Reading and Vocabulary Skills Assessment

Melinda Cuyul, Andrew Mobbs, Ciarra Proulx, and Alison Scherer

Northern Arizona University

Abstract

This report describes the test design and analysis project conducted for the authors' ENG 638

(Assessment for Second Language skills) course. This project entailed the construction,

administration, and results analysis (i.e., item analysis, B-index calculation, distracter analysis,

and reliability measurement) of the second Skills Assessment (SA) for the Reading and

Vocabulary Level 5 (RV5) course during the Fall 2014 semester in the Program in Intensive

English (PIE) at Northern Arizona University (NAU). During this process, investigated research

questions concerned the reliability of the test, the extent to which students would be prepared for

the summative Achievement Test (AT), and which course objectives (i.e., the subconstructs on

the SA) would require more targeted instruction. Reliability was primarily measured through

determining the KR-20 coefficient. The total scores of the SA were correlated with the total

scores of the AT that was administered two weeks later to determine the extent to which the SA

prepared students for the AT. Finally, item analysis and distracter analysis revealed the course

objectives on which these students could benefit from more instruction. The results that we

provide offer a more in-depth analysis of these issues.

Keywords: Reading, Vocabulary, Skills Assessment (SA), Achievement Test (AT)

Reading and Vocabulary Skills Assessment

Background

The purpose of this project was to develop and analyze a valid, reliable, formative Skills Assessment (SA) instrument for the Reading and Vocabulary Level 5 (RV5) course in the Program in Intensive English (PIE) at Northern Arizona University (NAU) to prepare students for the summative Achievement Test (AT) at the end of the semester. The SA was designed as part of the curriculum for RV5 at the PIE. The PIE teaches English for Academic Purposes with the goal of providing tertiary-level international students with academic readiness skills. Examining RV5 on a macro scale, RV5 is the fifth of six levels at the PIE, meaning it is a class taken by advanced-level students. The SA was designed into two main parts: reading and vocabulary. Each item in both parts was constructed reflect individual course objectives.

According to Norris (2014), assessment can be seen as a process, rather than a single instrument. To this end, the SA developed for RV5 was only the end result of a meticulous process in which many steps were taken and in which many people played a part. Additionally, the SA was situated within an assessment-development sequence in which the AT followed the SA. The assessment process often begins with the identification of course objectives or learning goals, then gauges the progress that test takers are making towards meeting those objectives, and concludes with a decision regarding how successfully those objectives were met (Miller, Linn, & Gronlund, 2013). To make the decision regarding to what extent test takers met objectives, the SA was developed as a formal, formative assessment. According to Brown and Abeywickrama (2010), a formal assessment is specifically planned to capture a sampling of students' attained knowledge and skills. With a formal assessment, student achievement may be measured. Part of the intended use of the SA was to help prepare students for the summative AT. To that end, the

SA was formative in nature because it was intended to be used as a step in the learning process, where part of the function of the assessment was to help teachers diagnosis learning gaps, and guide in the planning of curriculum until the end of the learning period (Fulcher, 2010). As a formative assessment, the SA was used to analyze learning progress as a result of instruction (Miller, et. al., 2013). The RV5 teaching team used the SA to understand their students' needs and successes. Finally, as part of a classroom-based assessment process, the SA was developed under the criterion-referenced test (CRT) paradigm. Students were individually rated on whether they successfully met each objective that was tested (Miller, et. al, 2013).

Research Questions

The following three research questions were addressed in this project: How reliable is the SA that was designed? How prepared will students be for the summative AT based on the results of the SA? Which course objectives addressed in the SA require more targeted attention from the classroom instructor?

Methods

Participants. This second SA was administered to four sections of Level 5 (i.e., advanced) international students in the PIE. Each of sections consisted of approximately 15 students (for a total of 64), all of whom were required to complete the assessment for a grade. Although all students were required to take the test, some students' data were not used, as per Institutional Review Board (IRB) guidelines. Only 49 students granted permission for their scores to be analyzed. Tests were administered to each section by their RV5 instructors.

Description of tasks. Five reading objectives and three vocabulary objectives were defined with a guiding table of specifications. For the reading portion of the test, the first objective was that of implied main ideas, where students were asked to determine the implied

main ideas of both the entire passage and of individual paragraphs of the adapted reading. The

second and third objectives, respectively, were identifying key details and making logical

inferences from the reading passage. The fourth objective, which was brand new, asked students

to recognize the author's point of view based on a specific segment of the passage. The fifth

reading objective was for students to be able to synthesize ideas between a passage from a text

that they have already read in class and a passage from the text provided for the assessment. For

the vocabulary portion, the first objective was that of semantic matching to test students'

knowledge of definitions. The second was paraphrasing, where students were asked to restate a

portion from the reading passage in their own words. The third vocabulary objective was word

usage in context in order to determine if students could choose the appropriate word given the

surrounding context of the sentence.

Test design. For all of the reading objectives, with the exception of identifying

supporting details, students were given multiple-choice questions with one correct answer and

two distractors. For the details section, students were provided with a prompt that portrayed a

key idea of the passage. Students then had space to fill in three details from the passage that

supported that idea, paraphrased into their own words. Regarding vocabulary, for the objective of

semantic matching, students were required to match the vocabulary word with the appropriate

definition. For the objective of paraphrasing, students were provided with a topic-related

sentence that contained a bold phrase for students to re-word while still maintaining the original

meaning. For the objective of word usage in context, students were presented with two

paragraphs with missing words and a word box above that contained a greater number of

possible words than there were blanks. Students could use the surrounding words to determine

the proper part of speech of the missing word as well as the gist of the sentence to decide which

vocabulary word would fit most appropriately. Each paragraph contained a word box with differing sets of vocabulary terms from which could choose.

Scoring procedures. Each of the four sections of Level 5 was taught by a different instructor, so each instructor was responsible for grading their own students' SA by hand. Instructors were allowed to give students half- and quarter-points at their discretion for items that involved a short answer. Total points were then added up and scored out of 34. As the majority of the test contained answers that would either be marked correct or incorrect, no rubric was needed or used. The score data were then entered into databases to document students' overall scores before being given to the researchers for analysis.

Results

Here is a brief summary of the results, beginning with the descriptive statistics of the SA:

Table 1

Descriptive Statistics for Level 5 Reading and Vocabulary Skills Assessment (Total Score=34)

Statistics	
k	32
N	49
Mean	25.05
Mode	27
Median	25.50
Min	9
Max	33
Midpoint	21
Range	24
SD (N>30)	5.18
Variance (N>30)	26.79
Skewness	-1.16
Kurtosis	1.78
KR-20 Coefficient	0.83
SEM	2.12

Item analysis revealed that the item facility of the items ranged from 0.3-0.9, and the item discrimination ranged from 0-0.56. Distracter analyses were also conducted on four of the most

difficult reading items and five of the most difficult vocabulary items. As indicated by the

Kurtosis reading, the results exhibit a very wide distribution of scores. The central tendency

measures (e.g., mean, mode, and median) and the skewness reading indicate that the test did not

appear to be extremely difficult.

Regarding the first research question, the KR-20 measurement for internal consistency

was approximately 0.83 with an SEM of approximately 2.12, which is comparatively reliable.

Regarding the second question, the SA adequately prepared students for the subsequent AT.

Although the SA had less items and contained less content than the AT, both tests were based on

the same course objectives, covered the same content, and included the same item types. The

mean score of the AT was 71.14 out of a possible 100 points, or 71.14%. This was similar to the

mean score of the SA, which was 25.05 out of a possible 34 points, or 73.68%. Regarding the

third question, based on item analysis, it was determined that the two main reading objectives

that merited further attention were the identification of implied main ideas and the identification

of the author's purpose; the IF for these items ranged from 0.3-0.6. The two main vocabulary

objectives that merited further attention were the ability to paraphrase and the ability to use

words in context; the IF for most of these items ranged from 0.4-0.7.

Relevance to PIE

If the SA was used to measure students' progress and to prepare them for the summative

AT, then these uses of the test call forth the concept of validity. A validity argument for the SA

can be made based on the facts that the SA was used to gauge student progress in understanding

and applying course content, and that appropriate interpretations were made from test scores to

predict student success on the summative AT. Furthermore, scores were used as inferences to

assist teachers in planning curriculum for the remainder of the course.

References

Alderson, J. C. (2000). *Assessing reading*. New York, NY: Cambridge University Press.

Bailey, K. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*.
    Boston, MA: Heinle & Heinle.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom*
    practices (2nd ed.). White Plains, NY: Pearson Education, Inc.

Carr, N. (2011). *Designing and analyzing language tests*. Oxford, UK: Oxford University Press.

Douglas, D. (2010). *Understanding language testing*. London, UK: Hodder Education.

Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.

Miller, D. M., Linn, R. L, & Gronlund, N. E. (2013). *Measurement and assessment in teaching*
    (11th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Norris, J. M. (2012). Purposeful language assessment: Selecting the right alternative test. *English*
    Teaching Forum, *39*(1), 41–45.