



Assessing vocabulary

Authenticity and Context

*Corey Gonzales
AL6730
Dr. Hanh Nguyen
May 3, 2012*

I. Introduction

The overall goal of this paper is to review principles in vocabulary assessment and the use of authenticity in item design as well as how these principles can be applied in the production, revision, administration and analysis of an assessment designed as a vocabulary test for ELL students in an American university classroom. The paper examines the setting of the project; the project overall including objectives and specifications; the results of the students; a reflection and discussion of this particular assessment, assessing the effectiveness of the task; and finally, future inquiries based on the aforementioned task/developed assessment.

II. Project Description – *Background Information*

Host class

The name of the host class is International Education, one of the courses in the Bridge Program. The class, taught by Dr. Brian Rugen, meets every Monday, Wednesday, and Friday from 11:50 to 12:45pm. The students' proficiency level is described as upper intermediate to advance. The students' needs in the program are to develop critical thinking and visual literacy skills, along with academic reading, writing, and vocabulary development. The objectives of the class are to examine and challenge students' beliefs and values regarding selected issues in educational systems in the U.S. and other regions through extensive readings, discussions, and mini-lectures. The instruction approach is content-based where the teacher provides materials of the topic. There are nineteen students in total and most of them are Asian.

Host institution

The host institution is Hawaii Pacific University located in downtown Honolulu, Hawaii. This course is listed under the International Bridge Program (IBP) which offers an opportunity for qualified international students to improve their English competency, academic skills, and

acquire content-based knowledge. The overall goal of the program is to prepare students for academic success at Hawaii Pacific University.

Group members

The group members for this project were Nga Loc from Vietnam, Corey Gonzales from Colorado, USA, and Yukiko Oki from Japan. Nga has taught EFL in Vietnam for ten years and has designed progress and final achievement tests for her high school students. Corey has never taught ESL/EFL before but has designed an achievement test which was a language arts tests for high school students. Yukiko has taught EFL to young learners for a year in Japan but has never designed any type of assessment before this one.

Language Assessment Instrument

This vocabulary quiz is an achievement test which was administered on Friday March 23, 2012. The last 30 minutes of class time was used for administering the quiz. Nga, Yukiko, and Corey were present for the administration. As five students were absent on this day, they took the quiz ten days later (after Spring Break) on April 2nd in the professor's office. They were given the same amount of time. The actual quiz administered can be found in Appendix A.

III. Objectives

As determined by Dr. Rugen and the group members, the objective for the test was to “assess understanding of the vocabulary words/terminologies (words given by the host teacher) that were covered in the module of Single-Sex Schooling.” While 32 words were given by the professor to be tested upon, only 29 of those words were selected for the actual exam as the host professor gave the test creators the option of which words should be used and tested, as long as they came from the list. The words not chosen were left out due to the difficulty in creating test items for them. Both the host professor and group members agreed upon this decision.

IV. Specifications

1. Specifications of content:
 - a. Operations:
 - Ability to complete sentences in context by choosing the right word/terminology
 - Ability to identify a word/terminology with its synonym/meaning
 - Ability to complete sentences using the given words in the right forms (consume, acquire, category, resource, and injure)
 - b. Types of text: Teacher-created sentences
 - c. Length of text: 3 pages, 24 sentences (with 14 words on average)
 - d. Addressees of text: Students learning academic English in relation to the topic of single-sex schooling in the U.S.
 - e. Topics: Single-sex-schooling, educational/academic settings
 - f. Readability (Flesh-Kincaid or grade level): Flesch Kincaid level 8 (taking into account that students are EFL/ESL learners)
 - g. Structural range: Intermediate-advanced
 - h. Vocabulary range: Academic, terminologies from the topic of single-sex schooling
 - i. Dialect and style: Standard American English dialect, formal style
2. Structure, timing, medium, and techniques:
 - a. Test structure: 3 sections
 - b. Number of items: 28 items
 - c. Number of passages: 24 sentences (with 14 words on average)
 - d. Medium: paper and pencil
 - e. Timing: 30 minutes for entire test
 - f. Testing techniques:
 - Section 1: 8 multiple choice items using a word bank
 - Section 2: 7 multiple choice items each with 4 options (one key)
 - Section 3: 13 gap filling items, base form of word provided
3. Critical level of performance: 21 correct out of 28 points (total) indicates a pass (75%)

4. Scoring procedure: All items are either correct or wrong. 1 point is given to each correct answer, 0 points for wrong. Candidates need to have correct spelling and form for a point in sections 1 and 3.
5. Sampling: Information on education and single-sex-schooling will be drawn from available publications such as online articles.

V. Student Results

The data chosen for the analysis of the exam focuses on the scores of all 19 students who took the exam, with 14 students taking the exam on March 23, 2012 and five students taking the exam on April 2, 2012. The average score for the quiz was 21.4. The lowest score was 11 and the highest score received was 27 of 28 total points. The most frequent score was a 25, which was received by four of the 19 students. A complete a list of all student scores can be found in Appendix B and descriptive statistical analysis of the exam is found in Appendix C.

VI. Reflection and Discussion

As Read (2000) explains, “vocabulary can be seen as a priority area in language teaching, requiring tests to monitor the learners’ progress in vocabulary learning and to assess how adequate their vocabulary knowledge is to meet their communication needs” (pp. 1-2). The test given to the students in Dr. Rugen’s IBP class proves as an example to this idea. As stated before in the classroom objectives, students were to assess understanding of the vocabulary words/terminologies that were covered in the module of Single-Sex Schooling. This exam requires students to recall information they had learned about single-sex schooling in the US and the terms used within this unit to explain said topic. By using their knowledge of ideas presented in the unit about single-sex schooling, students showed their knowledge of the vocabulary words chosen by their professor.

Examining the test as a whole, it appeared to be an effective means to test these students’ understanding of the words’ meaning, and, on occasion, form. It is first important to note that the

host teacher for this classroom requested that the group creating this exam provide all of the words to be tested and that student production of the vocabulary words was not directly tested. More about that will be discussed later. Each section of the test brought forth the use of context for students to choose responses on the exam. The first and second section focused on the use of classroom, and specifically single-sex schooling in the US as context for the sentences provided, while the third referred on contexts that differentiated from specific classroom topics.

The first section of the test used context to see if students were able to choose the correct word to fit in the various sentences, i.e. gap filling. In this section, students were required to choose the correct words from a word bank to fill in the blanks in sentences. The sentences were drafted from model sentences from an article the students read from the publication *The Atlantic*. In this section students had to rely on their understanding of not only the content learned about single-sex schooling, but also how specific vocabulary words fit into that context. In some cases, knowledge of form of the vocabulary words was also tested. The very first sentence students were required to complete was:

- (1) Is sending kids to single-sex schools better even though the tuition may be expensive and out of reach ?

As it was requested by the host teacher to provide the words for students, the members of the group decided on choosing the classroom context as the context for the gap-filling sentences. Bachman and Palmer (1996), as cited in Uçkun, B. (2008), explain that the “syllabus-based definition is appropriate when vocabulary assessment takes place within a course of study. Within this framework, the lexical items and the vocabulary skills to be assessed are specified in accordance with the learning objectives of the course,” i.e. single-sex schooling (p. 104). Additionally, according to Singleton (1999), as cited in Read, J. & Chapelle, C. A. (2001), “it is no longer justifiable to restrict vocabulary research to measures involving knowledge of

individual content words, and the scope of vocabulary teaching needs to be similarly expanded” (p. 2). Using a context about a new topic, perhaps unfamiliar to students, allows this idea to be incorporated into language testing.

Overall, the average I.F. for the first section was 0.684. While that is not ideal in this situation, it is on the upper end. It also appeared that while students were able to choose the correct word(s) to fit into the blank, sometimes having to produce the correct form of the word is where they had problems. For example, the fourth item on the test required that the students choose the word “proponents” to correctly fill in the gap. While eight students produced the correct form, four of the students who chose the correct word only wrote “proponent,” missing the plural ‘-s’ as the ending. It would appear that students knew the correct word, just not the correct form of the word to complete the gap. It is unclear if whether they did not have understanding about the sentence to know ‘s’ was needed, if they didn’t understand directions, or if they didn’t follow the directions. The test graders decided not to give credit for this incorrect response, as the focus of this section was to have students not only produce the correct meaning of the word, but also the correct form.

Section two of the test was multiple choice, requiring students to choose the synonym for the vocabulary words underlined in the sentences. This section also included classroom context for students as the word was in a sentence that had the single-sex classroom theme. For example:

9. Middle school was a fragile time for me.

- a. memorable **b. delicate** c. fun d. broken

It appeared that the context was quite useful for the students to decipher the meaning of the underlined word and choose the word that was synonymous as opposed to no context; although to analyze this, another exam without the context would have to be taken (either by the same students or by another similar group of students.) Overall, this section had an average I.F.

of 0.774; a great deal better than the first section. Most likely, this is because even more context was given for students and students did not have to demonstrate any production skills. There could be some miscalculation in final scores due to guessing; however four options were presented for students to “reduce the effect of guessing” (Hughes, 2003, p. 77). While this type of item is not necessarily favored (as it is not communicative,) the use of provided context is viewed as practical as “a word can have different meanings and be used as more than one part of speech” (Read, 2000, p. 99). The test items for this section all provided contextual information for students so that a correct choice could be made and that students would be able to see an example of how the vocabulary word can be used in a correct sentence in English.

As mentioned above, the host teacher was not concerned with students’ production abilities of vocabulary words for this specific test. That being said, section three of the test required the most production from students. In this section, students were required to generate the correct forms of the provided words in a particular context. As the words in this section are more academic and not necessarily used explicitly to explain or describe single-sex schooling, the sentences created did not include that context specific to that theme.

This section had the highest I.F. of the three test sections with an average of 0.851. While not ideal, it is an improvement from the previous two sections. In this section, the base form of the word was provided and students were required to produce the correct form of the words in the sentences. For example:

16. (3 pts.) **injure**

While playing football after school, Mark was injured and this horrible injury kept him from participating in the tournament. Mark learned the lesson that playing football can injure you.

It seemed that students were able to generate correct responses when they knew that each blank had to be a different form of the word and the words being tested were academic words they had most likely seen before, outside of the context of single-sex schooling and possibly in various situations.

Because the average of this test being described as a C+ (above average,) there were some areas of the exam that could have been created differently in order for students to have had more success. The authors of the exam chose not include any demonstration of production by the students. This is one aspect of the exam that could be changed if the test were recreated. Read (2000) explains that there is great need for a “communicative approach to language testing” (p. 4). This would mean that students are required to complete tasks “simulating communication activities that the learners are likely to be engaged in outside of the classroom” as opposed to the discrete-point approach (Read, 2000, 4). Basically, the authenticity of an exam that focuses on the communicative approach rather than the discrete-point approach increases immensely since students are being required to complete a task that they would be more likely to come into contact with in real-life situations when their skills in English are needed.

Realistically, this test would have had to take on a completely different structure to meet a more communicative approach. In an ideal situation, that could have been possible. The students in this classroom were preparing to give debates after spring break (and after the vocabulary test) which would have placed them on different sides of arguments about single-sex schooling. While this fact was unknown to all of the members of the group who created the test until the day the test was given, this might have been a great place to test vocabulary knowledge and practical use. If students had been required to demonstrate their knowledge of specific words while having a discussion with their classmates about the provided topic, a more authentic and more

communicative test would have been created. Unfortunately, this was not an option offered to test creators. Additionally, it would have been hard to test some of the vocabulary words since students would then need to know all the words required, perhaps having a list of the words needed. This might force the students to use the vocabulary in unnatural and unauthentic ways.

To reexamine and change the test without completely altering the form, a few components could have been adjusted. Starting from the beginning of the test, there was a question in section one that was found to have two possible correct answers. Item number two was written so that “shortchanges” was to be the correct response, however when the test was returned to the host teacher, he brought up the fact that “shortchanged” could also fit as a possible correct response:

(2) Some studies attempt to prove how coeducation shortchanges girls.

Despite the sentence being grammatically correct, the creators of the test did not think that “shortchanged” seemed to fit—as coeducation is still occurring, and therefore by means of the test item so are the studies; therefore, the response could not have been in the past tense. To create a sentence that could not have more than one correct response, perhaps the item could have read as “Some studies attempt to prove how *the current* coeducation system shortchanges girls.” This would explain that the studies are still happening, as is coeducation, and the correct response should stay in the present tense.

An additional area in section one where the students’ I.F. was quite low was on test item number four:

(4) Proponents of single-sex schools think all-girls schools are good.

This item had eight students choosing the correct response and four students choosing “proponent” but not changing the form of the word. Five other students chose the word “opponent” as their response. While there does not appear to be an option as test creators to

change this item, there would definitely need to be an area for washback post exam. In class following the test, the teacher might opt to do a mini-lesson on the difference between the two words so that students would be more likely to know the difference in any situation, not just one relating to single-sex schooling.

A third test item that proved significantly difficult for the students was item number ten, which appeared in section two of the test:

10. To some degree, it is very hard for girls and boys to come to a consensus in some situations.

a. settlement b. conversation c. discussion d. commitment

Of all items on the test, this one had the lowest I.F. of only 0.211. While all distractors appeared to be successful, (this was the only MC item that had a distractor chosen by at least one student,) perhaps the distractors were *too* successful. Four of the students chose the correct response; and 11 of the students chose “commitment.” The word fits grammatically, but it does not make as much sense as settlement. One possible reason for most of the students choosing “commitment” could be due to the collocation of the word “commitment” between men and women in society today. In one of the previous drafts of this test (a) had been “agreement” as opposed to “settlement.” It was noted that because “agreement” begins with a vowel it would not fit the stem where the vocabulary word “consensus” fell in the sentence. Creating the test item so that it read: “To some degree, it is very hard for girls and boys to come to a/an (or a(n) consensus in some situations.” might give students reason to think that the only possible answer could be the one starting with the vowel. To create a better item, perhaps this item could have been changed in the fact that more context was provided for students. Another possible solution would have been to underline “a consensus” and then have “an agreement” as an option as a replacement. As test constructors discovered, the multiple choice section was the one where

items were most difficult to create. Read (2000) and Hughes (2003) agree: creating successful multiple choice items can be quite difficult for various reasons. Using the single-sex schooling context forced the test creators to a very limited context for sentences. The fact that all options needed to fit them stem also proved difficult, making this section the most laborious.

VII. Future Inquiries

After creating, revising, administering, assessing, analyzing a vocabulary test, and reading about vocabulary assessment in general, some additional questions have been raised. As Read and Chapelle (2001) mention, “the ultimate learning objectives of contemporary language teaching programs, especially those for learners with specific purposes, focus on communicative use of language in particular contexts” (p. 22). The context was clear in creating the exam, but what was test successful in helping students reach a realistic goal of obtaining knowledge of the vocabulary used in the context of single-sex schooling in the US? It might have been useful to readjust the test so that the content focused on education as a whole. Perhaps that is truly an ideal situation since the topic for the class was chosen and the test was to be created about that topic specifically.

Read (2000), examines the use of Cloze tests to incorporate context in a vocabulary assessment. Oller (1973, 1979) as cited in Read (2000) noted the highly effective strategy of testing students’ knowledge of second-language proficiency by their knowledge of vocabulary or grammatical items (p. 102). To create a test like this, more than just content words need to be tested, as well as the fact that the test needs to be in a discourse structure. In the future, it would be beneficial to create a cloze test that incorporates the words chosen specifically for a particular context, like same-sex schooling for the test created for this project. It would be interesting to see

how effective that type of exam would be compared to a test with three separate sections like this one.

Considering again the role of content-based vocabulary assessments and the idea of creating tests that incorporate context, as in the test created for the IPB class, the use of a test that measured production skills would be very successful, especially when considering the communicative approach to language teaching. Is it possible to create a test that successfully measures production skills as well as vocabulary knowledge specific to a particular context? How could that test be created to incorporate authenticity and validity particular to a content-based syllabus or course? These questions may be answered with the aide of future research and data collection.

References

- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Read, John. (2000). *Assessing vocabulary*. J. C. Anderson, & L. F. Bachman, (Eds.). Cambridge: Cambridge University Press.
- Read, J. & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing* 18(1), 1-32.
- Uçkun, B. (2008). How does context contribute to EFL learners' assessment of vocabulary gain. *The Asian EFL Journal* 10(2), 102-130.