

| | |
|------------------|---|
| Title | Analysis of placement tests for the assessment of reading ability : text topics, test items and test takers' performance |
| Sub Title | プレイスメントテストによる英語読解力分析：トピック、テスト項目による受験者の反応分析 |
| Author | 中村, 優治(Nakamura, Yuji) Murray, Adam |
| Publisher | 慶應義塾大学日吉紀要刊行委員会 |
| Publication year | 2016 |
| Jtitle | 慶應義塾大学日吉紀要. 言語・文化・コミュニケーション (Language, culture and communication). No.47 (2016.) ,p.31- 38 |
| JaLC DOI | |
| Abstract | Locally based placement tests often focus on specific skills that a program deems important and which needs to be assessed. The purpose of the present study is two-fold : 1) to discuss important issues concerning second language reading assessment in academic settings, and 2) to analyze the reading section of an in-house university placement test. The two issues will be dealt with by examining the inhouse placement test results and several readability statistics. |
| Notes | |
| Genre | Departmental Bulletin Paper |
| URL | https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20160331-0031 |

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Analysis of Placement Tests for the Assessment of Reading Ability: Text Topics, Test Items and Test Takers' Performance

Yuji Nakamura
Adam Murray

Summary

Locally based placement tests often focus on specific skills that a program deems important and which needs to be assessed. The purpose of the present study is two-fold: 1) to discuss important issues concerning second language reading assessment in academic settings, and 2) to analyze the reading section of an in-house university placement test. The two issues will be dealt with by examining the in-house placement test results and several readability statistics.

I. Introduction

It is impossible to directly observe the act of reading. In fact, there is ongoing discussion about how the component subskills and strategies involved in the process of reading actually work together so that a person creates meaning. However, there is consensus among researchers that reading is an essential language skill, perhaps the most paramount, since so much of our information comes to use in written form. Students need to comprehend all written information required for their language development, as well as other skills, such as listening and speaking (Hubley, 2012).

When it comes to assessing reading, most assessment formats, including standardized tests and commercialized tests, can provide the foundation for placement and diagnostic assessment. However, the key point about assessments for placement and diagnostic purposes is that they are usually more locally driven (Grabe, 2009a). The class, program, and institute have specific placement needs and

for the diagnosis of their students' strengths and weaknesses. As Grabe (2009a) claims, locally based placement tests often focus on specific skills that a program deems important and which need to be assessed. For reading, these skills might include comprehension of academic material, discourse awareness of difficult texts, as well as vocabulary knowledge and grammar knowledge.

II. Purpose

1) To discuss important issues in second language reading assessment in academic settings, and 2) to analyze the reading section of an in-house university placement test. The two issues will be dealt with by examining the in-house placement test results and several readability statistics. These statistical measures include the Gunning-Fog Index, Flesch Reading Ease score, lexical density, and total word count.

III. Method

Subjects

Beginning in the spring semester of 2006, Keio University Faculty of Letters has administered an in-house placement exam for incoming freshmen and new sophomore students. Students are given placement tests twice a year, once at the beginning of the academic year and again at the end. For this analysis, the results of 10 exams (approximately 800 test-takers each) are examined.

Materials/Instruments

The placement test is a 60-minute examination which consists of 50 questions in four sections: grammar, vocabulary, gap-fill (cloze), and reading. For this study, only the reading section was examined. The reading section consists of 15 multiple choice questions (in three different categories A, B, C) with four options. The contents of 10 placement tests, 150 test items in total, were examined.

Analysis Procedures

In the IRT analysis, 110 test items were examined. The total number of test items was not 150 because each test form contains common items (anchor items) to apply common-item nonequivalent groups design using IRT analysis. The items were examined in terms of the Gunning-Fog Index, Flesch Reading Ease score, lexical density, and total word count as well as the content of each test item.

IV. Results and Discussion

Several readability statistics were recorded for each of the twelve reading passages of the placement tests. These include the Gunning-Fog Index, Flesch Reading Ease score, lexical density, and total word count. By comparing the readability statistics for each passage with student performance, it may be possible to determine which statistic is the best predictor of student performance.

Flesch Reading Ease

The Flesch Reading Ease (FRE) formula uses a 100-point scale to describe how difficult a text is likely to be. Texts between 90–100 are considered very easy, while those between 0–29 are extremely difficult or confusing.

Table 1: Flesch Reading Ease (FRE) Scores

| | | | |
|-----------|------|-----------|------|
| Passage B | 72.5 | Passage C | 46 |
| Passage L | 65.8 | Passage J | 44.7 |
| Passage K | 63.3 | Passage G | 40.4 |
| Passage I | 63.2 | Passage F | 27.1 |
| Passage P | 62.7 | Passage N | 11.1 |

Note: The passages with the highest and lowest scores.

Based on this scale, 7 of the passages received scores in the range 60–69, considered to be of “standard” difficulty. Thirteen passages were between 40–59 and can be considered to be fairly difficult. Passage N received a score of 11.1 and is considered extremely difficult.

Gunning-Fog Index (Table 1)

The Gunning-Fog Index is a widely used measurement that estimates the number of years of education (based on the U.S. system) that would be required in order to understand a passage when read for the first time. Based on this index, the reading passages ranged from 7.7 to 17.3.

Table 2: Gunning-Fog (G-F) Scores

| | | | |
|-----------|-----|-----------|------|
| Passage B | 6.7 | Passage R | 11 |
| Passage L | 7.7 | Passage U | 11.2 |
| Passage S | 7.9 | Passage J | 12 |
| Passage H | 7.9 | Passage F | 13.6 |
| Passage M | 8.6 | Passage N | 17.3 |

Note: The passages with the highest and lowest scores.

All but two of the passages have scores less than 12. It is expected that the passages that are higher on the index (around 9 and above) will be substantially more difficult than passages lower on the index. Passage N scored substantially higher on the index than the other passages with a score of 17.3 and is likely to be exceedingly difficult for test-takers.

Lexical density

Lexical density is used to determine how many unique words are used in a text. The formula for calculating Lexical Density is (Number of unique words / Total number of words) \times 100. More importantly, the proportion of content (lexical) words to total words is calculated. Texts with a lower density are more easily understood. As a guide, a lexically dense text has a lexical density of around 60-70% and one that is not dense has a lower lexical density score of around 40-50%.

Table 3: Lexical Density

| | | | |
|-----------|------|-----------|------|
| Passage I | 51.1 | Passage P | 59.4 |
| Passage F | 51.3 | Passage A | 59.5 |
| Passage H | 52.8 | Passage M | 60.3 |
| Passage R | 52.9 | Passage Q | 61 |
| Passage C | 54.1 | Passage U | 63.9 |

Note: The passages with the highest and lowest densities.

Half of the passages had a lexical density below 57%, while the remaining half had densities above that number. It is worth noting that Passage N has the highest G-F score and the lowest FRE score but is not the most lexically dense with a rating of 57.3%.

Passage length

There is considerable variation among the 22 passages in terms of word length. The shortest text is 324 words in total, whereas the longest text is 537 words. It is expected that test-takers will have greater difficulty with longer texts than with shorter texts.

Table 4: Passage Lengths

| | | | |
|-----------|-----|-----------|-----|
| Passage U | 324 | Passage P | 458 |
| Passage A | 343 | Passage Q | 462 |
| Passage M | 365 | Passage V | 479 |
| Passage O | 367 | Passage I | 493 |
| Passage J | 367 | Passage R | 537 |

Note: The passages with the highest and lowest word counts.

Test taker performance

Student performance was measured for each passage according to a “points” system. For comprehension questions in which 80% or more of test-takers responded with the correct answer, 3 points were assigned. For questions in which 60–79% answered correctly, 2 points were assigned. For questions in which 59% of students or less responded correctly, only 1 point was assigned. Each passage contains 5 comprehension questions, for a range of possible scores between 5 and 15.

Table 4a: Student Performance

| | | | |
|-----------|----|-----------|---|
| Passage B | 13 | Passage R | 8 |
| Passage H | 13 | Passage Q | 7 |
| Passage D | 12 | Passage T | 6 |
| Passage O | 11 | Passage C | 5 |
| Passage A | 11 | Passage N | 6 |

Note: The passages with the highest and lowest scores.

As a whole, the passages were of moderate difficulty, with scores of 11 points or less. Passages C and N were difficult and students scored 5 and 6 points respectively.

Table 4b: Student Performance in the order of difficulty using the point system

| | | | |
|------------------|----|------------------|---|
| Natural Sciences | 13 | Natural Sciences | 8 |
| Social Sciences | 13 | Humanities | 7 |
| Natural Sciences | 12 | Humanities | 6 |
| Humanities | 11 | Humanities | 5 |
| Humanities | 11 | Social Sciences | 5 |

Note: The passages with the highest and lowest scores.

The table shows that there are no specific differences in terms of topic difficulty among the three genres (humanities, social sciences and natural sciences). All three genres can be either difficult or easy.

As indicated above, there are both easy and difficult test items among three genres (humanities, social sciences, and natural sciences). In other words, the difficulty difference in each genre may not be caused by the genres but rather by the five questions of each test reading article. The way of presenting the test questions or choices, or the content of the test items will determine the difficulty. For example, in the genre of humanities, if the questions are about such abstract concepts as metaphors or similes, or paradoxical words, the difficulty level will be raised. In other words, the students find it difficult to get the correct answer, while the questions are straightforward even in the same genre of humanities; as a result, the difficulty level will be set lower than perhaps is practically accurate.

Furthermore, when the questions are about the implicit meaning of words or sentences, students find them rather difficult.

V. Conclusions and Implications

By comparing the readability statistics for each passage with student performance, it may be possible to determine which statistics are the best predictors of student performance.

One finding in this research was that the difficulty level of each test was almost at the level that the test writer intended. The order of the difficulty (easy to difficult) is close to the test writers' intention. Therefore, if the test writers are experienced and well informed of the test takers' ability and educational settings, they should choose appropriate topics, write reasonable test questions, and use valid test items without any difficulty. Classroom teachers could be one of the good predictors. In other words, general language proficiency tests, if the item writers are not familiar with the test takers or students in class, could provide the test items or questions which deviate from the students in reality.

Also, item difficulty should be taken into consideration when examining the test takers' performance.

Reading skills are broken down into sub skills, as suggested by Hubley.

Adjusted from Hubley's list of reading subskills (Hubley, 2012, pp.213–214)

| Level | Subskill | Typical way to test |
|----------------------------|---|---|
| Whole passage | Comprehend the main idea or gist meaning | What is the best title for this passage? What is the reading mainly about? |
| | Recognize author's attitude and biases | With which of the following statements would the author mostly closely agree? |
| | Distinguish fact from opinion | Which of the following is not a fact? |
| Section | Understand logical organization | Where would be the best place for this sentence? |
| Paragraph | Identify main idea and supporting details | List main ideas and match with paragraphs |
| | Understand topic sentences | The most important idea in Paragraph 4 is... |
| Across sentence boundaries | Understand the function of discourse markers | In line 22, what does "in addition" mean? |
| | Identify pronoun reference | What does "its" in line 6 refer to? |
| Sentence level | Guess the meaning of unknown words in context | Which word is closest to the meaning of "nomadic" in line 40? |
| | Paraphrase wording | Which phrase means the same as "brothers and sisters"? |

Furthermore, attention must be paid to the clear categorization of test items such as factual, inferential, discoursal, grammatical ability, or the ability to draw conclusions.

For future directions, the following six points can be taken into account (cf. Hubley 2012).

- a. The Internet has changed the availability of written material, the scope of sources, and even the ways in which people read.
- b. Learning to recognize sites and sources which are useful has become an essential life tool, and as a result skimming and scanning are no longer primarily classroom strategies.
- c. In addition other reading subskills, such as identifying key words and concepts, and separating fact from opinion are now skills found in everyday life, not just the classroom.
- d. Written texts are now often accompanied by graphs, maps, photographs, and video. This requires students to be able to comprehend information across multiple media.
- e. An international language, English is becoming highly non-standardized and readers now encounter a variety of abbreviated and other forms online.
- f. The challenge for instructors in the future is to assess relevant skills for reading today and their means of assessment.

Note:

An earlier version of this paper was presented at the 19th Conference of Pan-Pacific Association of Applied Linguistics (PAAL) in Tokyo on August 11, 2014.

Acknowledgment:

This work was supported by the Grant-in-Aid for Scientific Research (KAKENHI) (C) (24520646).

References

- Grabe, W. (2009a). *Reading in a Second Language: Moving from Theory to Practice*. Cambridge, USA: Cambridge University Press.

- Grabe, W. (2009b). Teaching and Testing Reading. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp.441–459). West Sussex, UK: John Wiley & Sons.
- Hubley, J. N. (2012). Assessing Reading. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge Guide to Second Language Assessment*. (pp.211–217). Cambridge: Cambridge University Press.