DEVELOPING AND VALIDATING EFL READING TEST

Anugrah Novendi Perkasa
SMAK Santo Albertus Malang, Indonesia
anugrahnovendi@hotmail.com
Indonesia

ABSTRACT
EFL students need to learn about reading comprehension because it is one of the essential abilities in learning English. To measure the students reading comprehension, a reading test is needed. Nevertheless, many students find difficulties in working the reading test. It happens because the reading tests which are given by teachers who do not analyze item difficulty, discrimination, distractor, and reliability. Those analyses are needed to prove that the reading test has met criteria of good tests. Criteria of good tests are needed in making the reading test because the reading test is used for measuring the achievement of EFL students. This study purposes at developing the reading test based on Indonesia National Curriculum basic competence, learning indicators, English lessons, and it examines to meet the criteria of good test. The researcher adapted Borg and Gall’s (1983) theory in developing the reading test.

Keywords: reading test, reading comprehension, learning, EFL

1. Research Background
EFL students need to master reading skill because they are expected to understand and comprehend the English texts. Students need to comprehend about reading comprehension because reading is needed to get information from written English discourses (Badan Nasional Standar Pendidikan., 2006). Notably, senior high school students need to master reading skills because they are expected to understand and comprehend written English texts (Emaliana, I., 2013a) Thus, reading is one of the important skills in learning English since it can determine the students’ achievement in understanding the written English discourses (Emaliana, I., 2013b); (Emaliana, I., 2011); (Puspitasari, I., Emaliana, I. Lailiyah, N, Lintangsari, A. P., Soewarso, P.N., 2019) Whereas, challenges in teaching reading comprehension are comprised of three domains such as students unequal linguistic competence, students’ critical thinking and also their motivation (Guntur & Rahimi, 2019). Besides, the reading comprehension assessment was dominantly dependent upon tools that were focus on narrow of skills. Whereas, the needs of the broader understanding beyond the comprehension are necessary to boost students reading comprehension. Providing reading inventories can supplement standardized test to widen the focus of reading comprehension skills, however, a consistent assessment framework is required to maintain the consistency but also dynamic and ongoing that let the readers making choices and activate the students metacognitive decisions (Woolley, 2008).

As aforementioned challenges found in reading comprehension teaching practices, providing a reliable, valid, yet dynamic and an ongoing reading test based on consistent framework is necessary to facilitate students in reading comprehension. A reading test is created based on basic competence and learning indicators which need to be examined by proving that the test meets the criteria of a good reading
The criteria of a good test are needed because the reading test is used for measuring students’ achievement. A good reading test is created by considering the criteria of the language test. There are five criteria for making a good test. The first criterion is the validity, it means that the reading test should be valid to which the reading test can be measured what it should be measured (Heaton, J.B., 1990) validity is conclusions made from the results of the test are appropriate, meaningful, and useful in terms of the purpose of the assessment. The test measuring what should be measured. The validity of a test is the extent, to which it exactly measures what it is supposed to measure (Heaton, J.B., 1990) A test has to aim to provide the actual size of a particular skill intended to measure not to the extent of measuring external knowledge and other skills at the same time.

The second criterion is the reliability, it that means the reading test should be reliable as a measuring instrument by verifying the consistency of the test results (Hughes, 2003). Reliability means the ability of assessment tools to produce stable and consistent results (Hughes, 2003). A test can be relied on if the teacher gives the same test to the same student on two different chances, the test must produce similar results. Therefore, reliability refers to the consistency of scores obtained by students. The third criterion is the discrimination, it means that the reading test can discriminate the students’ performance (Heaton, J.B., 1990). (Heaton, J.B., 1990) stated that the test must have the capacity to discriminate individual understanding in the learning group. (Heaton, J.B., 1990) divided the difficulty of a test into ten levels: extremely easy items, very easy items, fairly easy items, items below average difficulty level, items of average difficulty level, items with above average difficulty level, fairly difficult items, difficult items, very difficult items, and extremely difficult items.

The fourth criterion is the practicality, it means that the reading test should have practical administration and efficiency test (Brown, H.D., 2004) Practicality is an effective practical test. This means that the test is not excessively expensive, stays within an appropriate time constraint, is relatively easy to administer, has a scoring or evaluation procedure that is specific and has time-efficient (Brown, H.D., 2004) Furthermore, details of the test administration must be clearly defined before the test. Next, all materials and equipment for the test must be ready. Moreover, test fees must be within the budgeted limit. Next, the evaluation or evaluation system must be appropriate within the teacher’s time frame. Students must be able to complete the test naturally within the specified time frame. Practicality can be simply defined as the relationship between available resources for the test, human resources, material resources, time, and resources which will be required in the design, development, and the use of the test. The last criterion is the authenticity, it means that the reading test should be authentic (Brown, H.D., 2004) authenticity is the degree of correspondence of the characteristics of a given language test task to the features of a target language task (Brown, H.D., 2004) Authenticity of a test may be present in the following ways: the language in the test as natural as possible; items are contextualized rather than isolated; topics are meaningful (relevant, interesting) to the learners; some thematic organization to items is provided, such as through a story or episode; and tasks represent, or closely approximate, real-world tasks.

In reality, based on the need analysis conducted by the researcher by interviewing English teacher in SMAK Frateran Malang, the main problem was the process of making the reading test. In making the reading test, the teacher has not been analyzed the item difficulty, item discrimination, effective distractor and item reliability of the reading test because of the time limitation. However, analyzing the difficulty, discrimination, effective distractor, and reliability of the reading test is needed because the reading test is used for measuring achievement for grade ten. Considering existing problem having no standardized English reading test for the tenth grade students, the researcher designed and developed the reading test that emphasizes on the criteria of a good reading test. The researcher adopted the theory from (Heaton, J.B., 1990), Hughes (2003), and (Brown, H.D., 2004) as the guideline in writing the reading test, (Borg, W.R. & Gall, J.P., 2003) as the research method, and (Emaliana, I., Tyas, P.A., Widyantingsih, G.E.N., Khotimah, S.K., 2019) and (Djiwandono, M.S. 1996), as the guidelines in evaluating the item of the reading test.
2. Methods

Participants
Participants were 27 students from some senior high school in Malang. The students were 15 or 16 years old. The students at the intermediate level in TOEFL score (425-450).

Procedure
In the procedure of Research and Development, the researcher applied the cycles proposed by (Borg, W.R. & Gall, J.P., 2003) which are: (1) Need analysis, (2) Planning, (3) Develop preliminary form of product, (4) Preliminary field testing, (5) Main product revision, (6) Main field testing, (7) Operational product revision, (8) Operational field testing, (9) Final product revision, and (10) Dissemination and implementation. However, due to a limited time, this research only use nine cycles in conducting the research. The steps that applied in this research are explained by figure 1.

Figure 1. Cycle of Research and Development, adapted from (Borg, W.R. & Gall, J.P., 2003)

3. Results

Need Analysis
Need analysis as the starting point of research and development is done in order to find out the discrepancy between the factual condition and a desired set of condition (Borg, W.R. & Gall, J.P., 2003) This step is profoundly prominent since the result becomes the foundation of the whole research. To collect the information needed in research and development design, there are numerous ways that can be chosen, such as: interview, questionnaires, observation, data collection, informal consultation with practitioner or expert, and others. In this research, the data from need analysis were obtained by applying interview.

Interview with the teacher was intended to gather information about the current reading test and the expected product of the reading test. The interview was conducted to the English teacher by using the interview guideline in collecting the data. From the result of conducting interview, the researcher had the concept in making the plan to write the reading test.

Planning
The researcher planned to make the reading test specification and blueprint. First, the researcher identified the objective of the reading test. Identifying the objectives of the course is important to do to ensure the validity content of the test. The test should meet the objectives of the course. Therefore, the researcher have to identify the objectives and then determine the objective of the test, the micro skills tests, and type of the texts. The objectives of the course for the students were designed to improve the students’ reading skills. The specific instructional objectives evaluate students’ understanding of the topic, main ideas, and word meanings. Second, the researcher made the types of tests, number of texts in the tests, number of items, time for the tests, equipment involved, and scoring method of the reading test. The type of test is a
A multiple choice test consists of 8 recount texts and 20 items. The texts consist of 110 to 272 words. The needed time to work the test is 30 minutes in which students worked individually. In scoring, the correct answer gets a score of 1, while the wrong answer gets a score of 0. Last, the researcher designed the blueprint and made indicators based on the Indonesia National Curriculum basic competence number 3.9 and 4.13 curriculum 2013 that are suitable for the needs of the students.

**Writing the Test**

After making the test specification and blueprint of the reading test, the next step was writing the reading test. The researcher wrote the reading test covering writing the instruction and the items. The instructions have to be clear to ensure that the students are not confused in reading the instructions. The question items are based on reading micro skills. The distribution of test item was as follows. There are fourteen implied detail questions. Furthermore, there are five words meaning questions. Last, there is one main idea question.

**Expert Validation**

The validation on reading test was done by one lecturer in English Language Teaching Program on Universitas Brawijaya. There are ten checklist to validate the reading test. The checklist consists of the construct of the texts, the content of the texts, the length of the texts, the difficulty of the texts, the number of items, the number of distractors, the micro skills, the questions, the directions, and the time allotment. She approved that the reading test is appropriate to be used with minor revision on word choices and grammar in the sentence.

**First Revision**

The researcher received feedbacks whether the reading test needed to revise or not. The main purpose of the revising is to improve the reading test before implementing it to the students. Based on that evaluation, the researcher revised the word choices and grammar in the sentences in the reading test. After the revision, the reading test is ready to be used for the students.

**Main Field Testing**

Due to the limited amount of time, tryout was once in main field testing. In the opening the activity, the researcher informed the students the purpose of the conducting the try-out. In the tryout activity, the researcher asked the students to work the reading test. They were asked to take a pen and keep the English text book and dictionary. The students were permitted to ask question if they met a problem in working the reading test. At the end of the activity, the students were asked to collect the reading test.

**Analyzing the Results of the Main Field Testing**

**Item Difficulty Analysis**

The researcher used the classification from (Djiwandono, M.S. (1996)., 1996) to find the difficulty level of the reading test. The classification is as follows:

<table>
<thead>
<tr>
<th>Classification</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000 – 0.250</td>
<td>Difficult</td>
</tr>
<tr>
<td>0.251 – 0.750</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.751 – 1.000</td>
<td>Easy</td>
</tr>
</tbody>
</table>

Based on the Item and Test Analysis Program (ITEMAN) version 4.3 results, the results of item difficulty analysis can be seen on this table.
From the table above, we can see that there are six items belong to the difficult item. They are the items number 4, 5, 6, 15, 16 and 18. The range is from 0.000 to 0.250. Furthermore, twelve items belong to the moderate item. They are the items number 1, 2, 3, 7, 8, 9, 10, 11, 13, 14, 17 and 20. The ranges is from 0.251 to 0.750. The other, two items that belong to the easy item. They are the items number12 and 19. The range is from 0.751 to 1.000.

**Item Discrimination Analysis**

The discrimination of the test items tells how well the items in test perform in separating the higher group and the lower group. The researcher used the classification from (Djiwandono, M.S. (1996)., 1996) to get the discrimination score of the reading test. The classification is as follows:

<table>
<thead>
<tr>
<th>Classification</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>D ≤ 0.1999</td>
<td>Very low</td>
</tr>
<tr>
<td>0.200 – 0.299</td>
<td>Low</td>
</tr>
<tr>
<td>0.300 – 0.399</td>
<td>Moderate</td>
</tr>
<tr>
<td>D ≥ 4.000</td>
<td>High</td>
</tr>
</tbody>
</table>

Based on the Item and Test Analysis Program (ITEMAN) version 4.3 results, the results of item discrimination analysis can be seen on this table.

**Table 4. Results of the Item Discrimination Analysis**
From the table above, we can see that there are seven items belong to the very low item. They are the items number 1, 2, 4, 5, 7, 8 and 19. The score is lower than 0.1999. Next, five items belong to the low item. They are the items number 3, 9, 10, 17 and 18. The range is from 0.200 to 0.299. Furthermore, three items belong to the moderate item. They are items number 6, 15 and 16. The range is from 0.300 to 0.399. The other five items belong to the high item. They are items number 11, 12, 13, 14 and 20. The score is higher than 0.400.

**Effective Distractors Analysis**

The data show that ten items of the reading test have effective distractors and ten other items have ineffective distractors. The researcher used the classification from (Djiwandono, M.S. (1996), to get the distractor score of the reading test. The classification is as follows:

<table>
<thead>
<tr>
<th>Classification</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000 – 0.010</td>
<td>Poor</td>
</tr>
<tr>
<td>0.011 – 0.050</td>
<td>Fair</td>
</tr>
<tr>
<td>0.051 – 1.000</td>
<td>Good</td>
</tr>
</tbody>
</table>

Based on the Item and Test Analysis Program (ITEMAN) version 4.3 results, the results of distractors analysis we can see that not all the options have effective distractors. The items number 1, 4, 5, 7, 8, 10, 15, 16, 18 and 20 or ten items have effective distractors and the items number 2, 3, 6, 9, 11, 12, 13, 14, 17 and 19 or ten items have ineffective distractors

**Reliability Analysis**

The researcher used the classification from (Djiwandono, M.S. (1996), 1996) to get the reliability score of the reading test. The classification is as follows:

<table>
<thead>
<tr>
<th>Classification</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000 – 0.400</td>
<td>Low</td>
</tr>
<tr>
<td>0.401 – 0.700</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.701 – 1.000</td>
<td>High</td>
</tr>
</tbody>
</table>
Based on the Item and Test Analysis Program (ITEMAN) version 4.3 results, it is found that the score of reliability of the reading test is 0.603. The range score is from 0.401 to 0.700, this means that the score is moderate. Based on the score it can be concluded that the reading test is reliable.

Final Revision

The researcher revised the test items based on the results of the item analysis because of the item difficulty, item discrimination and distractor did not meet the required standard. Based on the item analysis results, there are three items that can be used directly, but fourteen others have to be revised, and three other items replaced with by items. Revised items are the items that have a low level of discrimination and distractors who have poor qualifications. For the level of difficulty is not revised because it has met the curve of the test difficulty. The revised items are explained as follows: first, the item number 1 has to be revised in term of the questions. Second, the item number 2 has to be revised in term of the questions and option E. Third, the item number 3 has to be revised in term of the questions and option C. Fourth, the item number 4 is replaced by new item. Fifth, the item number 5 is replaced by new item. Sixth, the item number 6 has to be revised in term of the questions and option D. Seventh, the item number 7 is replaced by new item. Eighth, the item number 8 has to be revised in term of the questions. Ninth, the item number 9 has to be revised in of term the questions and option D and E. Tenth, the item number 10 has to be revised in term of the questions. Eleventh, the item number 11 has to be revised in term of the option E. Twelfth, the item number 12 has to be revised in term of the option D. Thirteenth, the item number 13 has to be revised in term of the option D and E. Fourteenth, the item number 14 has to be revised in term of the option E. Fifteenth, the item number 15 can be used directly. Sixteenth, the item number 16 can be used directly. Seventeenth, the item number 17 has to be revised in term of the questions option C. Eighteenth, the item number 18 has to be revised in term of the questions. Nineteenth, the item number 19 has to be revised in term of the questions and option E. Last, the item number 20 can be used directly.

Final Reading Test

After going through and completing all development processes, those produced the final reading test. The final reading test consisted of 20 items that focused on implied details. Validity, reliability, discrimination, practicality, and authenticity (Heaton, 1990; Hughes, 2003; and Brown, 2004). Moreover, the final reading test has been evaluated based on theories from Emaliana et al. (2019) and Djiwandono (1996). It can be concluded that the final reading test as the product of this research is ready to be applied for testing English reading to EFL students.

4. Discussion

Generally, the reading test items development and validation appears to be feasible. Novice reading test items writer will find the development and validation procedures intuitive and easy to use. By starting deciding test specification, especially test blueprint, finding it useful for structuring the writer’s plan for developing the reading test. The test writer produced 35 items in roughly 3 days of at-home-working time, plus 6 hours of meeting review time. After the validation, some of the items were not usable, therefore, the final test consists of 20 items. The results of this reading test items development and validation are encouraging, although they are limited by several factors. First, the try out was done once only after the test items were made and consulted to the expert for language assessment due to limited time and permission grant from the school. Second, the test developed is for formative test than summative test, therefore, future researchers need to consider this for their future research. Finally, judgments about the quality of the selected test items were made by only one time measurement because the try out was only done once.

Cite this as:
5. Conclusion

As a Research and Development design, this study develops a reading test for EFL students. The reading test was designed through stages of development proposed by Borg, W.R. & Gall, J.P., (2003). The stages are conducting a need analysis, planning, developing a preliminary form of a product, preliminary field testing, main product revision, main field testing, operational product revision, and final product.

The reading test is designed by referring to the Indonesia National Curriculum basic competence number 3.9 and 4.13 curriculum 2013. Furthermore, the reading test is created by considering the criteria of the language test. They are validity, reliability, discrimination, practicality, and authenticity (Heaton, J.B., 1990); Hughes, 2003; and (Brown, H.D., 2004). Moreover, the reading test has been evaluated based on theories (Emaliana, I., Tyas, P.A., Widyaningsih, G.E.N., Khotimah, S.K., 2019) and (Djiwandono, M.S., 1996). It can be concluded that the reading test as the product of this research is ready to be applied for testing English reading for EFL students.

After becoming the final reading test, the reading test is ready to be applied for testing reading comprehension. For the teacher, this reading test is expected to help the teacher to find out the reading test that is appropriate with the standard reading test to be applied in testing reading comprehension. For the further development of the reading test, hopefully it can be developed for other materials, with other various techniques, and to other schools. The material can be changed related to the different competences that will be learned based on the syllabus.

Acknowledgement

I wish to thank to Dr. I've Emaliana for her valuable and constructive suggestions during the planning and development of this research work. Her willingness to give her time so generously has been very much appreciated.

6. References

Appendix 1. Sample of the Test Blueprint

This blueprint is adapted from *kisi-kisi soal Bahasa Inggris kelas X kurikulum 2013 Blue Print*

**Kompetensi Standar:**

3.9 Menganalisis fungsi sosial, struktur teks, dan unsur kebahasaaan pada teks recount sederhana tentang pengalaman/kejadian/peristiwa, sesuai dengan konteks penggunaannya.

4.13 Menangkap makna dalam teks recount lisan dan tulis sederhana.

**Materi:** Recount Text

**Indikator Soal:**

Menentukan informasi rinci tersurat dari teks berbentuk recount Mengidentifikasi main idea dari sebuah teks recount Menentukan makna kata dari teks berbentuk recount Mengidentifikasi kejadian dalam teks recount yang dibaca

**Reading Micro Skill:** Implied details, Main Idea, Word meaning

**Bentuk Soal:** Multiple Choice

**Jumlah Soal:** 20 Soal

---

<table>
<thead>
<tr>
<th>No. Soal</th>
<th>Rumusan Butir Soal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Messi Biography</td>
</tr>
</tbody>
</table>

Luis Lionel Andres Messi, born June 24th, 1987, is an Argentinian football player for F.C Barcelona. He is not very tall, mainly, due to the growing problem he had when he was younger. His eyes are brown. He never has short hair.

Lionel Messi started playing football at a very early age in his hometown’s Newell’s Old Boys. From the age of 11, he suffered from a hormone deficiency and as Lionel’s parents were unable to pay for the treatment in Argentina, they decided to move to Barcelona, Spain.

In the 2003-2004 season, when he was still only 16 years old, Messi made his first team debut in a friendly match with Porto that marked the opening of the new Dragao stadium. The following championship-winning season, Messi made his first appearance in an official match on October 16th, 2004, in Barcelona’s derby win against Espanyol at the Olympic Stadium (0-1). And now, in 2010, 2011, and 2012 he is best player in the world.

*Sources: Soal Ujian Nasional (UN) SMA Tahun 2017.*

**Reading Micro Skill:** Implied details

1. According to the text, Messi’s parents moved to Barcelona ....
   A. because they were very poor in Argentina
   B. because they wanted Messi to be successful in soccer
   C. so that Messi could learn in the best soccer club
   D. to get Messi’s health problem cured
   E. to find the best treatment to cure Messi’s health problem

**Kunci Jawaban:

D**

---

Appendix 2. Sample of Final Reading Test

**DIRECTIONS:**

1. Write your name and student number on the answer sheet provided
2. Read the texts and write the answers on the answer sheet provided by using capital letter (A, B, C, D, or E)

---

Cite this as:

3. You may not use dictionary
4. You have 30 minutes to finish the test
5. Do not write anything on the test book

TEXT 1. For Questions No. 1-2

Messi’s Biography

Luis Lionel Andres Messi, born June 24th, 1987, is an Argentinian football player for F.C Barcelona. He is not very tall, mainly, due to the growing problem he had when he was younger. His eyes are brown. He never has short hair.

Lionel Messi started playing football at a very early age in his hometown’s Newell’s Old Boys. From the age of 11, he suffered from a hormone deficiency and as Lionel’s parents were unable to pay for the treatment in Argentina, they decided to move to Barcelona, Spain.

In the 2003-2004 season, when he was still only 16 years old, Messi made his first team debut in a friendly match with Porto that marked the opening of the new Dragao stadium. The following championship-winning season, Messi made his first appearance in an official match on October 16th, 2004, in Barcelona’s derby win against Espanyol at the Olympic Stadium (0-1). And now, in 2010, 2011, and 2012 he is best player in the world.

Sources: Soal Ujian Nasional (UN) SMA Tahun 2017.

1. What was Messi problem when he was eleven years old?
   A. because they were very poor in Argentina
   B. because they wanted Messi to be successful in soccer
   C. so that Messi could learn in the best soccer club
   D. to get Messi’s health problem cured
   E. to find the best treatment to cure Messi’s health problem

2. What is the main idea of the third paragraph?
   A. he doesn’t really attractive face, but he’s a very good football player
   B. Lionel Messi is a good player for F.C Barcelona
   C. his best characteristics are on the foot has competitiveness
   D. he appears to be a quite good and modest person
   E. he scored goals to Espanyol

BIOGRAPHY OF AUTHOR

First author
Anugrah Novendi Perkasa
An English Teacher in SMAK Santo Albertus Malang, Indonesia. He is teaching for senior high school students. His research interests are English language teaching, reading, and language assessment