

ESL Test Development Project

Matthew Barbee
Kaitlyn McClaren
Lisa Park

University of Hawai'i at Manoa
Department of Second Language Studies
SLS 490: Dr. Thom Hudson

Test Development Project

Introduction

The purpose of this project was to create a language test from the ground-up, grounded in ESL pedagogical theory and based on a defensible construct. Our project began with a brief needs analysis of the population to be tested in the context of the language course that that population is taking, which led to the formation of a construct for testing, the writing of test item specifications, and ended with the piloting of our test on an equivalent population. Briefly, the context of our test is an end-of-unit, reading comprehension and vocabulary achievement test for ESL students in the US who want to enter an American university. It is a criterion-referenced test that focuses on two constructs, reading comprehension skills and vocabulary retention. The purpose of our test is to evaluate students on their mastery of new vocabulary from a language unit that they have recently studied as well as their comprehension of a reading passage with the same theme, length, and difficulty as other passages from that same unit. Our test includes a reading passage where the students have to demonstrate reading comprehension and two vocabulary sections, one matching and one productive, where students have to demonstrate their knowledge of the meanings of words presented and taught in the unit.

From the results of the unit test, test administrators will be able to make diagnostic decisions regards the students' mastery of learner objectives. If the test shows that the students have not mastered the tested objectives, those students will be given further supplemental instruction and ultimately re-tested. Overall, while the purpose of our test is to to assess what the students have learned from one unit, it, along with other unit tests within the context of the entire course, will not only provide the teacher a basis for making decisions about the students' learning,

but also give feedback to the teacher about their own teaching, e.g., pace, delivery, methods, etc., across the entire term.

Context of the Assessment

Regarding the specific context and population of the audience for our test, our test is intended for ESL students attending an academic language program in the United States. The students have enrolled in this program primarily because their TOEFL scores are below the entry level requirements (below 500) of the attached American university that they wish to enter. Based on scores from admissions and placement tests other than their TOEFL, all students who make up the audience for our test have been placed into the 100 level language courses, reading, writing, listening and speaking, and grammar. Our test focuses on academic language and objectives from their 100 level reading course. Found on the actual program's website, the description for the 100 level reading course that the students are taking is as follows:

COURSE DESCRIPTION: Reading 113

This course provides a strong foundation in basic reading skills for new or beginning-level learners of English. Your reading fluency and comprehension will increase as you study and practice effective strategies, develop your vocabulary, and read many interesting stories and articles. (HELP, 2013)

Also, from the website, the course has six student learning objectives (SLOs). They show that by the end of the course students will:

1. Increase reading speed (to approximately 100 wpm)
2. Locate main ideas and specific pieces of information when reading a short, modified text
3. Recall some facts or information from a short, modified text

4. Develop strategies or ways for guessing the meanings of new words from context
5. Increase knowledge of general English vocabulary
6. Respond to reading passages through discussion and writing

(HELP, 2013)

From these SLOs, our test will focus on three of the six objectives, (a) SLO 2: locate main ideas and specific pieces of information when reading a short, modified text, (b) SLO 3: recall some facts or information from a short, modified text, and (c) SLO 5: increase knowledge of general English vocabulary. Because these objectives do not specify the actual level of the students, their language ability level, or the nature of the texts that the students will study, we also used the actual textbook for the course as a source for establishing the learner needs of the class. The textbook for the course is Bonesteel's (2011) *Real Reading 1*. We had access to the student's book, the teacher's manual, and the CD that accompanied the student's book. According to the teacher's note found in the introduction to *Real Reading 1*, the textbook is

[t]he first book in a four-level (beginning, low intermediate, intermediate, and high intermediate) intensive reading series for learners of English. The books in the series feature high-interest reading that have been carefully written or adapted from authentic sources to allow effective comprehension by learners at each level. The aim is for learners to be able to engage with content in a meaningful and authentic way, as readers do in their native language. For example, learners who use *Real Reading* will be able to read to learn or feel something new, to evaluate information and ideas, to experience or share an emotion, to see something from a new perspective, or simply to get pleasure from reading in English. High-interest topics include superstitions, shyness, neuroscience, sports, magic, and technology, among others. (Bonesteel, 2001, p. VIII).

For the purposes of our test, the unit that we chose from the textbook is Unit 2: Folktales. Unit 2 consists of chapters 3 and 4. The content and length of reading passages in both chapters is consistent and in alignment with the theme of the entire unit, folktales. Reading passages found in the text are approximately 300-400 words long. We will write the test based on the vocabulary from the unit. Chapter 3 and 4 have ten vocabulary words each.

Construct of the Assessment

Our test will measure two constructs, (a) reading comprehension and (b) vocabulary retention. These constructs are reflected in the three SLO's that will be assessed by our test. SLO 2 and 3 will be assessed through the reading comprehension construct and SLO 5 will be assessed through the vocabulary construct.

Reading Comprehension Construct

According to Grabe (2009), "Reading assessment has great power to inform researchers, teachers, administrators, and policy makers" (p. 352). Our test will measure the construct of reading comprehension by using five true and false items. Grabe (2009) classifies this item type as a dichotomous assessment task format, and is one of twenty assessment formats that he lists. "Because reading assessment practices can cover a range of purposes and uses" (Grabe, 2009, p. 356), we narrowed the range of purposes to only include SLOs 2 and 3 from above, *locate main ideas and specific pieces of information* and *recall some facts or information from a short, modified text*, respectively. These SLOs are in alignment with Grabe's (2009) list of major component abilities for reading comprehension. Specifically, he lists fourteen component abilities. Three of those, (a) main ideas comprehension, (b) recall of relevant details, (c) inferences about the main idea of a text, directly correspond with the reading comprehension SLOs from our target audience's class.

Vocabulary Construct

Theoretically, while vocabulary is not one of the four major skills, i.e., listening, reading, speaking, and writing, it is however an important component in English pedagogy and "vocabulary tests are used for a wide range of instructional purposes" (Read & Chapelle, 2001, p. 5). That said, many theories exist on the teaching and assessment of vocabulary in ESL education (Read & Chapelle, 2001; Grabe, 2009; Walter, 2004). Of those theories, several sources comment on the use of other skills in assessing knowledge of vocabulary both directly and indirectly, e.g., vocabulary tests using reading comprehension, and more direct approaches to test vocabulary that include matching synonyms, and fill-in-the-blank/sentence completion (Read & Chapelle, 2001). Specifically, Read and Chapelle's (2001) article looks at the construct of vocabulary and the ways it is measured. Our test measures what Read and Chapelle calls a *selective* vocabulary construct. This means that particular target words were chosen as the focus of the assessment. The assessment is then used as a means of tracking how much knowledge of specific words a group of learners acquired through encountering them in their reading of a particular passage or a unit of instruction (Read & Chapelle, 2001).

Operationalization of the Construct

As stated previously, our test will assess the two constructs of reading comprehension and vocabulary. The reading comprehension construct will be operationalized by the students answering multiple choice questions based on a new passage. The passage is from the teacher's manual and is of the same literary genre as the passages from the unit, i.e., the same length, and the same level of difficulty according to the textbook. Vocabulary from the unit will be tested in two ways. First, vocabulary will be tested directly using a multiple-choice style format where students have to match vocabulary words to synonymous phrases within sentence-level context.

Second, also at the sentence level, students will be asked to complete sentences through a fill-in-the-blank format. The students will be asked to fill in the vocabulary word that corresponds with the blank from a word bank.

Item Specifications

Before creating our test, we wrote three separate item specifications, one for each section and item type on our test. A summary of those three item specifications can be found below (for complete item specifications see Appendix A):

1. reading passage with a dichotomous items (true/false)
2. vocabulary in context with multiple-choice questions
3. fill-in-the-blank sentence completion with vocabulary from a word bank

Scoring the Test

There will be twenty items on the test with five items in the reading section, seven items in the first vocabulary section, and eight items in the second vocabulary section. Each test item will be worth one point each for a maximum of twenty points.

Test-Writing Procedure

In writing our test (See Appendix B for completed test. See Appendix C for answer key), the process was completed in four stages. First, we conducted a small needs analysis of the context of our targeted audience, a content analysis of the course our audience was taking along with the textbook they were using, and a literature review of articles pertaining to our construct. In the first stage we also defined our construct for the test, limited its purpose as a student evaluation tool to measure SLO achievement, and established the number of item types we needed to operationalize our construct.

For the second stage, we wrote our item specifications and began to create our test items. During the second stage, all decisions regarding test design were made, decisions such as prompt and distractor format, number of items per section, and choosing the passage for the reading comprehension section. One decision in particular regarded the number of option choices for the multiple-choice section of the test. We decided to use a three-option format for two reasons, (a) our test is not a high stakes test and (b) 4-option and 5-option items tend to both test test-taking skills (which is not a part of our construct) and provide more opportunities for error and poor distractors (Lee & Winke, 2013).

In the third stage, we finalized our pilot test and conducted a content analysis of our test using researcher triangulation and a peer-review to check its validity in measuring the construct. Minor changes were made and the finished test was created.

In the fourth stage, we piloted our test on a population similar to our target audience. The pilot test will be discussed further in the task analysis section below. Also, in the fourth stage we analyzed the results of our pilot test, which can be seen in the next section.

Test Evaluation

Because our test is a criterion-referenced, end-of-unit test, we would ultimately want to evaluate the success of the test by determining how the test items compare to course's SLOs, and how the test shows that these SLOs were met. While this type of evaluation is most easily conducted using a difference index analysis, because this assignment does not allow for such assessment procedures, we have opted for more general evaluation procedures to determine the validity of our test: content analysis and task analysis. Beyond these types of evaluation, we will also propose further data analysis that could be conducted after a full version of the test was administered and scored.

Content Analysis

According to the assignment for this project, content analysis requires that test makers think logically about how direct and complete their tests are. According to Brown (2012), "the goal of content analysis for a CRT is to determine the degree to which each item is measuring the content that it was designed to measure, and the degree to which that content should be measured at all" (p. 77). For this type of analysis, we asked ourselves two questions,

- 1 How relevant are our test item types to the ability that we want to measure?
- 2 How well does our test items represent all of the aspects of the definition of the language ability that we want to measure?

To answer these questions, we utilized researcher triangulation and feedback from peer-reviewers. Individually, each of us answered these questions separately and then came together and discussed "which items should be kept in the revised version of the test and which should be reworked or thrown out" (Brown, 2012, p. 77) while thinking about how each item measured what we wanted to measure and to what extent. Beyond changing the wording of a few items, we also changed the item format of the reading comprehension section. Instead of using a multiple-choice format, as we had initially intended, we decided on a True/False format. We felt that this format was more appropriate for the language level of our target audience. Another item that we decided to take out and replace required students to have outside knowledge of the genre of storytelling and folktales. While this information would be taught during the course of the unit, it did not fit within the construct of our assessment.

Peer-review Feedback. Another way we analyzed the content of our test was to get peer-review feedback. For this feedback, we asked two students who had previously taken the SLS490: Testing course to take and review our test. After each-peer reviewer took the test, we

asked each of them to, in a sense, work backwards from the test design process. We asked them, "What is the construct or constructs that you think is being measured with this test?"

Unanimously, they both predicted that we were measuring reading comprehension and vocabulary with a strict focus on beginner-level ESL students. One peer reviewer even noted that our reading comprehension section was measuring students' ability to distinguish main ideas and details from a passage. In addition, we asked the peer-reviewers to answer a few questions on the clarity of the test items and instructions. After the following comment from a reviewer, we decided to change the instructions for the second section of our test:

Not all of the instructions were clear. The instructions for the first vocabulary section, where students had to circle the correct word with the same meaning as the underlined word in the sentences, were a bit wordy and ambiguous. The students may think they are able to circle two choices the way the wording stands now.

After completing our content analysis, we made changes and finalized the pilot version of the test. We then administered it to a population similar to our target audience.

Task Analysis

In piloting our test, while we wanted to use the actual target population, however that was beyond the scope of this project. Ultimately, we piloted our test using nine participants from the same program as our target audience. The difference being that our pilot-tested population had already completed the 100-level reading class, and are now in the middle of their 200-level term. In essence, because we didn't have the resources to pilot our test before and after 100-level students' exposure to Unit 2, which provides the content for our test, pilot testing on a 200-level population simulates a post-test environment in which participants have already experienced (been taught and learned) content from Unit 2 of the *Real Reading* textbook.

During the administration of our pilot test, we all observed the students while they were taking the test and asked them to complete a 6-item survey. The survey questions are as follows:

1. Are you a native speaker of English? If not, what is your first language?
2. How long did it take you to complete this test?
3. Were the instructions easy to understand?
4. Did you have any problems while taking this test?
5. Do you think this test is good for low-level ESL students at HELP? Why or why not?
6. Do you have any comments or suggestions for improving this test?

First we will discuss the results of the observation. During the observation, our attention was drawn to two items, 6 and 9. For item 6 students needed to match the word *smart* to its synonym *clever*. However, we had overlooked that *clever* was spelled wrong, and *cleaver* was what the students saw. Despite this error, 100% of the students still chose it as the correct answer. We have no data to indicate if the students were aware of this error during the pilot test or not. The issue with item 9 that was raised was of a cultural nature. That is, while students had to match the prompt *says sorry* to its synonym, *apologize*, another distractor was *say excuse me*. This was raised as an issue by the Japanese students due to the fact that the Japanese *sumimasen* can have both meanings. Despite that concern, item 9 had an item facility score of 1.0.

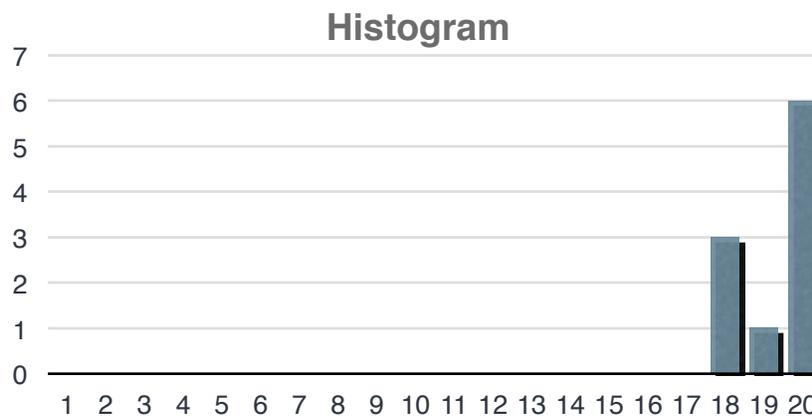
The complete feedback from the examinees' questionnaires can be seen in Appendix E. Data from the questionnaire show that all students were non-native speakers of English with a high variety of L1s (Japanese, Portuguese, Vietnamese, Polish, Arabic, Chinese). All students completed the test within 10 minutes, which was expected. All students found the instructions clear and did not have problems while taking the test. Of the suggestions for improving the test, one student mentioned that the test could be more difficult. While this is evidenced by the high

item facilities and scores, this result is expected in a post-test situation. Of interest, one student mentioned that our test was similar to other tests in their experience, while another said that "it's a good test because the reading sections does not require reading strategies such as scanning."

Proposed Data Analysis

After piloting our test, we scored each test and utilized test analysis procedures such as calculating descriptive statistics and finding the item facility of each item (See Appendix D). The test statistics (i.e. mean, standard deviation, etc.) as well as a histogram of the score distribution can be seen below:

N	9
k	20
Mean	19.44
Mode	20
Median	20
Midpoint	19
low-high	18-20
range	3
SD	0.88



From the descriptive statistics and the histogram we can see a strong negative skewing of the scores. This is consistent with our expectations for a post-test, CRT situation.

Beyond the scope of the pilot test, we would like to propose that further piloting be done in a pre-test situation so that a difference index can be calculated. This could be used to improve overall reliability and strength of the test. Also, further validity studies could be conducted by checking internal consistency. We suggest that internal consistency be determined using Cronbach's alpha because not all subsections of the test are dichotomous.

Individual Researcher Reflections

Researcher 1 (Matt)

When we were first told that our last assignment would be the development of a test, I thought it would be relatively easy. Because I have nine years of language teaching experience, I have written a lot of language tests, standardized, performative, norm-referenced, criterion-referenced, etc. I thought that the development of a test for this course would be a simple process that I had done many times before. I was wrong. While I may have had some experience as a teacher and test-writer, I had had no foundation in the theories behind test development and test analysis. Who knows how many of the beautiful tests I made may have been inefficient, measured the wrong construct, did not correspond with student expectations, or contained any other of a number of possible errors. In the end, while luckily the process of developing a test based on theory and proper test analysis strategies supported some of my common sense approaches to test making in the past, I have learned a lot more about the entire test development and analysis process than I had expected. I will never view another test in the same way again.

On a specific note, I didn't expect to become as invested I was in the development of this test. I really wish we had been able to pilot our test as a pre-test as well. After piloting our test in a post-test environment, I knew that the only way to fully check the decision-making power of

our test would be to also test it on students who were at the level of our target audience without exposure to the unit content. For this scenario, we would have to wait until the Fall 2013 semester--not possible.

Researcher 2 (Kaitlyn)

In the beginning of this project, I felt that it was going to be a lot of work. During our first meeting, we were making it too difficult when it didn't have to be. Once we chose the textbook to base our test on the process became much easier. I felt that writing the test was a good experience since I never had to write a test before. Having material to base it on rather than making it from scratch prevented the process from being too difficult. Also having a group to work with helped immensely. We used Google Drive for most of the project and we could constantly spell-check and edit errors so that the writing process for this paper went smoothly. I would have liked to have had at least one of the test-takers be at the target level of the test so that we could have had more accurate results, but the fact that we had nine people pilot the test was a great surprise and helped in collecting data. In that aspect, I feel we went above and beyond what was required.

Researcher 3 (Lisa)

The process of this test development project helped me to better understand some of the concepts of second language testing that had been discussed during the course of this semester. During the construction phase of the test, articulating the definition our construct was the most difficult. We each had different ideas for the direction of this project, and during the brainstorming process, we might've been making this more difficult than it needed to be. Once there was a consensus on the purpose of our language test, the target level of the examinees, and what exactly we wanted our test to test for, our group was able to see the light at the end of the

tunnel. Experiencing a mini version of the planning and development process of test-making gave me new esteem for teachers and educators alike. I didn't realize how meticulous the process was. Reviewing the administered tests and questionnaires was the rewarding part of this project as well as eye-opening. I felt like we had successfully accomplished our goal. We were able to observe possible distractors, calculate item facility from the scores, and were given feedback on the construct on the target audience. Also, I didn't expect that we would have any problems with items on the test because of our awareness while making the items. However, while observing the examinees, we realized that we were not as careful as we had thought.

Conclusion

In conclusion, we all agreed that it was a great experience overall and enjoyed working with each other.

References

- Bonesteel, L., & Nation, P. (2011). *Real reading 1: Creating an authentic reading experience*. White Plains, NY: Pearson Longman.
- Brown, J. D. (2012). *Testing in language programs*. McGraw-Hill.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Hawai'i English Language Program Homepage. (2013). University of Hawaii at Manoa. Retrieved April 2, 2013, from <http://manoa.hawaii.edu/eslhelp/wordpress/>
- Lee, H., & Winke P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, 30(1), 99-123.
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1): 1-32.

Appendix A: Item Specifications

Summary

We will have 3 item specifications.

1. reading passage with True or False (Lisa)
2. vocabulary synonym multiple-choice with word bank (Kaitlyn)
3. fill-in-the-blank sentence completion (Matthew)

Item Specification 1:

- General Description: After reading a passage of 375-400 words, students will answer comprehension questions based on the passage.
- Sample Item:

Directions: Read the following passage. After you have finished reading, answer the questions that follow the passage.

Reading:

The Pot of Gold

1 Tom Fitzpatrick was out walking one morning when he saw a very, very small man with a green jacket and a red cap sitting under a rock on the grass, fixing a shoe.

5 Tom was very excited because he knew that the little man was a leprechaun. Leprechauns are magical people. They say that if you follow a leprechaun, you will find a pot of gold. But if you take your eyes off him for one moment, he will disappear.

10 “Good morning,” said Tom to the little man. “What are you doing?”

“That’s not your business,” said the small man in an unfriendly way.

15 Tom got angry and picked up the little man, saying, “Tell me now where you have put your gold, or I’ll never let you go!”

The little man was frightened. “All right,” he said. “I apologize.” I’ll show you where the gold is if you let me go.” And Tom

20 agreed.

So off they went. The leprechaun went in front, and Tom never took his eyes off him. They walked for a long, long way. Finally

25 they came to a field that was full of ragweed.¹ The leprechaun pointed to one of the largest plants and said, “See that big plant there? Dig under it and you’ll find a pot full of gold.”

¹ **ragweed**: a plant that grows wild in the countryside

Tom knew it was true because leprechauns

30 never lie. He couldn’t wait to get his hands on the gold. But he didn’t have a shovel to dig a hole. He would first have to go home and get one. So he took a red handkerchief from his pocket and tied it around the plant so that he

35 would remember it. “Now,” he said to the leprechaun. “Promise me that you will not remove this handkerchief² until I come back?”

“I won’t,” said the leprechaun. May I go now?”

40 “All right then,” said Tom. “You may go, and good luck to you!” He went back to his house to get a shovel, and then he ran back to the field as fast as he could, with his shovel on his shoulder.

45 But as soon as Tom returned to the field, he knew that the clever leprechaun had played a trick. Every plant in the field had a red handkerchief on it, and each plant looked exactly the same!

50 Tom dug first under one plant, and then under another, for many hours, but in the end he had to go home, dirty and empty handed.

² **handkerchief**: a small piece of fabric that you use for blowing your nose

2. The father carried the little girl on his _____ . (shoulders)
 3. Most people eat three _____ a day: breakfast, lunch, and dinner. (meals)
 4. Where were you? Tell me the truth. Don't _____ to me. (lie)
 5. When we did something wrong, the teacher _____ us. (punished)
 6. Here's some candy. Please _____ it with your sister. (share)
 7. The boy took the money and put it into his _____. (pocket)
 8. The children have been on the beach, and now they're _____ with sand. (covered)
- Prompt Attributes:
 1. The prompt for each item will consist of a 10-15 word sentence or set of sentences with a missing word.
 2. The missing word from each sentence will be replaced with a line (blank space that is underscored).
 3. The line in each sentence will all be 20 spaces long to accommodate the longest missing word and not give a hint as to the length of the correct word.
 - Response Attributes:
 1. Answer choices found in the word bank will come from Unit 2: Folktales in the *Real Reading* textbook.
 2. While the word bank will have ten words, the students will not use two of them.
 3. To answer each item, students will physically write the word that completes the sentence in each blank.
 - Specification Supplement (optional):
 1. The word bank for these items are from Unit 2: Folktales in the *Real Reading* textbook.

Appendix B: Test

Name: _____

Score: ____/20

Date: _____

UNIT 2 TEST

I. Reading Comprehension

Directions: Read the following passage. After you have finished reading, answer the questions with T for True or F for False. Answer the questions using information from the passage.

The Pot of Gold

- 1 Tom Fitzpatrick was out walking one morning when he saw a very, very small man with a green jacket and a red cap sitting under a rock on the grass, fixing a shoe.
- 5 Tom was very excited because he knew that the little man was a leprechaun. Leprechauns are magical people. They say that if you follow a leprechaun, you will find a pot of gold. But if you take your eyes off him for one moment, he will disappear.

“Good morning,” said Tom to the little man. “What are you doing?”

“That’s not your business,” said the small man in an unfriendly way.

- 15 Tom got angry and picked up the little man, saying, “Tell me now where you have put your gold, or I’ll never let you go!”

The little man was frightened. “All right,” he said. “I apologize.” I’ll show you where the gold is if you let me go.” And Tom agreed.

- So off they went. The leprechaun went in front, and Tom never took his eyes off him. They walked for a long, long way. Finally
- 25 they came to a field that was full of ragweed.¹ The leprechaun pointed to one of the largest plants and said, “See that big plant there? Dig under it and you’ll find a pot full of gold.”

¹ **ragweed:** a plant that grows wild in the countryside

- Tom knew it was true because leprechauns
- 30 never lie. He couldn’t wait to get his hands on the gold. But he didn’t have a shovel to dig a hole. He would first have to go home and get one. So he took a red handkerchief from his pocket and tied it around the plant so that he
- 35 would remember it. “Now,” he said to the leprechaun. “Promise me that you will not remove this handkerchief² until I come back?”

“I won’t,” said the leprechaun. May I go now?”

- 40 “All right then,” said Tom. “You may go, and good luck to you!” He went back to his house to get a shovel, and then he ran back to the field as fast as he could, with his shovel on his shoulder.

- 45 But as soon as Tom returned to the field, he knew that the clever leprechaun had played a trick. Every plant in the field had a red handkerchief on it, and each plant looked exactly the same!

- 50 Tom dug first under one plant, and then under another, for many hours, but in the end he had to go home, dirty and empty handed.

² **handkerchief:** a small piece of fabric that you use for blowing your nose

- _____ 1. Tom wanted the leprechaun to show him the pot of gold.
- _____ 2. The leprechaun was happy to help Tom.
- _____ 3. Tom put a red handkerchief on the plant.
- _____ 4. When Tom got back, the leprechaun was still there.
- _____ 5. In the end, Tom found the gold.

II. Vocabulary

A. Multiple-Choice

Directions: Read each sentence. Then look at the answer choices. Circle the choice that means the same as the underlined word(s) in the sentence.

6. Foxes are very smart.

- a. dumb b. clever c. tricky

7. No two folktales are exactly alike. They often have different endings.

- a. the same b. different c. unique

8. The woman felt bad when her daughter behaved badly during the meal.

- a. distant b. good c. ashamed

9. Chris should say sorry to his mother for his bad behavior.

- a. apologize b. say excuse me c. talk

10. The man told us to continue on this road for about five minutes.

- a. rest on b. follow c. stop on

11. When the children saw the old lady, they ran away.

- a. How b. Then c. As soon as

12. Please take off your shoes in the house.

- a. put on b. remove c. turn off

B. Fill-in-the-Blank

Directions: Complete the sentences with the words from the word bank. Two of the words will not be used.

Word Bank:

characteristics	grass	meals	punished	shoulders
covered	lie	pocket	share	trick

13. One of the _____ of spiders is that they have eight legs.
14. The father carried the little girl on his _____.
15. Most people eat three _____ a day, breakfast, lunch, and dinner.
16. Where were you? Tell me the truth. Don't _____ to me.
17. When we did something wrong, the teacher _____ us.
18. Here's some candy. Please _____ it with your sister.
19. The boy took the money and put it into his _____.
20. The children have been on the beach, and now they're _____ with sand.

Appendix C: Answer Key**I. Reading Comprehension**

1. T
2. F
3. T
4. F
5. F

II. Vocabulary**A. Multiple Choice**

6. B (clever)
7. A (the same)
8. C (ashamed)
9. A (apologize)
10. B (follow)
11. C (As soon as)
12. B (remove)

B. Fill-in-the-Blank

13. characteristics
14. shoulders
15. meals
16. lie
17. punished
18. share
19. pocket
20. covered

Appendix E: Questionnaire Results

Unit 2 Pilot Test Questionnaire Results

Student	Native or non-native English speaker? L1?	How long did it take to compete the test?	Were the instructions easy to understand?	Did you have any problems taking this test?	Do you think this test is good for low-level ESL student at HELP?	Comments or Suggestions for improving this test?
001	non-native, Japanese	10 min	very clear	I did not know a few words.	I think that it's a good test because the reading sections does not required reading strategies such as scanning.	
002	non-native, Japanese	8 min	yes	no	Yes, I didn't need to guess the word "Leprechauns" because the meaning is explained in the plot.	
003	non-native, Arabic	10 min	yes	no	Yes, it has simple vocab, understandable, clear instructions.	The test is just like my regular tests.
004	non-native, Portuguese	10 min	yes	not much	Yes, it is adequate.	It could be a little more difficult.
005	non-native, Polish	9 min	yes	no	Yes, it helps to improve vocabulary.	
006	non-native, Japanese	9 min	yes	no	Yes.	No.
007	non-native, Portuguese	8 min	yes, they are	no :)	Yes because it works with comprehension and vocabulary.	N/A
008	non-native, Chinese	I don't know	yes, I think so	no	It depends. Every students have different knowledge background.	No time to think about it deeper.
009	non-native, Vietnamese	10 min	yes	no	Yes because it's easy to understand and vocabulary isn't difficult.	No.