Validating a Self-assessment Questionnaire on Vocabulary Knowledge

Seyyed Mohammad Alavi
Associate professor, University of Tehran
Is'haaq Akbarian
Lecturer, University of Qom

Abstract

Language testing researchers suggest various test formats and procedures to measure vocabulary knowledge. This study investigated a) the relationship between EFL learners' self-rating of vocabulary knowledge on a newly designed self-assessment questionnaire and their performance on Vocabulary Levels Test, and b) the possibility of using self-assessment report as a valid basis for placement purposes. The results from three groups of participants (N=295) with different vocabulary proficiency levels show a) a high internal consistency among questionnaire items and b) a moderate correlation between self-assessment and VLT scores. The low group, knowing the most frequent 3,000 word families, tended to overestimate and the high group, with over 5,000 word families, tended to underestimate their vocabulary knowledge. The middle group with over 3,000 and below 4,000 word families was more realistic. Apparently, less proficient learners may have little knowledge about what they do and more proficient ones may be aware of the limits of their knowledge. This study suggests a) test developers to design innovative and systematic instruments for self-assessment, b) teachers to provide EFL learners with practice in self-assessment, and c) language testers to use self-assessment data elicitation

procedures along with a valid test for grading purposes and providing complementary information.
Key words: Vocabulary knowledge; self-assessment; EFL learners; Vocabulary Levels Test; questionnaire

1. Introduction

Vocabulary knowledge appears to have a crucial role in almost all aspects of language education. In recent literature, some reasons are suggested for pursuing further vocabulary studies enthusiastically. First, there is the lack of a comprehensive theory of how vocabulary is acquired (Jiang, 2000) and many learners are not developing their lexicons adequately in many English language contexts (Barrow, Nakanishi, & Ishino, 1999). Also, vocabulary is still not being given a position of importance by the second language acquisition (SLA) community (Hunt & Beglar, 2005). More importantly, there is the question of whether "there are any special considerations for assessing vocabulary for students learning English as a second language" (Pearson, Hiebert, & Kamil, 2007, p. 290).

Research on vocabulary acquisition entails having an understanding of what 'knowing a word' means so that one can use appropriate tools and procedures to measure vocabulary knowledge (Bogaards, 2000). Researchers offer some assumptions and concepts about the dimensions of word or lexical knowledge or competence (Henriksen, 1999; Nagy & Scott, 2000; Nation 1990, 2001; Richards, 1976) that may usefully be employed as the content of vocabulary tests. Meara (1996) believes that it is ideal from an assessment perspective to construct a vocabulary test entailing measures of each of these types of knowledge of particular words. In practice, however, it would be difficult to do this for more than a handful of items. As Greidanus, Beks, and Wakely (2005) state, there are no tests concurrently examining all aspects of lexical knowledge like form, position, function, and meaning.

One possible solution is a self-assessment questionnaire that intends to tap, at least generally, the learners' own judgment about various components of their own vocabulary knowledge. In the present study, 'self-assessment' means judgments or beliefs of the

learners about their foreign language learning abilities and performance, including assessment for placement (Brantmeier, 2006; Heilenman, 1990; Oscarson, 1980; Schwartz, 1985).

2. Review of the Literature

Self-assessment can be beneficial to learning for many reasons, and thus should be encouraged in language classes. The successful use of self-assessment tools for placement purposes in some settings and their ease and limited expense of administration indicate the value of reconsidering self-assessment (LeBlanc & Painchaud, 1985; Wesche, Paribakht, & Ready, 1993). Also, a learner-centered curriculum falls short of its definition if learners are involved in decisions regarding the content of the curriculum and how it is taught but excluded from the process of evaluating curriculum outcomes, including their own learning achievement (Little, 2005).

Self-assessment motivates students to look at their strengths and weaknesses and become more autonomous (Coombe & Canning, 2002), broadens the learners' experience within the realm of assessment (Oscarsson, 1989), and is a prerequisite for a self-directed learner (Todd, 2002). As a result, the more learners are able to identify their strengths and weaknesses during a task (self-assessment), the more likely they will be able to feel a critical sense of mastery on that task (self-efficacy) (Coronado-Aliegro, 2006). In addition, self-assessment increases student and teacher motivation (Ross, 1998) and is highly motivating in terms of goal-orientation.

Formal contexts of language education provide learners with languages that are to be used in the world beyond the classroom. This demands a capacity for accurate self-assessment allowing learners to turn occasions of target language use into opportunities for further explicit language learning (Little, 2005). At this part, we firstly present some studies favouring self-assessment, followed by others questioning it.

2.1 Studies Favouring Self-assessment

Some empirical studies highlight the potential of language self-assessments. The first research line deals with self-assessment and skills. Clark (1981) compared self-assessments of speaking,

reading, listening, and writing with Foreign Service Interview scores and reading and listening test scores, and found correlations of almost 0.60, accounting for about 36 per cent of the shared variance. Also, LeBlanc and Painchaud (1985) found correlations of 0.80 and 0.82 between proficiency test scores and self-assessments of the four skills. Thus, 64 per cent and 67 per cent of the shared variance are accounted for by the self-assessment results alone. Moreover, Wilson (1999) administered the Test of English for International Communication TOEIC to some 900 Swiss workers enrolled in ESL training programs and collected self-ratings that were translated and administered in the participants' native languages (either French or German). Self-ratings correlated 0.75 and 0.70 with TOEIC listening and reading scores, respectively. Their squared results explain about 56 per cent of R = 0.75 and 49 per cent of R = 0.70. Tannenbaum, Rosenfeld, Breyer, and Wilson (2000) administered the TOEIC and 75 can-do statements (15 each in listening, reading, speaking, writing, and interactive – speaking/listening) to some 8,000 examinees and found correlations of TOEIC reading with self-ratings across the five domains to be 0.65, explaining about 0.42 of the shared variance.

Some other researches lend support for the hypothesis that self-assessment can be accurate for placement. Experimenting with ESL students, Krausert (1991) recommended that educators should utilize self-assessment instead of standardized exams for ESL placement in university programs. Birckbichler, Corl, and Deville (1993, as cited in Brantmeier, 2006) utilized a self-assessment questionnaire as part of the placement exam administered at Ohio State University in the USA and reported that self-assessment correlated higher than any other variable with the placement scores. Hargan (1994) compared a traditional multiple-choice placement test for grammar and a self-placement procedure with university students and found that both instruments indicated the same level of placement. Deville and Deville (1999) suggest self-assessment as part of the procedures for estimating a starting point for computer-adaptive testing for L2 placement.

Further lines of research indicate the importance of self-assessment in other areas of language assessment. For instance, Brown (2005) examined the usefulness of annotated samples of student writing as guides for learners to evaluate their own task-based writing performances. With a small sample of students, it turned out to be accurate, reliable, and useful.

Oscarson (1978) used a variety of rating instruments for self-assessment including scaled descriptions of performance and showed that adults studying EFL were able to make fairly accurate appraisals of linguistic abilities. Heilenman (1990) reported a correlation of 0.33 between course grades and undergraduate students' self-assessments of their French language skills (grammar, vocabulary, accuracy, and fluency), accounting for a small amount of the shared variance, $R^2 = 0.109$.

Ross (1998) carried out a two-phase study of self-assessment. Phase one was a meta-analysis of studies (60 correlations). The average correlation appeared strong for receptive skills. For reading, he located 23 correlations, with an average of $r = 0.61$. The average correlation and effect size for reading appear robust. The magnitude of self-assessment of listening for 18 correlations equals 0.65. However, learners are actually less adept at estimating their own productive skills. The correlation for 29 studies on self-assessment of speaking appeared 0.55 and 15 studies on that of writing skill was 0.52. So Ross (1998) echoed Blanche and Merino's (1989) conclusion that "self-assessment typically provides robust concurrent validity with criterion variables" (p. 16).

In phase two, he analyzed the validity of a self-assessment instrument. 236 'just-instructed' EFL learners completed self-assessments of functional English skills derived from instructional materials and from general proficiency criteria. The learners' teachers also assessed each of the 236 learners. The criterion variable was an achievement test written to assess mastery of the just-completed course materials. He then concluded, "Provided that the content validity requirement is met, the overall picture indicates that there is clear potential for predictive accuracy of criterion skills based on self-assessment measures" (p. 17).

Brantmeier (2005) found that levels of self-assessed abilities positively correlated with levels of enjoyment. In Roever and Powers' (2005) study, 115 participants – German, Mexican, Korean, and Taiwanese – completed self-assessments in both their native languages and English. They found comparable responses in both languages in terms of reliability, level, and variation.

2.2 Studies Questioning Self-assessment

Comparably, fewer studies question the potential in self-assessment. For example, Bayliss (1991) was unable to reproduce the results obtained by LeBlanc and Painchaud (1985), referred to earlier.

Examining the relationship between learners' self-assessment and scores on an online reading placement test as well as subsequent reading performance, Brantmeier (2006) found that self-assessment of L2 reading ability, as measured before and after reading via a 5-point scale, was not an accurate predictor variable for placement or subsequent performance. Her findings do not offer conclusive evidence about self-assessment value as a predictor of performance on Computer Based Testing or as an indicator of subsequent classroom performance, but she suggests that a more contextualized, criterion-referenced self-assessment instrument may be more beneficial for placing advanced readers in the USA.

Some researchers (e.g. Davidson & Henning, 1985, as cited in Roever & Powers, 2005) consider an inherent weakness of self-reports of language abilities in that the self-assessors might be prone to overestimation. Todd (2002) states that most teachers do not use self-assessment because learners do not rate any real-world language performance, rather rate their own beliefs and perceptions with little or no evidence on which to base their assessments.

Statement of the Problem

The inconsistent findings of research on self-assessment highlight the need for more research. A similar plea was also voiced by some scholars (e.g. Brantmeier, 2006; Coombe, 2002; Oscarson, 1997; Ross, 1998), acknowledging the scarcity of research in the area of self-assessment in FL or L2 context. Therefore, the present research is an attempt to provide empirical evidence concerning the use of

self-assessment factor as a predictor of vocabulary knowledge, operationalized here as vocabulary size. Vocabulary size forms an important part of language proficiency. More importantly, vocabulary size tests can discriminate between groups of learners (Meara, 1992) and are useful in admissions and in placing students into appropriate institutional placement levels within a program (Laufer & Nation, 1999; Schmitt, 1994). If appropriately used, they can allow teachers to identify and remedy deficiencies in their students' vocabularies (Schmitt, 1994) and can assess a wide range of words in a relatively short amount of time. This enhances their usefulness in assessing students with widely varying proficiencies.

One kind of evidence for the validity of our newly designed vocabulary knowledge self-assessment questionnaire is to see if it distinguishes between different levels of language proficiency, as vocabulary size tests such as Vocabulary Levels Test does. Therefore, the following research questions (RQ) guided the study:

1.Is there any relationship between participants' self-assessment ratings of vocabulary knowledge and their performance on Vocabulary Levels Test?

2.Does the relationship hold the same for males and females?

3.Can EFL learners' vocabulary knowledge self-assessment report be a valid basis for grouping them into different levels of vocabulary knowledge?

3. Method
3.1 Participants
Participants were from different Iranian universities. The data for analysis were provided by 295 of the initial 306 EFL participants. Table 1 shows the participants' profile.

Table 1: Participants of the Study

Gender	N	Level	N
Male	133	Undergraduate students	236
		Graduate students	64
Female	173	MA holders	6
Total	306	Total	306

N = number

3.2 Materials

The participants were invited to respond to two instruments of this study, i.e. Vocabulary Levels Test (henceforth VLT) and Vocabulary Knowledge Self-assessment Questionnaire, as described below:

3.2.1 Vocabulary Levels Test (VLT)

Version 2 of VLT, revised and validated by Schmitt et al. (2001), was used in this study. VLT views vocabulary of English as consisting of a series of levels based on frequency of occurrence. Each level has the most frequent 1000 words, the next most frequent 1000 words, and so on (Nation, 1990). The levels are highly scalable.

Using factor analysis and personal interviews, Schmitt et al. (2001) showed that the test is essentially unidemensional, that examinees accept the test, and that answers on the test do reflect underlying lexical knowledge. In the present study with 295 participants, a reliability analysis at the four 1000-, 3000-, 5000-, and 10000-word frequency levels shows a coefficient of 0.96.

VLT uses word-definition matching format and measures word knowledge at five levels: 2000, 3000, 5000, 10,000, and Academic Vocabulary section. Academic Vocabulary section was not used in this study since its words are different in kind from the other levels and should not be included in the profile comparison. They are based on the criteria of coverage and range across a variety of academic texts (Coxhead, 2007; Schmitt et al, 2001). Each level contains 30 items. A sample of the test appears in Figure 1.

Participants must choose the right word that goes with each meaning. They must write the number of that word next to its meaning. Here is an example.
1 business
2 clock ----- part of a house
3 horse ----- animal with four legs
4 pencil ----- something used for writing
5 shoe
6 wall

Figure 1: *A sample of Vocabulary Levels Test (VLT)*

Test takers were asked to select three out of the six words to match the three definitions on the right. In total, at each level, 30

definitions need to be matched to 30 words out of 60 (Schmitt et al., 2001). All the words at each level belong to the same part of speech, VLT does not group words and definitions that are related in meaning. The test is not intended to require the testees to differentiate between semantically related words or to show an awareness of shades of meaning. The ratio of the nouns, verbs, and adjectives at each level is 3: 2: 1 (Read, 1988, 2000).

VLT appears to be practical, economical, easy to administer and interpret, and can be completed in a short time. Each answer is marked as correct or incorrect. Answering 24 items correctly is the criterion for being competent at a level (N. Schmitt, personal communication, May 9, 2008). If a learner reaches the criterion at a level, say, the 5000-word level, he or she has most probably mastered the 2000 and 3000 levels as well (Read, 1988).

3.2.2 Vocabulary Knowledge Self-assessment Questionnaire

As there was no instrument to serve the purpose of this study, it was required to design, develop, and validate an instrument for investigating EFL learners' self-assessment on vocabulary knowledge. To meet this purpose, it was felt that a well-established and widely accepted format in vocabulary research was needed, thus defining the construct of vocabulary knowledge and the design of self-assessment questionnaire for this study.

Researchers recognize many dimensions and degrees of knowing a word. Cronbach (1942, as cited in Henriksen, 1999, p. 305) refers to a multidimensional model of word knowledge, including knowledge of a word's referential meaning as well as its different intentional or sense relations to other words in the vocabulary, such as paradigmatic (autonomy, synonymy, hyponymy, gradation) and syntagmatic relations (collocational restrictions). This model also entails knowledge of the syntactic and morphological restrictions and features of a lexical item.

Considering the nature of lexical competence, Richards (1976) suggests a) the degree of probability of encountering a word in speech or print, b) the limitations imposed on its use according to variations of function and situation, c) the syntactic behavior associated with it, d) its underlying form and the derivations made

from it, e) the network of associations between that word and other words in language, f) the semantic value of a word, and g) many of the different meanings associated with it. Nation (2001, p. 27) illustrates what is involved in knowing a word in details. See Table 2 for the aspects of the model.

Table 2: *Aspects of Word Knowledge*

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency …)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note: In column 3, R = receptive knowledge, P = productive knowledge.

Pearson et al. (2007) cite some important features of word knowledge identified by Nagy and Scott (2000) as *incrementality*, *multidimensionality*, *polysemy*, *interrelatedness*, *heterogeneity*, *the ability to reflect on and manipulate vocabulary*, illustrating the complexity of vocabulary. Commenting on a few studies of vocabulary that attend to these variables systematically, Pearson et al. (2007, p. 287) state that "these variables do, however, suggest important new directions for exploration in vocabulary research."

It can be argued that there is a need for a global instrument to assess all the aspects of vocabulary knowledge simultaneously. The

conceptual framework of vocabulary knowledge for designing vocabulary knowledge self-assessment questionnaire (henceforth the VKS questionnaire) is largely based on the collective strength of some of these dimensions, as discussed above.

To initiate the development of the instrument, the items in the existing instruments dealing with beliefs, attitudes, and other individual factors in language learning (Brantmeier, 2006; Roever & Powers, 2005; Sakui & Gaies, 1999) were critically reviewed. Borrowing a few items from these instruments, some other items were written based on the above conceptual framework of vocabulary knowledge. These items were presented to five English university instructors, who were asked to 1) evaluate the items in terms of their relevance, comprehensiveness, appropriateness, precision, clarity, brevity, and coverage in the area of vocabulary knowledge, 2) judge the relevance of the items for college-level English education, and 3) suggest additional topics for the items or revisions to the existing items.

Some revisions and a few topics suggested by these instructors were used to write new items, and the total pool of items was then examined to eliminate any remaining problems. Then the questionnaire was administered to a group of university students studying English Language and Literature at BA level.

Triangulation provides researchers greater opportunities to gain better insights into what they are researching. As Brown (2001) suggests, combining quantitative survey questionnaires with qualitative techniques helps researchers to better understand the quantitative data. In line with these suggestions, the VKS questionnaire contains 44 items presented in three parts. The 34 items in Part 1 are of the Likert type, with a statement to which respondents would indicate one of the four responses. The responses have been worded according to the content of the item provided on the basis of their go-togetherness and the correspondence of the item and the response, but not a fixed response format, such as *Strongly disagree*, *Disagree*, *Agree*, or *Strongly agree*. A fixed format of choices sometimes does not illustrate the precision with which the respondents should respond

to a questionnaire. We had purposely and intentionally selected to offer four, rather than five choices in order to avoid the "middle-of-the-road syndrome" (Decapua & Wintergerst, 2005; Wintergerst, DeCapua, & Itzen, 2001) and to deter students from selecting the middle or non-committal response. Wintergerst et al. (2001) had noticed a tendency on the part of respondents to choose the middle choice. By offering only four, they hoped to have respondents evaluate more precisely the statements and their own feelings. Also we tried to avoid the use of too strong and too negative words, such as *never* or *always*. These terms make an absolute statement when circumstances themselves are not absolute. Therefore, we substituted *seldom* for *never* and *much of the time* for *always*.

The eight items in Part 2 ask the respondents to read and score their ability from 1 to 10. They correspond to some of the items existing among the items numbering from 1 to 34. Each item from numbers 35 to 42 restates one or two of the items in Part 1. This was intended to see if respondents were consistent in self-rating.

Finally, there are two open-ended items that "allow respondents to express their own thoughts and ideas freely, and thus may result in more unexpected and insightful data" (Mackey & Gass, 2005, p. 93). In writing items, we tried to be to the point and focused for the sake of brevity (Coronado-Aliegro, 2006; Harris, 1997).

3.3 Procedures
Before conducting the study, VLT and the VKS questionnaire were piloted on a group of EFL students at the state University of Qom, Iran. They were selected from among freshmen and seniors to see if the wording of the questionnaire was appropriate to all participants at different proficiency levels. This would also show the time needed to complete VLT and the VKS questionnaire as well as the difficulty that the students had in completing these measures. Due to time restrictions and the fact that this study was part of a larger project with four instruments, TOEFL reading and VLT were administered in one session, and IELTS reading and the VKS questionnaire in another to manage time and for the participants to cover all the items. Some participants were absent in session two and thus did not respond to the questionnaire. Some others told that

they could only attend one session, but not both. Therefore, they were only provided with VLT and the VKS questionnaire.

3.4 Research Design and Data Analysis

In addition to descriptive statistics, the statistical procedures applied to the data set include Cronbach Alpha, Spearman rho correlation coefficient, paired-samples *t*-test, Pearson product-moment correlation coefficient, cross-tabs procedure, and Somers' d.

4. Results

4.1 Internal Consistency

The internal consistency of all the 42 items in the questionnaire, excluding the open-ended items, is 0.912 with 295 respondents. Further, Cronbach Alpha was performed on Part 1 and 2 of the VKS questionnaire separately. Cronbach Alpha on Part 1 with 34 items is 0.885 whereas it is 0.870 for Part 2 with eight items. The high reliability coefficients of the whole questionnaire and those of Part 1 and 2 imply an acceptable and justifiable consistency of items.

4.2 Equivalent items in Part 1 and Part 2

Each item in Part 2 restates one or two of the items in Part 1. Table 3 only presents the number of items in Part 2 (row one) along with their corresponding items (row two). To compare the correlation between the two parts, Spearman rho correlation coefficient was conducted, showing a positive and moderate correlation coefficients ($\rho = 0.672$, $p = 0.000$) between the two parts. A further analysis showed that the correlation with respect to sex is not noticeably different ($\rho = 0.623$ for males and 0.691 for females, $p = 0.000$).

Table 3: Items in Part 2 and the Corresponding Items in Part 1

Items in Part 2	35	36	37	38	39	40	41	42
Corresponding items in Part 1	17	7	9	21, 22	20	13, 14	27	24, 25

To investigate the difference in self-assessment (in Part 2 and the corresponding items in Part 1) for mixed participants as well as that of male and female participants, a paired-samples *t*-test was performed. The total obtained scores in the two parts were converted to the scale of 100 to meet the requirement of paired-samples *t*-test. The paired-samples *t*-test for mixed participants

indicates that there is no significant difference ($t_{(294)}$ = −0.405, p = 0.686) between the performances of participants on the corresponding items in the two sections of the questionnaire. The small effect size of −0.095 implies that there is no difference between these two distributions of scores. Also, the paired-samples *t*-test of male ($t_{(124)}$ = −0.206, p = 0.837) and female ($t_{(169)}$ = −0.356, p = 0.722) performances on the two sections of the questionnaire shows no significant difference. The small effect size of −0.0078 for males and −0.011281 for females mean that both groups tend to be very similar and overlap almost entirely in their responses to Part 2 and the corresponding items in Part 1.

4.3 Open-ended Items in the VKS Questionnaire

The first open-ended item was concerned with how the participants guess the meaning of unknown words. 295 participants responded to the VKS questionnaire. However, several respondents did not answer the first open-ended item at all whereas some respondents only provided one answer to it, and still some others had come up with two or even more. Table 4 shows the categorized commonalities or recurrent themes in the responses, frequency of responses, etc. for the first item, as worded by the respondents, with few linguistic changes to the original wording to the extent possible.

From the 403 responses emerged some commonalities that were then reduced to the following five major categories upon close analysis, i.e. contextual clues, morphological attributes, semantic attributes, discoursal attributes, and syntactic attributes.

Table 4: The Categorized Commonalities or Recurrent Themes

Categories	Commonalities or recurrent themes	Number of responses	% out of responses	% out of 295 respondents
Contextual clues	Guessing from the context	135		
	Guessing from the surrounding words	52		
	Guessing from the sentences related to the word	4		
	Guessing the meaning from the text, the passage or the text style	37		
	Using the (contextual) clues or information available in the context	24		
	Guessing from the sentences coming before and after the difficult word	23		
	Total	275	68.24	93.22
Morphological attributes	Guessing from word formation	3		
	Guessing from rules of abbreviation	1		
	Guessing from the history of the words	1		
	Guessing the meaning from the root or origin	49		
	Guessing from etymology, derivation, suffixes and prefixes, and infix	30		
	Total	84	20.84	28.47
Semantic attributes	Guessing via analogy	1		
	Relying on collocations	1		
	Guessing from the cognates and sometimes from mother tongue	4		
	Guessing from similar words, parallel paraphrases, synonyms and antonyms available in the context	18		
	Total	24	5.96	8.14
Discoursal attributes	Guessing with regard to what one expects them (words) to say	1		
	Focusing on the punctuation and the words such as *like*, *such as*, etc.	1		
	Guessing the meaning from the general information and one's background knowledge	5		
	Considering the topic, theme, the purpose of the text or relating the words to the main idea	3		
	Total	10	2.48	3.39
Syntactic attributes	Guessing from its parts of speech	5		
	Detecting the position we use the word	5		
	Total	10	2.48	3.39
	Grand total	403	100%	100%

The second item required participants to indicate what other aspects they assumed to be involved in vocabulary knowledge, if any, and how much they were capable in them. 24 respondents out of all the participants in the study had answered it in all. Some participants had referred to the same points or content already raised, such as pronunciation, spelling, the context where the words are used, root, and using suffixes and prefixes. Therefore, such responses were excluded from analysis and only the points, more or less, different from those raised in the VKS questionnaire were kept. Table 5 shows learners' judgements on vocabulary knowledge.

Table 5: Vocabulary Knowledge Aspects Perceived by Learners

• Phrasal verbs	• Association of a word to what it signifies
• Special terms	• Association of a word with a picture to learn the meaning
• Visual imagery	
• Stress or intonation	• Distinguishing the formal and informal form of the words
• Transitivity or intransitivity	
• The stories behind the words	• The way of learning new words and increasing vocabulary
• Imaging a word in one's mind	
• Denotative and associative meanings	• Using delexical verbs with certain nouns (like make a decision)

Reported above as worded originally, each aspect was given by only one or in some cases two of all the 24 respondents, meaning that few respondents shared these responses.

4.4 Vocabulary Knowledge Self-assessment and Performance on VLT
RQ 1 and 2 asked whether there is any relationship between participants' self-assessment ratings of vocabulary knowledge and their performance on VLT and whether the relationship holds the same for males and females. RQ 3 asked if EFL learners' vocabulary knowledge self-assessment report could be a valid basis for grouping them into different levels of vocabulary knowledge.

The correlation analysis of the participants' scores in VLT and the VKS questionnaire for mixed group shows a positive and moderate relationship ($r = 0.504$) which is not strong enough for us to claim that we can predict the participants' level of vocabulary

knowledge only through their self-rating report without the use of a well-established standardized test.

The positive and moderate correlation between VLT and Part 1 $(r_{(293)} = 0.513, p < 0.000)$ and between VLT and Part 2 of the VKS questionnaire $(r_{(293)} = 0.422, p < 0.000)$ show a relative common general construct of these measures.

There exists a relatively similar correlation coefficients between the performances of male participants on VLT and the VKS questionnaire $(r = 0.499, p < 0.000)$ and that of female $(r = 0.472, p < 0.000)$. This indicates a common variance in the performances of male and female participants on these two instruments.

The correlation between the responses of male participants to VLT and Part 1 and between VLT and Part 2 of the VKS questionnaire indicates that $r_{(123)} = 0.457$, and $0.467, p < 0.000$ for males, respectively. However, for females $r_{(168)} = 0.507$ and 0.368, $p < 0.000$, respectively, for the relationship between VLT and Part 1 and that between VLT and Part 2 of the questionnaire. Almost all the relationships are moderate, except the weak relationship (0.368) between VLT and Part 2 of the VKS questionnaire for females.

Given the high correlation between Part 2 and the corresponding items in Part 1 of the VKS questionnaire on the one hand and the moderate correlation between VLT and the VKS questionnaire on the other, it appears that the participants overestimated or underestimated their vocabulary proficiency on the questionnaire. Otherwise, their performance on VLT, acting as evidence, should have corresponded more with their self-assessment results. To explore the relationship between VLT and the VKS questionnaire, cross-tabs procedure was performed. It is the level of one's vocabulary knowledge – the independent variable in the column – that affects how participants self-assess their vocabulary knowledge – the dependent variable in the row – but not the reverse.

Schmitt (personal communication, May 9, 2008) suggests the cutting point for the acquired level on VLT to be 24. It means that if participants answer 24 (80%) items correctly, they acquire the level. Otherwise, they do not reach the level. While the basis for this assertion is not clear from published sources, it remains the basis for

establishing vocabulary level in studies using VLT (see, for example, Xing & Fulcher, 2007). Apparently, a learner's percentage score on a level of VLT roughly indicates the number of words known at that level (Laufer & Nation, 1999; Read, 2000). On the other hand, some respondents answered, say, 22 items of one level, thus not meeting the point for acquiring that level, but some items of the upper levels. Taking these points into account, in crosstabulation, the sum of the correct items in all were included, irrespective of the cutting point for any level. The scores on VLT and the VKS questionnaire were converted to 100. Then the participants were divided into four levels for both VLT and the VKS questionnaire.

The columns show the levels based on VLT whereas the rows show them based on the VK questionnaire. Table 6 only includes level 2, 3, and 4 since no participant falls in level 1 based on either VLT or the VKS questionnaire. As column one shows, 60 participants fall in level 2 in accordance with their vocabulary knowledge. This does not hold true for row one, showing only 15 participants. There are 167 participants in level 3, column two, whereas there are 240 in row two. So only 167 learners as placed by VLT were shown to be in level 3. But the participants falling in level 3, row two, as placed by the VKS questionnaire outnumber those in column two as placed by VLT. There is a big difference. In level 4, column three, VLT places 68 participants whereas 40 participants fall in this level based on the VKS questionnaire.

Somers' d was then applied to the results of crosstabulation to measure the association between vocabulary knowledge as the independent variable and self-assessment of vocabulary knowledge as the dependent variable. The value for Somers' d is 0.345, indicating a weak positive association.

Table 6: Self-assessment versus VLT Cross tabulation

			Vocabulary Levels Test			Total
			2	3	4	
The VKS Questionnaire	2	Count	9	6	0	15
		% within VLT	15.0%	3.6%	.0%	5.1%
	3	Count	50	146	44	240
		% within VLT	83.3%	87.4%	64.7%	81.4%
	4	Count	1	15	24	40
		% within VLT	1.7%	9.0%	35.3%	13.6%
Total		Count	60	167	68	295
		% within VLT	100.0%	100.0%	100.0%	100.0%

Note: Columns are levels based on classified standard scores of VLT and rows are based on the VKS questionnaire in like manner.

5. Discussion and Conclusion

The present research pursued two specific objectives. Firstly, it was attempted to design and validate a global self-assessment questionnaire on vocabulary knowledge in an EFL situation. Secondly, it was tried to estimate its efficacy in placing the EFL learners into different levels of vocabulary knowledge.

The relatively strong correlation ($r = 0.67$) between Part 2 and the corresponding items in Part 1 indicates the one-dimensional nature of the VKS questionnaire. However, the square of this correlation shows that the proportion of shared variance is only about 0.452 per cent which is a rather moderate proportion of shared variance. Considering the point that the questionnaire is a first attempt of its kind, the proportion is promising. Future research might well increase the index.

Qualitative information generated by students on the self-assessment questionnaire provides additional support for the usefulness of self-assessment by FL instructors (Coronado-Aliegro, 2006). The open-ended items were thus raised for the participants to indicate some issues on vocabulary knowledge that might have fallen out of consideration in the study or the items in the

questionnaire might not have dealt with them properly. These items confirmed that the participants took the questionnaire seriously, thought about their own vocabulary proficiency, tried to reflect them as accurately as the questionnaire permitted, and in many cases suggested some issues not included.

The responses to the first item (Table 4) are quite relevant to the VKS questionnaire in the sense that the respondents almost unanimously point to the very concepts that all the items (from 1 to 42) refer to in some way. In fact, the responses are restatements of the content of our questionnaire items. Table 7 shows our items addressing the commonalities under each category in Table 4.

Table 7: Categories of the Participants' Answers Corresponding to the Questionnaire Items

Categories	Items in the VKS questionnaire addressing the commonalities under each category
Contextual clues	Item 9, 13, 14, 18, 21, 37, and 40
Morphological attributes	Items 3 and 4
Semantic attributes	Items 6, 7, 8, 16, 18, 19, 21, 22, 36, and 38
Discoursal attributes	----
Syntactic attributes	Items 5, 17, and 35

There are, however, some responses worth considering. Under *morphological attributes*, one is 'rules of abbreviation', following from knowledge of any specific field particularly. It might well be part of one's specialized knowledge. Very few common words such as *street*, *especially*, etc. are abbreviated in non-specialized ordinary language with which we are all familiar. Besides, such a specific item seems appropriate for a situation- or task-specific questionnaire and not a global one on vocabulary knowledge.

Under *semantic attributes*, one respondent mentions 'guessing from analogy' and does not explain any more. So one does not know what kind of analogy he or she has in mind. So this theme cannot be taken as a new feature about vocabulary knowledge.

The responses under *discoursal attributes* are not vocabulary knowledge per se. Rather, they are strategies in comprehending texts. On the other hand, there might be some overlap between vocabulary knowledge abilities and text comprehension strategies.

So this category might not be vocabulary-specific altogether. Consequently, the 403 responses provided raise no novel components and indirectly confirm the construct validity of our conceptual framework of vocabulary knowledge as shown above.

Some answers to the second open-ended item were near restatements or identical to the content of our items. Novel points in the responses (Table 6), however, are as follows: 'Phrasal verbs', a large portion of English language vocabulary, calls for a specific item in the VKS questionnaire. 'Special terms' are associated with one's specialized field of study and thus fall out of the scope of this questionnaire that deals with one's overall vocabulary knowledge. 'Intonation' is concerned with speech in context or is at the sentence level whereas 'stress' is at the word level. However, item 1 (When I see a word, I know how that word is pronounced) carries the issue of stress in that when one knows how a word is pronounced, they should know how it is stressed as well. 'Transitivity or intransitivity' is a feature of verbs, a part of the vocabulary. Item 35 generally tackles this issue. If we had included a specific item for verbs, we should likewise have done so for the other content words, hence lengthening the questionnaire beyond practicality.

It is worth noting that some points as raised by participants could best be described as vocabulary learning strategies, not vocabulary knowledge self-assessment. Learning strategies are specific procedures or techniques a person uses in solving a problem or in learning or teaching new words. (Wintergerst et al. 2001). Self-assessment, in contrast, is concerned with judgments about one's own ability (Brantmeier, 2006), here about one's own knowledge of vocabulary, but not about how one learns or teaches vocabulary. The former resembles a *process* whereas the latter is more like a *product* in the mind that is static. So, given this distinction, the responses in Table 6, such as 'visual imagery', 'learning the stories behind the words', 'imaging a word in one's mind', 'associating a word with what it signifies' and 'associating a word with a picture to learn the meaning', and 'the way of learning new words and increasing vocabulary' might well be regarded as strategies in vocabulary learning and teaching.

By 'denotative and associative meaning', the respondent most probably refers to denotative versus connotative meaning. "The *denotation* of a word is its primary significance or reference", such as the dictionary meanings; "its *connotation* is the range of secondary or associated significances and feelings which it commonly suggests or implies" (Abrams, 1993: 36). This aspect was not taken care of. 'Distinguishing the formal and informal form of the words' is parallel to item 10 (I can easily distinguish more common words from less common ones) though we have used the term common and less common for formal and informal. Also items 8 and 25 indirectly refer to the same thing since simplifying what you want to say is usually done with more frequent words than less frequent, less common or formal words. 'Using delexical verbs with certain nouns (like *make a decision*)' is the last aspect indicated. 'Make a decision' is a collocation because the noun 'decision' collocates with the verb 'make', but not with 'do' or any other verb. This theme is addressed by items 6, 7, and 36, as went above.

The responses to the second item indicate that a few more items need to be added to the questionnaire. It also appears that a specific item should be imbedded to ask about the issue of idiomaticity. As Anglin (1993) states, more than half of the compound entries are idioms, giving colour, feeling, charm, and precision to the language.

The correlation between VLT and the VKS questionnaire ($r = 0.504$) is moderate, implying that the VKS questionnaire by itself is not powerful enough to place participants in different groups of language proficiency. However, VLT and the VKS questionnaire results might provide complementary information, and thus we recommend both for grading purposes. Males and females do not perform any differently from one another in their self-assessment.

From a different angle, the crosstab procedure confirms our findings as well. Similar number of participants is expected to fall in each level through both VLT as the independent variable in the column variable and the VKS questionnaire as the dependent variable in the row variable. To clarify Table 6, we describe level 2 as an example. VLT places 60 participants in level 2 through crosstabulation. However, out of this number, 9 participants

(15.0%) fall in level 2, 50 (83.3%) in level 3, and 1 (1.7%) in level 4. So only 15.0% of the participants in level 2, as placed by VLT, self-assess their vocabulary knowledge correctly. 85.0% of the participants in level 2, as placed by VLT, overestimate their vocabulary knowledge. This holds true for other levels as well, though in a different direction.

It is worth noting that the higher the participants' vocabulary knowledge, the lower their self-assessment of their vocabulary knowledge. As Table 6 shows, 85.0% of the participants in level 2 overestimate their vocabulary knowledge. A different direction occurs for level 4 as placed by VLT. 68 participants fall in level 4 of VLT in crosstabulation. However, only vocabulary knowledge self-assessment of only 35.3% of the respondents corresponds with their VLT level. It means 64.7% of the participants in level 3, as placed by VLT, underestimate their vocabulary knowledge to be in level 3. Interestingly enough, 146 participants (87.4%) of 167 falling in level 3 of VLT in Table 6 correctly self-assess their vocabulary knowledge, 15 (9.0%) overestimate, and 6 (3.6%) underestimate their vocabulary proficiency. It indicates that the learners who almost know between 3000 and 4000 word families – roughly lying in level 3 of VLT – will be more realistic in self-assessment.

Overall, the application of Somers' d to crosstabulation results reveals a weak association between vocabulary knowledge and vocabulary proficiency self-assessment. If we square the result of Somers' d ($d = 0.354$), the shared variance is not so large, $d^2 = 0.119$. So there are other factors at work since there is 0.881 per cent unexplained variance in self-assessment.

The results, so far, caution us against using EFL learners' vocabulary knowledge self-assessment report alone as a valid basis for grouping them into different levels of vocabulary knowledge, hence the answers to the research questions. However, instructors can use students' self-guided appraisals as information to enhance future teaching methodology. One way to bring about the self-evaluation attitude in FL learners is to administer self-assessment questionnaire along with a valid and reliable instrument. Studies recommends that when more concrete and descriptive scales are

provided, students are able to assess themselves more accurately (Jansen-van Dieten, 1989; LeBlanc & Painchaud, 1985).

It appears that, in Iranian culture, the teacher is perceived as the one responsible for preparing, administering, and grading the assessments. Students appear to be the passive recipients of knowledge and do not participate in the evaluation process. Although classes are supposedly student-centred, teachers control evaluation, maybe because they do not trust the accuracy of students' self-assessments.

The value in establishing score comparisons between VLT and self-assessment questionnaire resides in its potential to bridge the gap between differences in judgments and, therefore, assist learners in identifying their own abilities to self-assess objectively. When students recognize their ability of accurate self-assessment, they'll be more confident and motivated to participate in self-assessment activities. Munoz and Alvarez (2007, p. 16) state that it "may be reasonable to think that grammar and vocabulary are more easily assessed by students because these aspects are more tangible…."

The results suggest that quantitative research or statistical findings alone are insufficient to ascertain the effectiveness and usefulness of a questionnaire. As the data in response to the two open-ended items reveal, several components were not addressed in the construct validity of the questionnaire. These issues need to be further investigated in future research. A factor that might influence the validity and reliability of questionnaires is that respondents might be often uninterested or bored in completing such a questionnaire. If respondents check answers merely to complete an instrument, they are not reflecting upon the questions or indicating their true preferences (Porte, 2002). So, combining quantitative survey questionnaires with qualitative techniques helps researchers to better understand the quantitative data (Brown, 2001).

In addition, triangulation allows researchers greater opportunities to gain better insights into what they are researching. Utilizing a self-assessment questionnaire, including a) an enough number of can-do statements, b) a second part that is the restatement of some of the original items in the first part, and c) some open-ended items,

along with an evidence-based vocabulary test, presents a fuller picture with regard to the comprehensiveness and precision of a vocabulary knowledge self-assessment questionnaire.

Though self-assessment is not a panacea for all testing problems and the field is fraught with problematic issues, yet future researchers can systematically illuminate the efficacy of self-assessment tools in testing and evaluation in EFL situations. Apart from what was mentioned above, it appears to us that the following factors affected our findings: First, self-assessment seems to be scarcely carried out in English educational system. Even graduate students rarely have the experience of assessing themselves. So we go along with Ross (1998) that "the experience factor is potentially an important source of variation in self-assessment" (p. 13). In other words, as Ross (1998) states, "the degree of experience learners bring to the self-assessment context influences the accuracy of the product" (p. 16). Second, VLT was not administered to the participants prior to administering the VKS questionnaire. Had it been tested before the questionnaire, VLT would have been effective in offering an objective criterion against which the participants could have weighed self-evaluation of their vocabulary knowledge, and have been useful in indirectly discouraging the participants from overestimating or underestimating themselves. Some of the participants observed that if VLT had been given first, they would have responded differently. In addition, the strong consistency of the scores in Part 2 and the corresponding items in Part 1 is evidence of their being consistent in a second evaluation of their own vocabulary proficiency. What is lacking seems to be a criterion and some practice. These learners, in particular, need practice to be more realistic in their self-assessment.

The study supports Laufer and Nation (1999) that it is important in the design of the vocabulary component of a teaching curriculum that instructors be able to determine the state of their learners' vocabulary knowledge and draw on a variety of vocabulary measures to investigate the nature of vocabulary growth. Self-assessment can be an alternative. Of course, it is a lengthy and complex process to adopt an innovation within an educational

system; individuals must first be presented with an innovation, then be given sufficient time and opportunity to understand, become convenient with, and finally internalize the proposed change (Sakui & Gaies, 1999). No matter how it is performed, seemingly the value of exploring learners' vocabulary knowledge self-assessment is uncontroversial for theoretical and practical reasons.

References

Abrams, M. H. (1993). *A glossary of literary terms* (6th Ed.). Fort Worth: Harcourt Brace College Publishers.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58* (10).

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Barrow, J., Nakanishi, Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, *27*, 223-47.

Bayliss, D. (1991, March). *The role and limitations of self-assessment in testing and research.* Paper presented at the Language Testing Research Colloquium, Princeton, NJ.

Beglar, D., & Hunt, A. (1999). Revising and Validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing*, *16*, 131-162.

Blanche, P., & Merino, B. (1989). Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning, 39*, 313-40.

Bogaards, P. (2000). Testing L2 vocabulary at a high level: The case of the Euralex French tests. *Applied Linguistics, 21*, 490-516.

Brantmeier, C. (2005). Non-Linguistic variables in advanced L2 reading: learner's self-assessment and enjoyment. *Foreign Language Annals*, *38*, 493–503.

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System, 34*, 15-35.

Brown, A. (2005). Self-assessment of writing in independent language learning programs: The value of annotated samples. *Assessing Writing, 10,* 174-191.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Clark, J. L. D. (1981). Language. In T. S. Barrows et al. (Eds.), *College students' knowledge and beliefs: A survey of global understanding* (pp. 25-35). New Rochelle, NY: Change Magazine Press.

Coombe, C. (2002). Self-assessment in language testing: Reliability and validity issues. Retrieved September 30, 2005, from http://www3.telus.net/linguisticsissues/selfassess. html.

Coombe, C., & Canning, C. (2002). Using self-assessment in the classroom: Rationale and suggested techniques. Retrieved September 30, 2005, from http://www3.telus.net/linguis ticsissues/selfassess2.html.

Coronado-Aliegro, J. (2006). The effect of self-assessment on the self-efficacy of students studying Spanish as a foreign language. Unpublished PhD dissertation. University of Pittsburgh.

Coxhead, A. (2007). Five practical tips to improve writing skills using the Academic Word List. *Oxford students' dictionary of English* (R10-R14). Oxford: Oxford University Press.

DeCapua, A., & Wintergerst, A. C. (2005). Assessing and validating a learning styles instrument. *System, 33*, 1-16.

Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, *19*, 273–299.

Greidanus, T., Beks, B., & Wakely, R. (2005). Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *The Modern Language Journal, 89*, 221-233.

Hargan, N. (1994). Learner autonomy by remote control. *System*, *22*, 455–462.

Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal, 51*, 12-20.

Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing, 7,* 174-201.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303-317.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1), 23-59.

Janssen-van Dieten, A. (1989). The development of a test of Dutch as a foreign language: The validity of self-assessment by inexperienced subjects. *Language Testing, 6*, 30-46.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*, 47-77.

Krausert, S. R. (1991). Determining the usefulness of self-assessment of foreign language skills: post-secondary ESL students' placement contribution. Unpublished PhD Dissertation. University of Southern California.

Laufer, B., & Nation, P. (1999). A productive-size test of controlled productive ability. *Language Testing, 16*, 33-51.

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*, 673-687.

Little, D. (2005). The common European framework and the European language portfolio: Involving learners and their judgments in the assessment process. *Language Testing, 22*, 321-336.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Meara, P. (1992). EFL vocabulary tests. *Tests/Evaluation Instruments*, Swansea: Wales University.

Meara, P. (1996). The vocabulary knowledge framework. Retrieved October 10, 2006, from www.swan.ac.uk/cals/vlibrary/pm96d.

Munoz, A., & Alvarez, M. E. (2007). Students' objectivity and perception of self assessment in an EFL classroom. *The Journal of Asia TEFL, 4*(2), 1 – 25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Oscarson, M. (1978). *Approaches to Self-assessment in Foreign Language Learning*. Council of Europe, Council for Cultural Cooperation, Strasbourg.

Oscarson, M. (1980). *Approaches to Self-Assessment in Foreign Language Learning*. Oxford: Pergamon Press.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *The encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 175-187). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Oskarsson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing, 6*, 2-13.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*, 282-296.

Porte, K. (2002). *Appraising Research in Second Language Learning*. Amsterdam, PA: John Benjamins.

Read, J. (1988). Measruing the vocabulary knowledge of second language learners. *RELC Journal, 19*(2), 12-25.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Roever, C., & Powers, D. E. (2005). *Effects of language of administration on a self-assessment of language skills*. (TOEFL Monograph series Report No. 27). Princeton, NJ: Educational Testing Service.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15*, 1-20.

Sakui, K., & Gaies, S. J. (1999). Investigating Japanese learners' beliefs about language learning. *System, 27*, 473-492.

Schmitt, N. (1994, March). Vocabulary Testing: Questions for test development with six examples of tests of vocabulary size and depth. *Thai TESOL Bulletin*, *6,* 9-16.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–88.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: SAGE.

Tannenbaum, R. J., Rosenfeld, M., Breyer, F. J., & Wilson K. (2000). *Linking TOEIC scores to self-assessments of English-language abilities: A study of score interpretation.* Unpublished manuscript.

Todd, R. W. (2002). Using self-assessment for evaluation. *English Teaching Forum, 40*(1), 16-19.

Wesche, M., Paribakht, T. S., & Ready, D. (1993). *A comparative study of four placement instruments.* Paper presented at the Paper presented at the Annual Language Testing Research Colloquium (15th, Cambridge, England, and Arnhem, The Netherlands, August 2-8, 1993).

Wilson, K. M. (1999). *Validity of a global self-rating of ESL speaking proficiency based on an FSI/ILR-referenced scale* (ETS RR-99-13). Princeton, NJ: ETS.

Wintergerst, A., DeCapua, A., & Itzen, R. (2001). The construct validity of one learning styles instrument. *System, 29*, 385-403.

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, *35*, 182-191.