. Costello and Osborne (2005) maintain that sampling adequacy is crucially important for EFA, and that a minimum subject to item ratio of 10:1 should be reached before EFA results could be trusted. Maximum Likelihood was the extraction method selected because it produces "more generalizable and reproducible results, as it does not inflate the variance estimates" (p. 6). Promax with Kaiser Normalization was chosen as the method of rotation because oblique rotation normally renders a "more accurate, and perhaps more reproducible, solution" (p. 3) for variables that are expected to correlate to a certain degree.

For the first round of EFA, separate EFA were performed on the following categories: metacognitive beliefs (12 items), metacognitive strategies (10 items), inferencing (10 items), dictionary use (9 items), note-taking (8 items), rehearsal strategies (11 items), encoding strategies (20 items), and activation strategies (5 items). In all cases, both Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity indicated that the data were appropriate for EFA (Field, 2009). The majority of these categories yielded factor patterns corresponding to the theoretical constructs. In other words, the items designed to represent a strategy loaded onto the same factor. Items with weak loadings below .30 or cross-loaded onto different factors (>.32) were removed.

Interestingly, the eight note-taking strategies revealed a factor structure different from the original two-strategy structure (meaning- vs. usage-oriented note-taking). Examined closely, the new structure was meaningful as well. Four items in one of the two factors represented a strategy to choose words for note-taking, while the other four items in the second factor represented a strategy to decide what information goes into the notes. The new factor structure was therefore adopted and new strategy labels given.

The largest category, encoding strategies, turned out to be the most problematic. The original construct had six sub-categories of encoding strategies represented by 20 items: *association*, *visual encoding*, *auditory encoding*, *use of word-structure*, *semantic encoding*, and *contextual encoding*. Instead of a six-factor structure, the first round of EFA produced a 5-factor structure, with two of the three items for *semantic encoding* loading weakly under *contextual encoding*, and the other item loading under *use of word-structure*. Another problematic strategy was *association*. Although all four items loaded onto the same factor, the loadings

were weak; and three out of four items cross-loaded under other factors. Removing the weak and cross-loaded items resulted in a four-factor structure, practically removing association strategies and semantic encoding strategies.

The remaining 62 items were subjected to a second round of EFA. The KMO measure of sampling adequacy was .899, well above the .50 cut-off suggested by Field (2009). Bartlett's test of sphericity was also acceptable (*Chi-square* = 17672.477, *df* = 1891, *p* = .000). Eigenvalue greater than 1 was used as the main criterion for factor extraction. Fifteen factors were extracted. The 15 factors with initial eigenvalues of 1.0 or above explained 67.42% of the total variance (see Figure 2). The extracted 15 factors accounted for 57.42% of the total variance.

Figure 2 Rotated factor pattern

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Commu-nality
1. Once the English words of all my native language meanings have been remembered, English is learned.		.45														.31
2. The best way to remember words is to memorize word lists or dictionaries.		.68														.47
3. The purpose of learning a word is to remember it.		.65														.51
4. A good memory is all you need to learn a foreign language well.		.69														.48
5. Repetition is the best way to remember words.		.72														.49
6. You can only learn a large vocabulary by memorizing a lot of words.		.71														.56
7. The meanings of a large amount of words can be picked up through reading.						.67										.47
8. Learners should pay attention to expressions (e.g., pick up) and collocations (e.g., heavy rain; strong wind) that go with a word.						.78										.68
9. Learners can learn vocabulary simply through reading a lot.						.89										.74
10. The least a learner should know about a word is its spelling, pronunciation, meaning, and its basic usage.						.49										.43
11. I know whether a new word is important in understanding a passage.																.20
12. I know which words are important for me to learn.																.24
13. When I meet a new word or phrase, I know clearly whether I need to remember it.																.19
14. Besides textbooks, I look for other readings that fall under my interest.					.42											.42
15. I wouldn't learn what my English teacher doesn't tell me to learn. (Reversed value)					.71											.57
16. I only focus on things that are directly related to examinations. (Reversed value)					.89											.75
17. I wouldn't care much about vocabulary items that my teacher does not explain in class. (Reversed value)					.83											.70
18. I make use of the logical development in the context (e.g., cause and effect) when guessing the meaning of a word.	.79															.60
19. I use common sense and knowledge of the world when guessing the meaning of a word.	.71															.58
20. I check my guessed meaning in the paragraph or whole text to see if it fits in.	.78															.61

Item									
21. When I don't know a new word in reading, I use my background knowledge of the topic to guess the meaning of the new word.	.78								.66
22. I look for explanations in the reading text that support my guess about the meaning of a word.	.96								.79
23. I make use of the grammatical structure of a sentence when guessing the meaning of a new word.	.59								.47
24. I make use of the part of speech of a new word when guessing its meaning.	.52								.44
25. When I see an unfamiliar word again and again, I look it up.		.48							.27
26. When not knowing a word prevents me from understanding a whole sentence or even a whole paragraph, I look it up.		.65							.39
27. I look up words that are important to the understanding of the sentence or paragraph in which it appears.		.71							.46
28. I pay attention to the examples when I look up a word in a dictionary.		.31							.41
29. When I want to have some deeper knowledge about a word that I already know, I look it up.		.65							.64
30. When I want to know more about the usage of a word that I know, I look it up.		.79							.62
31. I check the dictionary when I want to find out the similarities and differences between the meanings of related words.		.76							.61
32. I make a note when I think the meaning of the word I'm looking up is commonly used.						.99			.81
33. I make a note when I think the word I'm looking up is related to my personal interest.						.80			.72
34. I make a note when I see a useful expression or phrase.						.57			.59
35. I write down the English explanations of the word I look up.				.79					.77
36. I write down both the meaning in my native language and the English explanation of the word I look up.				.94					.52
37. I note down examples showing the usages of the word I look up.				.42					.54
38. I go through my vocabulary list several times until I remember all the words on the list.					.75				.66
39. I make vocabulary cards and take them with me wherever I go.					.67				.54
40. I make regular reviews of new words I have memorized.					.80				.70
41. When I try to remember a word, I say it aloud to myself.							.34		.48
42. When I try to remember a word, I repeat its pronunciation in my mind.							.85		.67
43. Repeating the sound of a new word to myself would be enough for me to remember the word.							.78		.66
44. When I try to remember a word, I write it again and again.								.71	.51
45. I memorize the spelling of a word letter by letter.								.33	.40
46. I write both the new words and their translation in my native language again and again in order to remember them.								.77	.63
47. I act out some words in order to remember them better (e.g., jump).			.53						.62
48. I create a picture in my mind to help me remember a new word.			.86						.83
49. To help me remember a word, I try to "see" the spelling of the word in my mind.			.72						.64
50. I put words that sound similar together in order to remember them.				.95					.79

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
51. When words are spelled similarly, I remember them together.								.80								.70
52. When I try to remember a new word, I link it to a sound-alike word that I know.								.65								.59
53. When I learn new words, I pay attention to pre-fixes, roots, and suffixes (e.g., inter-nation-al).												.51				.42
54. I intentionally study how English words are formed in order to remember more words.												.76				.61
55. I memorize the commonly used roots and pre-fixes.												.74				.61
56. When I try to remember a word, I also try to remember the sentence in which the word is used.							.85									.71
57. I put words in set expressions or sentences in order to remember them.							.90									.79
58. I remember a new word together with the context where the new word appears.							.78									.65
59. I make up my own sentences using the words I just learned.			.52													.53
60. I try to use the newly learned words as much as possible in speech and writing.			.90													.72
61. I try to use newly learned words in real situations.			.86													.74
62. I try to use newly learned words in imaginary situations in my mind.			.78													.73
Initial eigenvalue	13.83	5.38	3.83	2.76	2.23	1.83	1.72	1.48	1.43	1.38	1.29	1.26	1.20	1.12	1.06	
Variance explained (%)	22.31	8.68	6.18	4.45	3.60	2.95	2.78	2.38	2.30	2.23	2.08	2.03	1.94	1.81	1.70	
Accumulated variance explained (%)	22.31	30.99	37.17	41.62	45.22	48.17	50.94	53.32	55.62	57.86	59.94	61.97	63.90	65.72	67.42	

The 15-factor structure neatly corresponded to the theoretical constructs. The two original inferencing strategies (i.e., guessing by making use of global clues, and guessing by the use of local clues) loaded onto the same factor and were combined. The two dictionary strategies (i.e., using dictionary for comprehension, and using dictionary for learning) also loaded onto the same factor and were combined into one strategy. Only one item cross-loaded under both *oral repetition* and *visual repetition*. The item is: "When I try to remember a word, I repeat it aloud to myself." Removing this item would have messed up the factor structure; and it does not make sense to be grouped under *visual repetition*. I therefore decided to retain this item and reword it to: "When I try to remember a word, I say it aloud to myself." Another problem was the metacognitive strategy of *selective attention*. The loadings for all three items were too low to form part of the 15-factor solution. However, selective attention for vocabulary learning has been proven to distinguish between successful and non-successful learners (Gu, 1994). It was therefore decided to retain the three items constituting the *selective attention strategy*.

4.2.2. Factor correlations

Another source of evidence for the construct validity of the questionnaire was to examine the inter-factor correlations. Intra-factor and inter-item correlations among the items that form the same strategy or item-total correlations show convergent validity. This will be reflected in the reliabilities section below. At the same time, inter-factor correlations that are not too high (.70 or above) would

indicate discriminant validity. In other words, these low correlations would indicate that the factors are indeed separate, though correlated, factors.

Table 2 Factor correlation matrix

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Guessing	1													
2. Dictionary	.39	1												
3. Memorise	-.15	-.09	1											
4. Activation	.40	.34	-.06	1										
5. Self-Initiate	.30	.30	-.43	.31	1									
6. Learn	.47	.36	-.16	.09	.23	1								
7. CodingVis	.18	.04	.04	.43	.05	-.04	1							
8. CodingCntxt	.49	.36	-.21	.65	.31	.17	.44	1						
9. CodingAud	.42	.29	-.01	.52	.23	.18	.33	.50	1					
10. NoteInfo	.25	.27	-.03	.56	.21	-.08	.29	.47	.34	1				
11. Wordlist	.22	.11	.13	.64	.18	-.17	.38	.44	.41	.61	1			
12. Notewhat	.48	.49	-.15	.44	.45	.44	.09	.44	.43	.37	.26	1		
13. CodingStr	.40	.31	-.06	.50	.22	.22	.30	.50	.50	.39	.39	.39	1	
14. RepeatOral	.47	.29	-.04	.40	.27	.29	.13	.35	.35	.33	.32	.43	.33	1
15. RepeatVis	.09	.14	.36	.23	-.16	.07	.24	.15	.20	.18	.29	.16	.20	.19

Note. Guessing = Inferencing; Dictionary = Dictionary strategies; Memorize = Vocabulary should be memorized; Activation = Activation strategies; Self-Initiate = Self initiation; Learn = Vocabulary should be learned and used; CodingVis = Visual encoding; CodingCntxt = Contextual encoding; CodingAud = Auditory encoding; NoteInfo = Choosing what information to note down; Wordlist = Using word lists; Notewhat = Choosing which word to note down; CodingStr = Using word structure; RepeatOral = Oral repetition; RepeatVis = Visual repetition

The factor correlation matrix in Table 2 shows good discriminant validity of VLQ6.4. Only three inter-factor correlations were above .6, the highest being .653, between *contextual encoding* and *activation strategies*. This is a reasonable finding, because part of activating a newly learned word is to put it into at least a sentential context.

4.3. Reliability

Cronbach's alpha for each strategy was obtained to collect evidence for the internal consistency reliability of the questionnaire. Table 3 below indicates that VLQ6.4 is a largely reliable instrument. Only one of the 15 strategies, that is, *visual repetition*, had an alpha of .638. The remaining strategies were all above .70. In fact, the overwhelming majority of them were above .80. The strategy that was not in the 15-factor structure, that is, *selective attention*, also showed low but acceptable internal consistency (alpha = .627).

Table 3 Internal consistency reliability (Cronbach's alpha)

Dimensions	Categories	Strategies	Number of items	Alpha
Metacognitive	Beliefs about vocabulary learning	Words should be memorized	6	.821
		Words should be learned through use	4	.813
	Metacognitive regulation	Selective attention	3	.627
		Self-initiation	4	.820
Cognitive	Inferencing	Guessing strategies	7	.884
	Using dictionary	Dictionary strategies	7	.816
	Taking notes	Choosing which word to put into notebook	3	.851
		Deciding what information goes into notes	3	.798
	Rehearsal	Use of word lists	3	.807
		Oral repetition	3	.735
		Visual repetition	3	.638
	Encoding	Visual encoding	3	.833
		Auditory encoding	3	.850
		Use of word-structure	3	.732
		Contextual encoding	3	.874
	Activation	Activation	4	.874

4.4. Predictive validity

It follows that if VLS are meant to improve vocabulary learning results, there should be evidence linking the VLQ to measures of vocabulary learning and other types of second language achievement in general. Based on previous research (e.g., Gu & Johnson, 1996), it is predicted that believing in memorization should negatively affect vocabulary learning. Likewise, visual repetition should also be negatively correlated with vocabulary size. All other strategies should show a positive correlation between strategy use and vocabulary size. In other words, for the majority of strategies, the more often a strategy is used, the larger the vocabulary size should be. In order to see if this predicted relationship can be replicated, vocabulary size was broken into three groups, that is, bottom, middle, and top groups, and one-way analysis of variance (ANOVA) was performed to check if a significant difference can be found between the group mean scores of vocabulary size.

Four hundred and fifty two (452) students provided their vocabulary size scores with a mean vocabulary size of 6,822.34. The vocabulary size for the 33.33 percentile was 5,900, and that for the 66.67 percentile was 7,300. Based on these percentile scores, the whole group was then divided into three subgroups. Group 1 (*n* = 153) included those with a reported vocabulary size of 5,900 or lower; Group 2 (*n* = 146) contained those who reported a vocabulary size between 5,901 and 7200; and Group 3 (*n* = 153) included those whose vocabulary size was 7,201 or above.

Table 4 Relationship between VLS and vocabulary size

		N	Mean	SD	F	p
Memorise	1.00	105	45.7143	17.83686	4.656	.010
	2.00	115	43.6449	16.63307		
	3.00	119	38.7031	18.82705		
	Total	339	42.5511	17.99107		
Learn	1.00	137	78.4982	11.65270	6.363	.002
	2.00	137	77.2445	13.35500		
	3.00	140	82.3500	12.22558		
	Total	414	79.3859	12.59012		
Selective	1.00	125	50.1227	16.13432	10.629	.000
	2.00	107	57.5732	15.46158		
	3.00	118	58.7655	15.80654		
	Total	350	55.3143	16.25128		
SelfInitiate	1.00	122	44.7480	20.66232	15.066	.000
	2.00	111	47.7477	16.25166		
	3.00	122	57.5533	19.59466		
	Total	355	50.0866	19.74810		
Guessing	1.00	118	64.0097	14.12093	6.453	.002
	2.00	117	66.5458	12.09791		
	3.00	130	70.1275	14.09587		
	Total	365	67.0016	13.69706		
Dictionary	1.00	120	60.7381	16.57932	5.492	.004
	2.00	116	62.6700	13.62583		
	3.00	126	67.0193	15.34798		
	Total	362	63.5434	15.44013		
NoteWhat	1.00	123	63.1897	19.23993	3.147	.044
	2.00	123	66.3225	17.56699		
	3.00	130	68.9231	17.71579		
	Total	376	66.1968	18.28496		
NoteInfo	1.00	122	47.2022	21.42532	2.405	.092
	2.00	119	48.0896	20.13105		
	3.00	128	52.5885	20.99162		
	Total	369	49.3568	20.94387		
Wordlist	1.00	124	38.8575	20.51620	3.220	.041
	2.00	125	38.0853	18.82364		
	3.00	126	44.0185	21.01941		
	Total	375	40.3342	20.26227		
RepeatOral	1.00	128	55.7500	18.37311	6.389	.002
	2.00	122	58.7896	16.69723		
	3.00	129	63.4677	17.15946		
	Total	379	59.3553	17.68176		
RepeatVis	1.00	128	55.2786	19.15698	.598	.550
	2.00	123	52.9079	18.83276		
	3.00	130	53.2410	18.09050		
	Total	381	53.8180	18.67403		
CodingVis	1.00	124	49.3333	22.62342	.333	.717
	2.00	123	48.7073	20.38726		
	3.00	138	50.8116	21.47693		
	Total	385	49.6632	21.47929		

CodingAud	1.00	127	53.6010	21.11003	6.294	.002
	2.00	124	54.0134	20.25922		
	3.00	136	61.5172	19.95195		
	Total	387	56.5151	20.71466		
CodingStr	1.00	125	55.3840	21.06940	9.426	.000
	2.00	122	57.7377	16.91556		
	3.00	137	64.9538	17.80574		
	Total	384	59.5460	19.07503		
CodingCntxt	1.00	119	49.5938	23.33641	3.305	.038
	2.00	120	52.9583	20.66796		
	3.00	133	56.6090	20.95395		
	Total	372	53.1873	21.78992		
Activation	1.00	127	48.9528	21.81414	2.819	.061
	2.00	115	48.6587	19.08534		
	3.00	126	54.1012	19.77069		
	Total	368	50.6236	20.39560		

Table 4 shows the predictive validity of VLQ6.4. Believing that vocabulary should be memorized showed a negative relationship with vocabulary size. Although the entire group tended not to hold this belief (M = 42.55), the bottom group (Group 1) had the largest mean score for this belief (M = 45.7), and the top group (Group 3) had the lowest mean score (M = 38.7). Another strategy variable that was predicted to hold a negative relationship was *visual repetition*. Although one-way ANOVA did not produce a significant overall F value, and even post-hoc comparisons between groups did not produce any significant differences, Table 4 still shows that the bottom group used this strategy more often than the other two groups. The only surprise was the *visual encoding* strategy, which included items such as "I create a picture in my mind to help me remember a new word." Although the top group used the strategy slightly more than the other groups, the three groups basically did not significantly differ from each other in using this strategy, and the overall mean score of 49.67 basically indicated that the Chinese students in this study had a lukewarm endorsement of visual encoding for vocabulary learning.

5. Using VLQ6.4

5.1. Potential uses of VLQ6.4

The VLQ has been designed as a research tool. It can be used to explore the range of strategies a group of students use for the learning of vocabulary. It can also be used to explore the relationship between different VLS and vocabulary learning outcomes. In a study focusing on strategy instruction for the improvement of strategic learning behavior, VLQ6.4 can be used as a pre-test and post-test to show the extent of learning of various strategies as a result of the instruction.

With this simplified and shortened ESL version, the VLQ should be a useful diagnostic tool for teachers and learners. For example, if a group of learners complain that their vocabulary is limited and that they do not know how best to learn vocabulary, VLQ6.4 can be quickly administered in this group of students to diagnose the repertoire of strategies and the frequency of strategy use as a group measure or for a specific student. The least the questionnaire can do is to raise the awareness of students about their vocabulary learning behaviors so that they can become more reflective about their own learning.

VLQ6.4 should be suitable for ESL learners with a basic grasp of the first 2,000 most frequent words in English. It is most suitable for those who are studying or preparing to study at the tertiary level. ESL learners beyond the basic level who need to examine their own vocabulary learning processes will also find the questionnaire useful as an awareness-raising or diagnostic tool. Teachers of English at the beginning of a course may use VLQ6.4 as a quick tool to help them gauge their students' VLS and make decisions about their vocabulary teaching. Teachers who have already found vocabulary learning problems among their students may use it to catch potential learning strategy causes of the problems they have discovered.

5.2. Administration, scoring and interpretation

It is very easy to set up the online slider bar version using an online survey service such as Qualtrics. The benefit of such an online tool is the convenience and flexibility for data collection. Once the survey is set up, the target population can be approached with an email with a link to the survey. Students can then do the survey at their own time and using their own laptops or even smart phones. Once it is done, the data can be downloaded for processing. Alternatively, a simple paper version with a scale from 0 to 100 can be easily reproduced, with spaces provided every 10 or 5 points. The only drawback is that the data will need to be manually entered for later processing.

Scoring is done in two steps. The first step is to obtain the average item score (the mean score) under each strategy, in other words, by adding all items up and dividing the sum by the number of items. For example, to get a respondent's strategy score for the *activation strategy*, all four items for the strategy are added up, and the sum is then divided by 4. Mapping out the mean score for each strategy on the VLQ would give the user a quick idea about the repertoire and level of strategy use for each sample or respondent observed. For research purposes, this mean score is enough. For diagnostic purposes, the second step will help.

To diagnose VLS problems, I suggest one standard deviation above and below the mean score as the cut-off points for "moderate use." Theoretically, this accounts for 34% above the mean and 34% below the mean. Another way of seeing

it is that we would see the level of strategy use being "moderate" if a student's strategy use falls within 68% around the mean score. This group's strategy use can be boosted further to improve strategic learning, but there is little to worry about. For a strategy that is supposed to be good for vocabulary learning, if a student's score is below -1 standard deviation, s/he belongs to the 16% at the bottom end, which begs for attention. Anyone above the +1 standard deviation belongs to the 16% at the top who already use the strategy with experience. On the contrary, for a strategy that is normally linked to low vocabulary achievement (e.g., *visual repetition*), high strategy use would be reason for alarm.

Table 5 Interpreting VLQ6.4 scores (slider bar version, 0-100)

	Mean	SD	Low strategy use	Moderate strategy use	High strategy use
Memorize	41	18	(41-18) 23 or below	24-58	(41+18) 59 or above
Learn	79	12	(79-12) 67 or below	68-90	(79+12) 91 or above
Selective	56	16	(56-16) 40 or below	41-71	(56+16) 72 or above
SelfInitiate	53	20	(53-20) 33 or below	34-72	(53+20) 73 or above
Guessing	67	14	(67-14) 53 or below	54-80	(67+14) 81 or above
Dictionary	65	15	(65-15) 50 or below	51-79	(65+15) 80 or above
NoteWhat	67	18	(67-18) 49 or below	50-84	(67+18) 85 or above
NoteInfo	51	21	(51-21) 30 or below	31-71	(51+21) 72 or above
Wordlist	42	20	(42-20) 22 or below	23-61	(42+20) 62 or above
RepeatOral	59	18	(59-18) 41 or below	42-76	(59+18) 77 or above
RepeatVis	53	19	(53-19) 34 or below	35-71	(53+19) 72 or above
CodingVis	50	21	(50-21) 29 or below	30-70	(50+21) 71 or above
CodingAud	57	20	(57-20) 37 or below	38-76	(57+20) 77 or above
CodingStr	60	19	(60-19) 41 or below	42-78	(60+19) 79 or above
CodingCntxt	55	21	(55-21) 34 or below	35-75	(55+21) 76 or above
Activation	53	20	(53-20) 33 or below	34-72	(53+20) 73 or above

For the busy classroom teacher or individual learner who does not want to do the tedious calculation of standard deviations, a rough guide is to add or delete 20 from the arithmetic average (Table 5). The average standard deviation of all the strategies in VLQ6.4 for this sample is 18.28. There may well be a slightly different distribution of data with another sample. Although the sample in this study is not large enough to play the role of a norm, it can still be a reasonable guide for quick diagnostic purposes. As a crude classification measure, it seems safe that each student's strategy use can be classified into high, moderate, and low categories by adding to or deleting 20 from the average score (mean) of each individual strategy.

6. Conclusion and remaining issues

The present update and validation effort has focused mainly on producing an ESL version of the VLQ, shortening the questionnaire, and on trialing the use of the 100-point slider bar. As such, conceptualization of the construct of VLS has not changed since the beginning of the instrument. Readers should be aware that the VLQ focuses on strategies for learning single words and not multi-word units. In addition, the VLQ does not distinguish between strategies for vocabulary breadth from those for the depth of vocabulary knowledge. The VLQ does not explicitly cover strategies for the development of automaticity and appropriateness in vocabulary use either. Furthermore, partly because of the focus on vocabulary breadth (size), the overwhelming majority of items cover strategies for the learning of receptive vocabulary. Only the last category with 4 items in the latest version attempts to elicit strategies for "activation." In addition, the VLQ covers metacognitive and cognitive, but not social strategies and affective strategies. Depending on the purpose, future instruments on VLS should aim at covering the areas of vocabulary learning that the VLQ does not cover adequately.

This report has mainly included the validation of the slider bar version of the VLQ. When I started exploring the slider bar as a survey scale, I had hoped that eventually it might be a replacement for the Likert scale. The reasoning was simple. If the slider bar was not perceived to be psychologically different from the Likert scale when the students responded to the items, and if the respondents saw the slider bar as providing more fine-grained choices than the Likert scale, the slider bar should be the right choice. Statistically, Hatch and Lazaraton (1991) suggested that "wider scales encourage more precision in rating and thus approach equal intervals" (p. 57). When I stretched the scale from seven points to a hundred points on the slider bar, I had hoped that the resulting data would be approaching interval more than ordinal data. In the piloting round for VLQ6.3 which was not reported in this article, 25 students answered both the paper version and the online version. While the students themselves did see the slider bar as giving them more choices and that they generally did not feel much of a difference, the judgment process in responding to the slider bar was still found to be very much ordinal (Fernandez, Liu, Costilla, & Gu, forthcoming). Despite the possibility of a seemingly continuous scale, human judgement is by nature imprecise, and the majority of the choices were placed on or near 0, 5, 10, 15, 20, 25, 30... and 100. In other words, the 100-point slider bar data is still more ordinal than interval, although it can be taken as similar to the marks we give our students for their written assignments.

The VLQ has shown considerable stability over the years. However, students from different cultural and educational backgrounds may demonstrate different

preferences in strategy choice and use. Responses to questionnaires are known to be dependent on the specific sample on which a questionnaire is administered (Horwitz, 2016). Therefore, even if the basic construct of *vocabulary learning strategies* remains the same, interpretations of the construct may vary among different samples. In this sense, there is no such thing as an absolute "validated version" of a questionnaire instrument. For research purposes, I suggest revalidation whenever possible, rather than adopting the "validated version" of the VLQ. This is especially true of reliability if the new study is examining the same construct as defined in the VLQ. Even if the exact version of VLQ6.4 is used, a re-calculation of reliability for each strategy category backed up by a complete description of the sample would be warranted. Diagnostic uses can of course exercise the license of flexibility.

References

Ahmed, M. O. (1989). Vocabulary learning strategies. In P. Meara (Ed.), *Beyond words* (pp. 3-14). London: British Association for Applied Linguistics, in association with Centre for Information on Language Teaching and Research.

Brown, C., & Payne, M. E. (1994). Five essential steps or processes in vocabulary learning. In *TESOL '94*. Baltimore, MD.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7). Retrieved from http://pareon line.net/getvn.asp?v=10&n=7

Fernandez, D., Liu, I., Costilla, R., & Gu, P. Y. (forthcoming). Assigning scores for ordered categorical responses.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Gu, Y. (1994). Vocabulary learning strategies of good and poor Chinese EFL learners. In N. Bird, P. Falvey, A. B. M. Tsui, D. M. Allison, & A. McNeill (Eds.), *Language and learning* (pp. 376-401). Hong Kong: The Education Department.

Gu, Y. (2003). Fine brush and freehand: The vocabulary learning art of two successful Chinese EFL learners. *TESOL Quarterly, 37*(1), 73-104.

Gu, Y. (2013). Vocabulary learning strategies. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Blackwell Publishing. https://doi.org/10.1002/9781405198431.wbeal1329

Gu, Y. (2015). Designing and validating a visual analogue scale version of the Vocabulary Learning Questionnaire. Presented at the Situating Strategy Use (SSU2015), Klagenfurt, Austria: Alpen-Adria Universität.

Gu, Y., & Hu, G. W. (2003). Cihui xuexi celue, cihuiliang, yu yingyu chengji de bianhua. In Q. F. Wen & L. F. Wang (Eds.), *Yingyu xuexi celue shizheng yanjiu* (pp. 388-409). Xi'an: Shaanxi Normal University Press.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*(4), 643-679.

Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education.* New York: Cambridge University Press.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* Boston: Heinle & Heinle Publishers.

Horwitz, E. K. (2016). Factor structure of the foreign language classroom anxiety scale comment on Park (2014). *Psychological Reports, 119*(1), 71-76. https://doi.org/10.1177/0033294116653368

Kojic-Sabo, I., & Lightbown, P. M. (1999). Students' approaches to vocabulary learning and their relationship to success. *Modern Language Journal, 83*(2), 176-192.

Little, A., & Kobayashi, K. (2015). Vocabulary learning strategies of Japanese life science students. *TESOL Journal*, *6*(1), 81–111. https://doi.org/10.1002/tesj.141

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. https://doi.org/10.2307/1175249

Mizumoto, A., & Takeuchi, O. (2009). Examining the effectiveness of explicit instruction of vocabulary learning strategies with Japanese EFL university students. *Language Teaching Research*, *13*(4), 425-449.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). New York: Routledge.

Parry, K. (1993). Too many words: Learning the vocabulary of an academic subject. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 109-129). Norwood, NJ: Ablex.

Ranalli, J. M. (2003). *The treatment of key vocabulary learning strategies in current ELT coursebooks: Repetition, resource use, recording* (MA thesis). University of Birmingham, Birmingham.

Ranalli, J. M. (2012). *The VVT Project: A web-based platform for strategy instruction and research into self-regulated learning of L2 vocabulary*. Doctoral dissertation, Iowa State University, Ames, Iowa.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *Modern Language Journal*, *79*, 15-28.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*(1), 78-102.

Zhang, X., & Lu, X. (2015). The relationship between vocabulary learning strategies and breadth and depth of vocabulary knowledge. *Modern Language Journal*, *99*(4), 740-753. https://doi.org/10.1111/modl.12277

APPENDIX

Vocabulary Learning Questionnaire (Version 6.4): Categories, strategies, items

Categories	Strategies	Items
Beliefs about vocabulary learning	Words should be memorized	1. Once the English words of all my native language meanings have been remembered, English is learned. 2. The best way to remember words is to memorize word lists or dictionaries. 3. The purpose of learning a word is to remember it. 4. A good memory is all you need to learn a foreign language well. 5. Repetition is the best way to remember words. 6. You can only learn a large vocabulary by memorizing a lot of words.
	Words should be learned through use	7. The meanings of a large amount of words can be picked up through reading. 8. Learners should pay attention to expressions (e.g., pick up) and collocations (e.g., heavy rain; strong wind) that go with a word. 9. Learners can learn vocabulary simply through reading a lot. 10. The least a learner should know about a word is its spelling, pronunciation, meaning, and its basic usage.
Metacognitive strategies	Selective attention	11. I know whether a new word is important in understanding a passage. 12. I know which words are important for me to learn. 13. When I meet a new word or phrase, I know clearly whether I need to remember it.
	Self-initiation	14. Besides textbooks, I look for other readings that fall under my interest. 15. I wouldn't learn what my English teacher doesn't tell me to learn. (Reversed value) 16. I only focus on things that are directly related to examinations. (Reversed value) 17. I wouldn't care much about vocabulary items that my teacher does not explain in class. (Reversed value)
Inferencing	Guessing strategies	18. I make use of the logical development in the context (e.g., cause and effect) when guessing the meaning of a word. 19. I use common sense and knowledge of the world when guessing the meaning of a word. 20. I check my guessed meaning in the paragraph or whole text to see if it fits in. 21. When I don't know a new word in reading, I use my background knowledge of the topic to guess the meaning of the new word. 22. I look for explanations in the reading text that support my guess about the meaning of a word. 23. I make use of the grammatical structure of a sentence when guessing the meaning of a new word. 24. I make use of the part of speech of a new word when guessing its meaning.
Using dictionary	Dictionary strategies	25. When I see an unfamiliar word again and again, I look it up. 26. When not knowing a word prevents me from understanding a whole sentence or even a whole paragraph, I look it up. 27. I look up words that are important to the understanding of the sentence or paragraph in which it appears. 28. I pay attention to the examples when I look up a word in a dictionary. 29. When I want to have some deeper knowledge about a word that I already know, I look it up. 30. When I want to know more about the usage of a word that I know, I look it up. 31. I check the dictionary when I want to find out the similarities and differences between the meanings of related words.

Taking notes	Choosing which word to put into notebook	32. I make a note when I think the meaning of the word I'm looking up is commonly used. 33. I make a note when I think the word I'm looking up is related to my personal interest. 34. I make a note when I see a useful expression or phrase.
	Deciding what information goes into notes	35. I write down the English explanations of the word I look up. 36. I write down both the meaning in my native language and the English explanation of the word I look up. 37. I note down examples showing the usages of the word I look up.
Rehearsal	Use of word lists	38. I go through my vocabulary list several times until I remember all the words on the list. 39. I make vocabulary cards and take them with me wherever I go. 40. I make regular reviews of new words I have memorized.
	Oral repetition	41. When I try to remember a word, I say it aloud to myself. 42. When I try to remember a word, I repeat its pronunciation in my mind. 43. Repeating the sound of a new word to myself would be enough for me to remember the word.
	Visual repetition	44. When I try to remember a word, I write it again and again. 45. I memorize the spelling of a word letter by letter. 46. I write both the new words and their translation in my native language again and again in order to remember them.
Encoding	Visual encoding	47. I act out some words in order to remember them better (e.g., jump). 48. I create a picture in my mind to help me remember a new word. 49. To help me remember a word, I try to "see" the spelling of the word in my mind.
	Auditory encoding	50. I put words that sound similar together in order to remember them. 51. When words are spelled similarly, I remember them together. 52. When I try to remember a new word, I link it to a sound-alike word that I know.
	Use of word-structure	53. When I learn new words, I pay attention to prefixes, roots, and suffixes (e.g., inter-nation-al). 54. I intentionally study how English words are formed in order to remember more words. 55. I memorize the commonly used roots and prefixes.
	Contextual encoding	56. When I try to remember a word, I also try to remember the sentence in which the word is used. 57. I put words in set expressions or sentences in order to remember them. 58. I remember a new word together with the context where the new word appears.
Activation	Activation	59. I make up my own sentences using the words I just learned. 60. I try to use the newly learned words as much as possible in speech and writing. 61. I try to use newly learned words in real situations. 62. I try to use newly learned words in imaginary situations in my mind.