

Developing Reading Test for the First Semester Students of English Education Department at University of HKBP Nommensen Pematangsiantar

DOI: <https://doi.org/10.47175/rielsj.v4i4.836>

| Irene Adryani Nababan¹ | Wiwiek Meisuri² | Anita Sitanggang³ |

¹ English Education Department,
University of HKBP Nommensen
Pematangsiantar, Pematang
Siantar, Indonesia

² Postgraduate Program,
Universitas Negeri Medan,
Medan, Indonesia

³ English Education Department,
University of HKBP Nommensen
Pematangsiantar, Pematang
Siantar, Indonesia

1ireneppbiuhnp@gmail.com

2meisuri@yahoo.com

3anitasinggang2019@gmail.com



This work is licensed
under a Creative Commons Attribution-
ShareAlike 4.0 International License.

ABSTRACT

The significance of reading comprehension in lifelong learning cannot be overstated. It serves as an interactive process that bridges the gap between the reader and the text. Irrespective of the academic level or subject, students must possess strong reading comprehension skills. The assessment and evaluation of students' reading comprehension abilities are integral components of testing. The tests administered to measure students' achievements should strive to provide accurate and reliable information. The purpose of this research is to create a practical, valid, and reliable reading comprehension exam for use in the English Education Department of the Faculty of Teacher Training and Education at the University of HKBP Nommensen Pematangsiantar. This research uses a quantitative method and carries out at the first semester students, academic year 2023/2024. In the development of reading tests, the preferred format utilized is the multiple choice test. There are 25 items. Relatively speaking that the 9,38 score is relatively high. It demonstrates how the exam appears to be able to distinguish between the test subjects. The reliability of this multiple choice reading comprehension test can be inferred from the KR-21 scores, which stand at 58%. The test's evaluation and the field trial's outcome confirm that the test's results are accurate, dependable, and useful in terms of content.

KEYWORDS

developing; test; reading comprehension; valid; reliable; practical

INTRODUCTION

An essential part of the teaching and learning process is the test. A test's ability to gauge the effectiveness of the teaching-learning process is one of its purposes. For this reason, a test and the teaching-learning process are inextricably linked. Notwithstanding the significance of assessments, it is still common for teachers to create exams without taking test development standards into account. This is the reason we frequently come across tests that are neither valid nor reliable because they are unable to consistently measure student achievement or to significantly measure what needs to be measured. The institution needs to be improved because the situation frequently occurs in the English Education Department of the Faculty of Teacher Training and Education at the University of HKBP Nommensen Pematangsiantar.

Enhancing the current syllabus more organized and in line with the goals of the English Education Department of Faculty of Teacher Training and Education at at the University of HKBP Nommensen Pematangsiantar's is one improvement that needs to be prioritized. As a component of the curriculum, assessment is the pivotal point to be developed in order to reach the stated educational objectives. To achieve this goal, it is therefore essential that

creating the evaluation form be one of the prerequisites. Listening, Speaking, Writing and Reading are the four most important courses, according to the syllabus of the English Education Department of Faculty of Teacher Training and Education at University of HKBP Nommensen Pematangsiantar. One of the deciding courses appears to be reading comprehension. It suggests that the reading comprehension component of the curriculum ought to take up more space than the other abilities or elements.

Reading comprehension courses, also known as Literal Reading, Intensive Reading, Extensive Reading, and Critical Reading, are offered in four semesters by the English Education Department of Faculty of Teacher Training and Education of the University of HKBP Nommensen Pematangsiantar (Kurikulum KKNi 4.0 PPBI). Literal Reading is the one of the four reading courses mentioned above that has been selected to be the topic of the subject of test development that will be administered in the first semester. The reading comprehension tests created by certain researchers largely failed to meet the standards of a valid test, according to the preliminary investigation the researcher conducted while she was a teaching assistant in this department. These details provide the researcher with motivation to design a thorough investigation to create an effective reading comprehension assessment.

The research issue addressed in this study is based on the background information provided "How to develop a test that measures the reading comprehension of the English Education Department students of the Faculty of Teacher Training and Education at the University of HKBP Nommensen Pematangsiantar in a way that is valid, reliable, and practical?"

RESEARCH METHODS

In this research, quantitative research methods were employed to construct an assessment tool capable of evaluating an individual's reading comprehension skills of first semester students of the English Education Department at the University of HKBP Nommensen Pematangsiantar during the academic year 2023/2024. The research focused on a population of 74 students, with a sample size of 32 students. The reading comprehension scores were presented in the form of the 25 items of multiple-choice test. The data collected from this research were used to evaluate the validity of the test in the English Education Department of the Faculty of Teaching Training and Education at the University of HKBP Nommensen Pematangsiantar in 2023.

RESULTS AND DISCUSSION

The reading test uses a multiple-choice test format, including 25 questions. The test items comprehensively encompass all skills relevant to reading comprehension, as indicated in Table 1.

Table 1. Specification for Test Items and Skills

Skills of Reading Comprehension	Level of Comprehension Domain	Number of Items	Test Item Number
Micro Skills			
1. Comprehending lexical as well as grammatical coherence	Literal	5 Items	2,9,13,17,22
2. Recognizing referents like comparison, aim, cause, and result.	Literal and Inferential	2 Items	4,11
3. Recognizing the sentence and clause structures	Literal and Inferential	1 Items	19
4. Recognizing discourse markers	Literal and Inferential	2 Items	8,25

Macro Skills			
1. Skimming : rapidly accessing a book, determining which section—or if section is relevant in its entirety—is necessary to establish	Literal and Inferential	5 Items	3,5,15,20,23
2. Scanning : deliberate act of attempting to identify particular information within a written piece.	Literal and Inferential	3 Items	1,6,24
3. Surveying for gist	Literal	3 Items	7,10,12
4. Recognizing key concepts & specifics	Literal	2 Items	14,16
5. Recognizing opinions and facts, as well as attitudes toward readers and subjects	Evaluative	2 Items	18,21
Total		25 items	

1. Initial Try-out Data

In this phase, the researcher acquired data from four lecturers who are affiliated with the English Education Department of the Faculty of Teaching Training and Education at the University of HKBP Nommensen Pematangsiantar. Some things should be taken into consideration based on their questionnaire responses. First, they have to do with the text's length; they believe it to be "enough.". Second, they believe that revisions to the instruction are necessary for practical reasons, which has to do with how clear it is. Third, they believe that the general information needs to be modified in order to accurately reflect the test taker's identity. Fourth, the test example has to be revised with new options. Fifth, each text should include a reference to the original source at the end. The test's allotted period of 100 minutes has been altered in light of the feedback. 90 minutes is the new allotment of time. The guidelines for the multiple-choice format encompass the essential steps of providing one's name on the allocated area of the answer sheet, thoroughly examining the texts provided and responding to each question appropriately by marking the respective column on the answer sheet that corresponds to the options of A, B, C, or D, representing the correct input. They also include advice on how to make the most use of the time. An example of how to answer the questions is provided to make the advice more useful.

2. Field Try-out Data

Following a first draft revision, the researcher conducted a field trial with 32 first-semester students of the English Education Department of Faculty of Teacher Training and Education at the University of HKBP Nommensen Pematangsiantar. For two reasons, it is believed that all of the pupils are equal. First of all, the department treats them equally academically as freshmen. Second, they both work in the English Education Department, which is the same department. The tryout results indicate that this result's mean score is 62,28. 65.12 is the median. 66.12 is the mode. The mean, mode, and median in this instance closely match one another. It demonstrates how the scores often approach a center value.

3. Validity

The test maker employments consistent prove to appear the legitimacy of the test since it is troublesome for him to show experimental confirmation within the shape of scores from another authentic test. Construct and content validity are therefore applied. The degree to which test results align with hypotheses derived from an ability theory, or construct, is known as construct validity (Bachman, 1990).

$$ID = \frac{\text{High group\#correct} - \text{low group\#correct}}{\frac{1}{2} \times \text{total of two comparison group}}$$

The provided data in table 1 portrays the comprehensive coverage of all the skills encompassed within the construct of reading comprehension. These skills will be further elucidated in the subsequent discussion. Understanding lexical and/or grammatical coherence is represented by 20% of the items, referent comprehension is represented by 10%, phrase and clause syntactic structure is represented by 5%, and discourse markers are represented by 15%. Since those particulars pertain to micro skills, literal and inferential comprehension levels are covered. Here are the specifics for macro skills: Skimming is represented by 13% of the items, scanning by 10%, surveying for the gist by 8%, grasping the major concepts and crucial information by 15%, and separating opinion from truth and identifying attitudes toward readers and subjects by 8%. Every degree of comprehension in these macroskills.

4. Reliability

In this research, reliability is established through the use of internal consistency reliability. The current study decided to determine reliability by looking at the consistency within the instrument for two reasons. First, because this study was conducted in a short amount of time, test-retest reliability would take longer to establish. Second, it appears that creating the equivalent form for determining reliability will also take some time. As a result, the test developer employed internal consistency for this investigation. As Kuder-Richardson 21 (K-R21) serves as a valuable instrument for evaluating the dependability of multiple-choice examinations, the researcher incorporated it into the test development procedure. 25 items of multiple choice, with a mean score of 18.32 and a standard deviation of 3.23. These results are based on experimental test evaluation results. Considering the dependability estimate's outcome The dependability coefficient, as determined by KR-21, is 0.58. In this instance, we may state that 58% of the scores are dependable or consistent.

5. Item Facility

According to research, the optimal range for the IF index is 0.20 to 0.80 (Djiwandono, 1996, p. 141). This is the IF calculation method for multiple choice questions borrowed from Brown. (2005, p. 68). The results show that a certain amount of goods are of poor quality. Item numbers 2 (0.6), 11 (0.8), 17 (0.7), and 21 (0.3) are below average. So, these

revision of items is necessary in order to comprehend the underlying causes of these occurrences and determine the appropriate methods for rectifying them. Consequently, a thorough analysis of distractors is performed as a means to achieve this objective.

6. Discrimination Analysis Items

Referred to as ID, this analysis is the measure of how effectively an item differentiates between test takers who perform well and those who perform poorly. The ID index values range from 1.0 to 0, with a value close to 1.0 indicating a high level of discriminating power and a value of zero indicating a low level. In this study, Ebel's (1979) categorization of ID was utilized :

- 0.40 and up ----- very good item
- 0.30 – 0.39 ----- reasonably good item
- 0.20 – 0.29 ----- marginal item

below 0.19 ----- poor item

The formula for computing the ID in multiple-choice questions, which was obtained from Brown (2004, p. 59), can be expressed in the following manner :

Using this formula, the result of current research indicates that the multiple-choice test items numbered 3, 8, 19, 22, and 25 have ID values less than 0.3 and are thus categorized as bad items; the other items, on the other hand, have ID values greater than 0.3 and are therefore categorized as excellent things. Items with ID values less than 0.30 should be altered to raise them over 0.30 so that they can better distinguish between the performance of the upper and lower groups in the test. Distracter analysis is carried out in order to determine why these occur and how to modify the items.

7. Distractor Analysis

The study's distracter efficiency is determined by analyzing the percentages of tests that choose each option. This analysis proves to be of great significance when it is able to present the proportion of each alternative selected in both the higher and lower cohorts (Brown, 1996). Compared to the lower group, the upper group should answer a higher frequency of accurate responses. It should also be less common for the upper group to select the distractions compared to the lower group. If one group answers a test question more correctly than another, or if one group selects distracters less frequently than the other, or if the distracters themselves do not elicit any responses, then the original question should be modified.

Tabulating the alternate answers provided will allow you to assess how well the distracter can divert test takers, it is found that item number 2, 5, 6, 8, 9, 11, 14, 15, 16, 17, 18, 19, 20, and 21 need to be changed because nobody is drawn to the distractions. In addition, as the distracters draw in the top group more, items no. 1, 7, 12, 23, and 25 also need to be changed.

8. Practicality

Practicality standards often center on aspects like economy, scorability, and administrability. A test in finance should be efficient in relation to the economy component. The test must to be affordable and reusable. Given that it is just seven pages long and can be reused, this test satisfies the requirements. In relation to the test's scorability, its short turnaround time makes it simple to check. Anyone may verify the exam, which is a multiple-choice question with the only way to know the answer is to cross out options a, b, c, and d. In terms of administerability, the exam is adequate because it is simple to administer and store for the students. Thus, it may be said that the exam is useful since it.

CONCLUSION

Previous studies have indicated that numerous students engaged in research are required to construct their own instruments for their study. This investigation is driven by this discovery. Similarities to this matter can be observed in specific reading comprehension research. No functional tool is currently accessible. The commercial reading comprehension tests that are currently on the market are designed to cater to a broad spectrum of test-takers' needs. They cannot achieve particular research goals because those goals require particular environments and subjects. With the help of this research, a reading comprehension test for the English Education Department students of Faculty of Teacher Training and Education at the University of HKBP Nommensen Pematangsiantar

will be created. These exam questions are derived from 501 Critical Reading Tasks. Test comes about are regarded substantial in terms of substance, tried and true, and valuable based on assessment of the test and field tryout comes about.

REFERENCES

- Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Hong Kong: Oxford University Press, Ltd.
- Brown, Douglas. H. (2004). *Language Assessment: Principles and Classroom Practice*. San Francisco State University: Pearson Education Inc.
- Brown, J.D. (1996). *Testing in Language Programs*. Upper Saddle River: Prentice Hall Regent
- Brown, James Dean. (2005). *Testing in Language Programs: A Comprehensive Guide To English Language Assessment*. New York: The McGraw-Hill Company
- Djiwandono, S. (1996). *Tes Bahasa dalam Pengajaran*. Bandung: Penerbit ITB Bandung
- Ebel. RL. (1979). 3rd Ed. *Essential of Educational Measurement*. Englewood, New Jersey : Prentice Hall.
- Kurikulum KKNi 4.0 PPBI Universitas HKBP Nommensen Pematangsiantar