A Rationale for Creating Items for the Vocabulary Section of the ELI Reading Placement Test
Nhi Chau Thi Nguyen and Alexis von Biedenfeld
December 11, 2005
A Rationale for Creating Items 2

Objectives:

Our task is to provide a vocabulary test of 20-25 items to be included in the vocabulary item bank for the ELI. This vocabulary item bank will serve as a source for the ELI to select vocabulary test items when they need to administer a reading placement test to the students of the University of Hawai’i at Manoa.

An overview of ELI’s Reading Placement Test and its vocabulary component:

The ELI Reading Placement Test (RT) is one of the four placement tests that the ELI provides to incoming international learners who score between 500 and 600 on the paper-based TOEFL. The purpose of the test is to measure learners’ relative proficiency so that they can be placed into either Level 1 or Level 2 of courses that develop learners’ proficiency in the areas of reading, speaking, listening and writing for academic purposes. For example, the result of the RT will help the ELI staff decide which course, Level 1 Reading or Level 2 Reading, a particular person who takes the test should be placed in.

ELI’s RT has two parts: one is to test reading comprehension and the other is to test learners’ vocabulary knowledge. Each section is made up of 25 multiple-choice items, and each item has four options: one key and three distractors.

For the vocabulary items, there are three basic formats:

1. *Match one word with other single words.* In this format, the stem consists of only one word, the one that is tested. The correct answer is also one word – the one that is synonymous with or close to the meaning of the stem. The distractors also consist of only
one word, which have similar meanings to the stem or the other distractors. The following example is taken from the ELI Reading Placement Test, Form B:

1. vendor
   A. adviser
   B. director
   C. maker
   D. seller

2. **Match a single word to two- or three-word phrases.** In this format, the stem also consists of one word, the one that is tested, but the key and the distractors can consist of two or three words whose meanings are closely related. The following example is taken from ELI Reading Placement Test, Form B.

2. exemplify
   A. to make louder [amplify]
   B. to represent
   C. to be free from [exempt from]
   D. to feel for [empathize]

3. **Match a two- or three-word definition to a single word.** In this format, the stem is a two or three-word definition and the key and the distractors consist of only one word. See the example below (ELI Reading Placement Test, Form B):

3. related to the brain
   A. casual
   B. central
   C. cerebral
   D. ceremonial

**Background Information (Devising a Strategy):**

Before we started developing items for the ELI Vocabulary Test, our first task was to determine what kind of items we wanted to write. Although this test features multiple-choice, context-free items, we wanted to confirm that this was the most efficient means to measure a
students’ knowledge of vocabulary. To this end, we focused on the environment and format of the test items.

First, we analyzed the effectiveness of placing the Vocabulary Test in the Reading section of the ELI Placement Test. According to Read (2000), there is a “strong, well-documented association between good vocabulary knowledge and the ability to read well (p.74).” The testing GA’s were able to confirm that there is a strong correlation between students’ test scores in the Reading Comprehension and Vocabulary section on the Placement Test. Therefore, including a Vocabulary Test in the Reading section is beneficial for both testing reading ability and reinforcing the validity of either test (through a comparison of scores).

Also related to how the items are situated is the question of whether or not the items should be context-dependent. At first, we speculated that vocabulary items should be included with each reading passage. Because some words have more than one meaning, and because the students are not likely to encounter difficult words out of context, we postulated that it would be more realistic to test the students on how well they knew the meanings of words in context. However, we eventually realized that we should not use context to test these vocabulary items.

First, this would increase the amount of time the test-takers needed to answer the question because they would have to re-read the passage in order to answer the question. The more time it takes a student to read the item, the fewer the items that can be tested (Read, p. 147). Additionally, in context-dependent vocabulary tests, the extent to which only the target vocabulary is being tested is not clear (Read, p. 119). Therefore, context might confuse even students who understand the meaning of the target word. There seems to be a need for more research on the role of context in vocabulary assessment (Read, p. 101). For the task at hand, we
decided to maintain the context-free environment for the test items and keep them in the Reading Section of the test.

We were also interested in determining the best format for the items. Using multiple-choice items has several potential disadvantages. In the first place, these kinds of items take a long time to create. In order to write a multiple-choice item, we must first take care to choose distractors that will not give the test taker a clue to the correct meaning of the word. Then, we must pilot the items to make sure that they discriminate well between low and high-proficiency students and that they are not biased against or toward any particular group of students. This is indeed a lengthy process.

Additionally, multiple-choice questions may not reflect a student’s knowledge of a word. Multiple-choice questions typically test only one meaning of a word, and it may not be the meaning with which the test taker is familiar. Also, distractors may confuse the test takers or give them clues to the meaning of the word. Finally, these types of questions do not reflect what Canale and Swain call strategic competence, or the ability to negotiate for meaning.

In the end, though, multiple-choice items have a number of advantages. Although they take longer to develop, they can be scored more quickly. This is especially important for the teachers at the ELI because they want to put the students into groups as quickly as possible before the semester begins. Also, even if multiple-choice items may not perfectly discriminate between low and high-proficiency students, the correlation between scores on the Vocabulary and Reading Comprehension tests indicates that both tests are consistently measuring the same construct. After our research, we were convinced that the environment and format in which items
A Rationale for Creating Items

on the Vocabulary Test are presented are conducive to achieving the goal of measuring
vocabulary knowledge among ELI students.

**Strategies for developing test items:**

One thing to notice here is that although the ELI claims to use distractors that are related
in meaning, as we can see from the examples above, the meaning of the distractors are not
always closely related. In example 1, what is similar in the four choices are the –er or –or
suffixes denoting the agent. Or in example 3, the four choices are similar in the –al ending while
the meanings cannot be said to be similar. Example 2 is a little complicated, but if you try to find
the one-word synonym to the definition given in the choices, you can find that in fact, these
options all have synonyms that sound like the word found in the stem. Therefore, the distractors
are similar in form to the stem or the correct answer. This is acceptable because if the learners
just memorize the word lists, they might end up confusing the meaning of the words which are
somewhat similar in form (i.e. having the same word ending, or sounding similar to each other).
This test may separate learners who actually know the meaning of the word from those who are
only familiar with the word form. So this is a good strategy to consider when developing
vocabulary items for a vocabulary test.

Another issue that we learned about from the testing GAs of the ELI is that the
vocabulary items that are tested are selected from the academic word list (AWL) developed by
Victoria University of Wellington ([http://www.vuw.ac.nz/lals/research/awl/index.html](http://www.vuw.ac.nz/lals/research/awl/index.html)). One of
the challenges in writing VT items is deciding which words to test. In the area of vocabulary
testing, it is a well-documented procedure to sample from a list of frequently-occurring words
(Read, 2000). This AWL is particularly suited to the task of choosing vocabulary items for the
ELI Vocabulary Test (VT) because it 1.) lists words from academic tests and 2.) lists words
A Rationale for Creating Items

according to frequency and word families. These two characteristics are necessary for both selecting test items and proving the validity of this test.

The purpose of the VT is to spread students out according to level of proficiency. This is done in order to place them in classes that help develop their academic vocabulary so that the learners can be successful in reading academic texts at UHM. So in a way, selecting the tested words from the AWL is a good idea because in the end, what we need to know is if students have enough academic vocabulary to deal with academic reading texts. Therefore, choosing words for the VT from a list of academic words is one way to prove content validity (c.f. Read, p. 95).

In addition, the AWL consists of 10 sublists arranged in descending order of frequency. John Read’s book Assessing Vocabulary features a Perkins and Linnville (1987) study on the features of the target word, not the distractors, that influence the difficulty of multiple choice vocabulary items, and the conclusion is that one of the best predictors of the learners’ performance on the test is the naturally-occurring frequency of the word that is tested (p.79). So the AWL serves as a tool that can facilitate test-developers’ choice of easy and difficult items using the sublists (the first sublist consists of words of the highest frequency and the last sublist consists of word of the lowest frequency out of the words identified as frequently occurring in academic texts in English). The strategy of choosing vocabulary test items from an established AWL is sound, for both accurately determining a students’ level of proficiency and establishing test validity.

VT item development procedure (words to be tested):

Because our aim is to develop 25 items for the ELI item bank, our strategy is to follow their format and use the same source of vocabulary.
First, we decided to select vocabulary randomly from the AWL. The AWL has 10 sublists altogether. Sublists 1-9 consist of 60 word families each, and Sublist 10 consists of 10 word families. The sublists are arranged in descending order of frequency: Sublist 1 has the words of the highest frequency, while Sublist 10 has the words of lowest frequency. Because we have to come up with 25 items, what we have decided to do is to select three word families from each list in Sublists 1-7 and two word families from each list in Sublists 8-9. So for Sublists 1-7, we will select the word whose ordinal number is the multiplier of 20 in the sublist, meaning word families numbered 20, 40 and 60 in the sublist. For Sublists 8-9, we have selected the word families for number 30 and 60. We have not chosen any words from Sublist 10 because it has only 10 word families, which does not suit our system of choosing word family numbered as a multiplier of 20 or 30.

VT item development procedure (item format):

After we decided on which word family to include in the test, we decided on which category of the word family should be considered in the test: noun, adjective, verb, adverb or conjunction. In terms of test format, we used the same test formats that the ELI used (c.f. discussion of ELI test format in the second section). We established the following table to use as a guideline to make sure that the categories of the word families and the item formats are equally treated in the test (see Table 1). In this table, in the column designated as test format, word-word means that the relative test item will have a one-word stem and four one-word options. Word-phrase means that the item has a one-word stem, but the options are a phrase; and phrase-word means the stem is a phrase giving the definition of something, and the four options are one-word.

In fact, we tried to distribute the categories and test formats evenly across the word families that we have decided to test; however, we could not achieve such a goal because of the
nature of the words on the AWL. For example, for words like *culture*, if we distribute the word categories evenly, we might try to test the adverb form of this word (i.e. *culturally*). Nevertheless, we found that for the word family *culture*, the noun category is the one that occurs most frequently and it is also easier to design a test on the noun than on the adverb; so we have decided to test the noun form of this word family rather than the adverb. With this kind of reasoning, you can see that the categories of the word families that are tested are somewhat subjective, but we still reach an almost equal number of nouns, verbs, adjectives and adverbs tested on the test.

**Table 1. A summary on the VT item development procedure:**

<table>
<thead>
<tr>
<th>Sublist</th>
<th>Order in sublist</th>
<th>Order of consideration</th>
<th>Selected word family</th>
<th>Selected category</th>
<th>Test format</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>1</td>
<td>environment</td>
<td>N</td>
<td>word – word</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>2</td>
<td>method</td>
<td>Adj</td>
<td>word – phrase</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>3</td>
<td>vary</td>
<td>V</td>
<td>phrase – word</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>4</td>
<td>culture</td>
<td>N</td>
<td>word – phrase</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>5</td>
<td>positive</td>
<td>Adv</td>
<td>word – word</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>6</td>
<td>transfer</td>
<td>V</td>
<td>phrase – word</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>7</td>
<td>dominate</td>
<td>V</td>
<td>word – word</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>8</td>
<td>partner</td>
<td>N</td>
<td>phrase – word</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>9</td>
<td>volume</td>
<td>Adj</td>
<td>word – phrase</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>10</td>
<td>emerged</td>
<td>Adj</td>
<td>phrase-word</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>11</td>
<td>overall</td>
<td>Adv</td>
<td>word – phrase</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>12</td>
<td>undertake</td>
<td>V</td>
<td>word – word</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>13</td>
<td>equivalent</td>
<td>N</td>
<td>word – word</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>14</td>
<td>objectively</td>
<td>Adv</td>
<td>word-phrase</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>15</td>
<td>whereas</td>
<td>Conj</td>
<td>word-phrase</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>16</td>
<td>estate</td>
<td>N</td>
<td>phrase-word</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>17</td>
<td>lecture</td>
<td>N</td>
<td>word-word</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>18</td>
<td>utilize</td>
<td>V</td>
<td>word-phrase</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>19</td>
<td>eliminate</td>
<td>V</td>
<td>word-word</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>20</td>
<td>paradigm</td>
<td>N</td>
<td>phrase-word</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>21</td>
<td>voluntarily</td>
<td>Adv</td>
<td>word-phrase</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>22</td>
<td>implicitly</td>
<td>Adj</td>
<td>phrase-word</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>23</td>
<td>widespread</td>
<td>Adj</td>
<td>word-word</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>24</td>
<td>manually</td>
<td>Adv</td>
<td>word-phrase</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>25</td>
<td>vision</td>
<td>N</td>
<td>phrase-word</td>
</tr>
</tbody>
</table>
The same is true for the test format. Some words are easier with the word-word format than the word-phrase or phrase-word format, so although we have tried to use the different item formats in an alternating pattern, we still used some personal judgment of when to use one format rather than the other. However, the final result is also to ensure that the test formats are used equally across the whole test, meaning 8-9 items with word-word format, 8 items with word-phrase format and 8 items with phrase-word format.

**VT item development procedure (distractors):**

To check the meaning of each word, we used the Microsoft Thesaurus Dictionary of American English (MTDAE) and the Microsoft Encarta Dictionary of North American English (MEDNAE). The advantage of these two dictionaries is not just that they give the definitions of words in American English, but also that the MIDAE gives the one-word equivalent form (which is very suitable for our word-word format), and the MEDNAE gives the definition of each word in phrases of more than two words, which is suitable for our word-phrase and phrase-word formats.

For the word-word format, we checked the meaning of the word on the MTDAE to come up with a list of synonyms. As we know that there are no perfect synonyms in English or any other language, we checked the meaning of each synonym to find if these synonyms have any other meaning other than the meaning stated by the stem of the test (the word that we want to test). In this way, we can easily come up with distractors that are close in meaning as the ELI wishes to have. In some items, we also tried to mimic the existing ELI VT by choosing distractors that are similar in form to the correct response (e.g. Appendix A, Item 19). This prevents test-takers from choosing an answer based on the fact that it stands out from the other distractors in some way other than being the correct answer.
For the word-phrase format, we began the same way as we did with the word-word format, but when we got the list of synonyms, we checked if the synonyms had any other meaning than the meaning of the stem. If one does, we chose the word that had a different meaning from the meaning of the stem, then looked it up in the MEDNAE for a phrase that had the same meaning as the one whose meaning is different from the stem, and used that phrase as a distractor. The correct answer is the phrase whose meaning is the same as the meaning of the word in the stem. As in the word-word items, something we kept in mind while creating distractors in the word-phrase items was that the distractors had to be similar in order to prevent the student from choosing the correct answer based on test-taking strategies, rather than actually knowing the correct answer. In order to achieve this, we tried to use distractors that were of similar length and form. For example, we tried to be consistent by starting each distractor with the same part of speech (e.g. Appendix A, Item 11).

For the phrase-word format, we started by checking in the MEDNAE for a phrase synonymous with the word family and category we would like to test. Then, we put that phrase in the stem and the word in question as the correct answer. For the distractors, we found the synonyms of the word in question on the MTDAE and then checked again for meanings other than the correct meaning to use as distractors. In order to make the distractors efficient, we employed the same methods as in the word-word items and word-phrase items.

After we found the words to test and four options for each item, we had to decide how to arrange the distractors. For some of the items, we decided to arrange the options in alphabetical order of the first letter of each item. For example, Item 1 (Example 1, below) tests the students on their knowledge of the word *environment*. The four options that we developed are *character*, *condition*, *position* and *surroundings*. As *character* starts with letters *ch*, it is placed before
A Rationale for Creating Items 12

*condition*, which starts with *co* and before *position* and *surrounding* which respectively start with *p* and *s*. The purpose of this arrangement is to ensure the random arrangement of the test options and therefore to reduce the probability of the test-takers getting the correct answers by chance.

**Example 1**

1. environment
   A. character
   B. condition
   C. position
   D. surroundings

   Another method we used to order the distractors was to list them according to length. This was especially useful for such items as Number 14, in which one distractor is considerably longer than the others. This strategy does not encourage the test taker to consider length as a means of guessing the answer; rather, he or she sees that the distractors have been listed in order of length and that their order will not affect which one is the correct answer. This strategy is also employed on the current ELI Vocabulary Test.

   One final way we decided the order of the distractors was to use a random number generator. There are a number of ways to generate random numbers; one of the easiest ways is to access an on-line random number generator, such as the one found at www.random.org. All we had to do was plug in the numbers 1, 2, 3 and 4 and the website generated a list of random numbers. Then, we assigned letter values to these numbers (i.e. A=1, B=2, etc.), and generated an answer key, based on random assignment. This is the primary method used in Items 15-25 (see Appendix A). Any of these three methods is useful for making sure that the order of the distractors does not indicate to the test taker which answer is correct.

   Finally, we had to make sure that the test takers were able to understand the meaning of the distractors. If the distractors were too difficult, we would be testing the students’
understanding of more than just the target vocabulary, which is beyond the scope of this vocabulary test. One way of ensuring that the distractors can be easily understood is to utilize a general word-frequency list, like the one found at www.edict.com/hk. This website includes the Brown Corpus of the first 2000 and next 3000 most frequently-occurring words. Because university-level words occur less frequently than these 5000 most–frequent words (cited in Read, p. 118-9), the 5000 most frequent words should be easily understood by students who are entering the ELI. Therefore, in developing distractors, we tried to use words that were either on the 5000 most-frequent word list or in one of the earlier AWL sublists.

**Discussion and Conclusion:**

In developing test items for the ELI Vocabulary Test, we had to consider many factors. First, we had to create items that would allow the teachers at the ELI to have enough data to place the students in classes according to proficiency. To this end, we used multiple-choice, context-free items that would minimize the amount of time it took students to finish the test, and therefore maximize the amount of vocabulary that could be tested. We also chose vocabulary words according to their level of frequency (and thereby difficulty) so that it would be easy to determine students’ range of vocabulary. In this way, we intended to spread students along a normal curve, which would then determine their level of placement.

We also had to consider practical issues. For example, although writing a multiple-choice/ matching the meaning item to test students’ ability to use conjunctions was difficult (because conjunctions tend to indicate a relationship more than a clear meaning), we included the word “whereas” in a multiple-choice item (Appendix A, Item 15) because other formats (e.g. cloze, context-dependent, etc.) would require students to take more time to read the question. Additionally, multiple-choice was the best format to use to ensure that ELI teachers would have
enough time to objectively score the ELI Placement Test before the start of the semester.

Although multiple-choice tests take time to design and pilot, they facilitate the scoring procedure.

Finally, we had to determine a method for creating and ordering distractors. This was challenging because we did not want to make the answer obvious, nor did we want to make the distractors too difficult for the students to understand. Also, we wanted to make sure that students who knew the target vocabulary would be able to score higher than students who had simply memorized words (not meanings) or students who guessed. When the items we designed are piloted, we will discover whether or not we were successful in creating item discrimination.

In conclusion, developing test items was a challenging task. Perhaps one of the most important things that we discovered about writing vocabulary test items was that more research needs to be done on this topic. That being said, we think that the items on the ELI Vocabulary Test are valid for the purpose of this test. We aimed to create efficient, unbiased items, based on the ELI’s existing VT, but whether or not we achieved our goal remains to be tested.
Appendix A: The test

VOCABULARY

INSTRUCTIONS: The problems in this section are of different types, as illustrated by the examples below:

EXAMPLE I:

feverish
A. evil
B. hot
C. hungry
D. poor

EXAMPLE II:

related to work
A. occupational
B. recreational
C. rotational
D. sensational

In Example I you find a test word, beneath which are four words marked A, B, C or D. You should choose the one word of the four whose meaning is nearest to the meaning of the test word. In this example the word nearest to the meaning of the word “feverish” is the word “hot”. If this were a part of the test, you would blacken the circle for the letter B on your answer sheet.

Example II gives you a test definition which defines one of the four words below it. The definition “related to work” tells you the meaning of the word “occupational,” which is choice A. Your answer sheet would be marked with a blackened circle for the letter A.

Please remember to make marks only on the answer sheet, NOT on the test booklet. If you understand the directions continue the test and answer all of the items numbered 26 through 50. If you have any questions, raise your hand and one of the teachers will be happy to help you.

Please turn the page and work on the vocabulary test.
1. environment
   A. character
   B. condition
   C. position
   D. surroundings

2. methodical
   A. being economical
   B. happening frequently
   C. well-behaved
   D. well-organized

3. undergo change
   A. contrast
   B. disagree
   C. modify
   D. vary

4. culture
   A. advance in values
   B. learning experience
   C. shared beliefs and values
   D. uncontrollable conditions

5. positively
   A. absolutely
   B. actually
   C. emphatically
   D. sincerely

6. give ownership of
   A. distribute
   B. move
   C. replace
   D. transfer

7. voluminous
   A. generous
   B. liberal
   C. loose
   D. sufficient

8. business associate
   A. equal
   B. fellow
   C. partner
   D. superior

9. dominate
   A. carry out
   B. fail to notice
   C. have power over
   D. let pass

10. appearing for the first time
    A. active
    B. emergent
    C. evolving
    D. rising

11. overall
    A. by and large
    B. in the usual way
    C. on a regular basis
    D. on many occasions

12. undertake
    A. employ
    B. endure
    C. perform
    D. suppose

13. equivalent
    A. complement
    B. counterpart
    C. opponent
    D. similarity

14. objectively
    A. with criticism
    B. with a purpose
    C. without giving a reason
    D. without personal influence
15. whereas
   A. for the reason that
   B. at the same place
   C. while in contrast
   D. in addition to

16. all of a person’s possessions
   A. loan
   B. fund
   C. estate
   D. principal

17. a speech
   A. a lecture
   B. a letter
   C. an election
   D. an element

18. utilize
   A. to speak
   B. to wait for
   C. to make use of
   D. to understand as

19. eliminate
   A. return
   B. relapse
   C. retain
   D. remove

20. point of view
   A. summary
   B. paradigm

21. voluntarily
   A. with effort
   B. with speed
   C. by force
   D. by choice

22. not directly stated
   A. implicit
   B. explicit
   C. imposed
   D. exposed

23. widespread
   A. long
   B. heavy
   C. educated
   D. common

24. manually
   A. by hand
   B. on foot
   C. with ease
   D. for fun

25. the ability to see
   A. vigor
   B. vision
   C. variance
   D. vitality
A Rationale for Creating Items 18

Answer key:

1. D
2. D
3. D
4. C
5. A
6. D
7. C
8. C
9. C
10. B
11. A
12. C
13. B
14. D
15. C
16. C
17. A
18. C
19. D
20. B
21. D
22. A
23. D
24. A
25. B

Reference