

Assessment in Indonesian Higher Education: Developing a Reading Comprehension Test for English Students

Harits Masduqi*, Fatimah

Department of English, Faculty of Letters, Universitas Negeri Malang

Department of Languages and Literature, Faculty of Cultural Studies, Universitas Brawijaya

*(*corresponding author*)

✉ harits.masduqi.fs@um.ac.id

Abstract: This paper describes the processes of designing, administering, and evaluating a reading comprehension test for intermediate-level students who have taken a General English course at the English language centre of an Indonesian university. Using R&D method, the writers developed the test which contains two authentic texts. The first text is *It isn't a Rehearsal, You Know*, written by Ray Connolly and the second text is *Ageing around the World*, written by Timothy Johnson. The reading comprehension test is a part of summative assessment which is aimed to measure the students' ability to comprehend two different types of texts based on skimming, scanning, and vocabulary skills. Having done the processes of developing the reading test, the writers have some important points to revise including level of difficulty of the texts, the test organisation, and the reading skills tested. The writers recommend that understanding and practicing appropriate procedures to develop a reading comprehension test is not only crucial for English teachers to measure their students' reading skills, but also necessary for their professional development.

Keywords: intermediate-level students, language assessment, reading comprehension test

INTRODUCTION

Assessment has become an imperative issue particularly in higher education. It is part of students' university life since students' works are formally assessed to measure their mastery of certain subjects or skills. Assessment, according to Butler and McMunn (2006), is an action of collecting information about individuals or groups of individual in order to understand them better. More specifically, the Department of Education and Training (2021) states that assessment is the ongoing process of gathering, analysing, and reflecting on evidence to make informed and consistent judgements to improve future student learning. Furthermore, Brown & Abeywickrama (2010) assert that assessment refers to an ongoing process to measure students' performances. A teacher can make assessment on students' performance whenever they respond to questions, give comments, or try out new words and structures. Assessment is, therefore, not the same as test because test is only of the ways that teachers can use to assess their students. This is in line with the statement of Griffin and Nix (1991) that in assessment, information is obtained from a series of facts and is not mainly taken from tests, measurements,

scores, or grades. Thus, it can be said that assessment is an ongoing process to measure students' learning need, progress, and achievement; and it can be obtained from various instruments.

Assessment can be informal or formal. *Informal assessment* is implemented during the teaching and learning process in classroom tasks which are particularly designed to elicit students' performance without recording the result. It includes unplanned comments or responses, impromptu feedback on students' oral performance, and suggestions on how to read or write better. On the other hand, *formal assessment* is systematic and planned procedures that are designed to know students' achievement. It can be in the form of a test, a portfolio assessment, or an academic journal of students (Brown & Abeywickrama, 2010; Yorke, 2003).

Butler and McMunn (2006) argue that assessment may serve two purposes; to provide feedback to students and give information on student's mastery of content, knowledge, or skills. The first purpose is usually achieved through *formative assessment*. In this assessment, students get feedback on their progress. Yorke (2003) mentions that formative assessment deals with obtaining information about students' performance and their learning. Brown & Abeywickrama (2010) add that in formative assessment, students are evaluated during the process of forming their knowledge or skills so that they can continue developing their progress. This kind of assessment is usually continuous, on daily basis and closely related to teaching and learning program (Butler & McMunn, 2006; Yorke, 2003). The second purpose is achieved through *summative assessment*. It is an evaluative process and usually carried out at the end of a course of program (Brown & Abeywickrama, 2010; Griffin & Nix, 1991; Scriven, 1967). Brown & Abeywickrama (2010) further explain that summative assessment informs what particular students have learned and how well they have accomplished learning objectives.

In relation to assessment in Indonesian higher education, most universities expect their students to take English courses in order to improve their English competence to comprehend academic texts in English. The goal makes sense since reading in university level requires students to be able to read and comprehend English texts from different sources effectively (Cahyono and Widiati, 2006; Lamb & Coleman, 2008; Masduqi, 2014; Masduqi, Fatimah, & Subiyanto, 2021; Setiono, 2004). This paper describes the processes of designing, administering, and evaluating a reading comprehension test for intermediate-level students who have taken a General English course at the English language centre of an Indonesian university. The reading comprehension test is a part of summative assessment which is aimed to measure the students' ability to comprehend different types of texts based on skimming, scanning, and vocabulary skills. The students or the test takers are assumed to have mastered the three skills in their previous General English course. The test is also used to identify whether the test takers or the intermediate-level students are ready to go to the next level of an English course. This will be discussed further in the following sections.

DEVELOPING THE READING COMPREHENSION TEST

In developing the reading comprehension test, the writers use Research and Development (R&D) Method. The term R&D was first coined by Borg & Gall (1983). They suggest that this method is utilized for developing and validating educational products (Borg & Gall, 1983, 2003). Borg & Gall (1983) proposed ten steps in R&D starting from Research and Information Collection, Planning, Developing Preliminary Form of Product, Revising Main Product, Main Field Testing, Revising Operational Product, Operational Field Testing, Revising Final Product, and Disseminating and Implementing. In its implementation, there are some simplifications and models proposed by other researchers. Plomp (1997) simplified the process into five steps. His model was considered flexible since the researchers can adjust the

steps in accordance to the nature of their research. His five steps consist of Investigation, Designing, Realization/Construction, Testing, Evaluation, and Revision and Implementation. In investigation stage, researchers need to analyze the problem and situation in order to know the need and find a proper solution. Having obtained the necessary information, the researchers then design a model and plan the steps to develop the models. The next step is developing the model. The developed model is then tested to obtain feedback. Afterwards, the researchers revise the model based on the feedback and evaluation made. Finally, the revised model is implemented.

The development of R&D method evolves and Fotis and Mentzas (2006) proposed nine steps in the R&D model, while Hoge, Tondora, & Marelli (2005) proposed seven steps. The writers followed Plomp's model (1997) since this model is suitable for developing reading comprehension test and the steps are clear and simple. The writers, however, only employed the first four steps (Investigation, Designing, Realization/Construction, Testing, Evaluation, and Revision). The writer decided not to include the final step which is Implementation since the implementation of the revised reading comprehension test will be held in the next similar program. Please see the figure below for detailed steps.

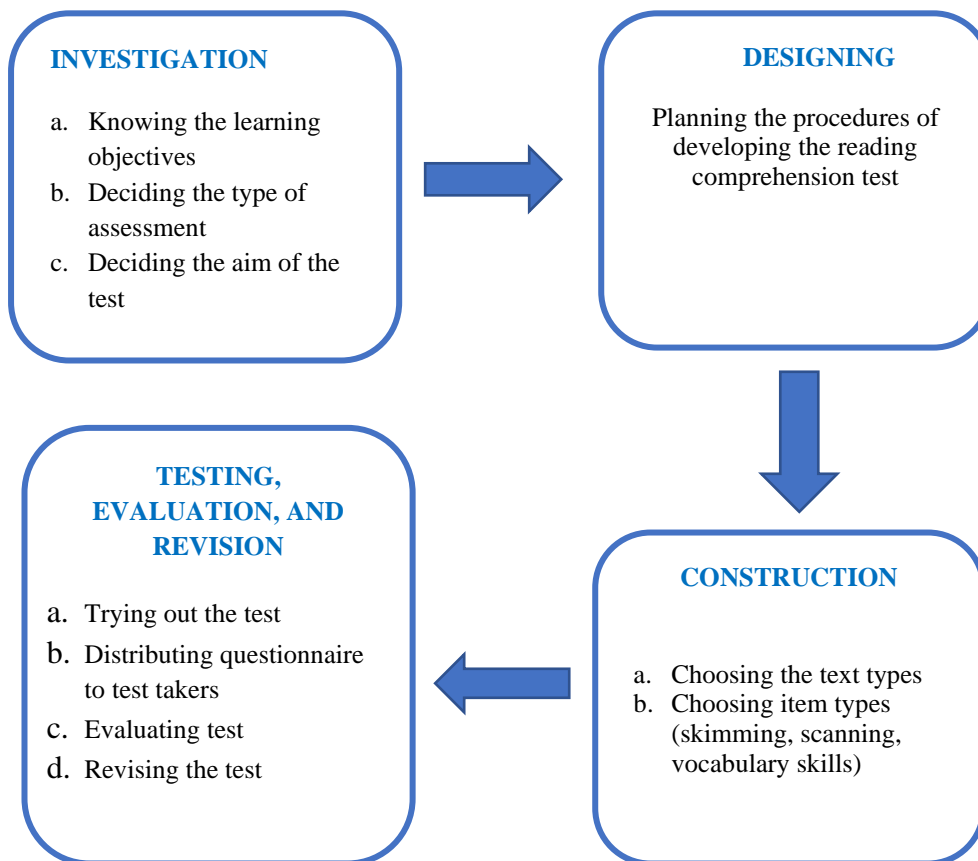


Figure 1. R&D Method in Developing Reading Comprehension Test

1. Investigation

The writers read the syllabus of the course to know the learning objectives that students have to acquire and the material coverage. Then, the writers agreed that the reading comprehension test is a part of summative assessment which is aimed to measure the students' ability to comprehend different types of texts based on skimming, scanning, and vocabulary skills. The students or the test takers are assumed to have mastered the three skills in their

previous General English course. The test is also used to identify whether the test takers or the intermediate-level students are ready to go to the next level of an English course. This will be discussed further in the following sections.

2. Designing

After the writers set the objectives of the reading comprehension test, the next stage is to design a model. The writers planned procedures to develop the reading comprehension test. In constructing the test, the writers chose the texts by considering several aspects, then the writers chose appropriate test items, tried out the test, and distributed questionnaire to test takers to obtain feedback and revised the model test based on the feedback and the writers' own evaluation.

3. Realization/Construction

In this step, the writers chose the texts by considering several aspects and chose appropriate test items. Each of the steps are explained below.

3.1 Choosing the Text Types

Two authentic texts with different genres are used in the test. The first reading passage, *It Isn't A Rehearsal, You Know*, which contains 815 words is a kind of reading for pleasure. It is written by Ray Connolly and was originally published in the *Standard* on 14 May 1984. The second reading passage, *Ageing around the World*, contains 610 words and is inclusive of reading for knowledge. Published in the *World Times* on 28 July 2002, the article was written by Timothy Johnson.

In terms of authenticity, the writers have two reasons to choose the two texts. First, both texts are published in well-known English magazines that are always up to date with the current development of western culture as well as journalism and therefore, suitable for English students. Secondly, the two texts consist of day-to-day discourses which happen in real life. The writers believe that the issue of authenticity is always an important aspect of any discussion on assessment. According to Bachman and Palmer (2014), good language assessment strives to use formats and tasks that mirror the types of situations in which students would authentically use the target language. Whenever possible, teachers should attempt to use authentic texts and materials in testing language skills. Moreover, language learners are motivated to perform when they are faced with tasks that reflect real world situations and contexts.

Regarding the different genre of the two texts, the writers argue that it is also a crucial point in language assessment. The rationale of using two texts with different types is mainly based on two ideas. The first comes from Allen (1997) that a good reading test which covers more than 20 test items should contain at least two passages. The passages should have different types and level of difficulty to avoid student's nervousness, boredom, and fatigue. Secondly, reading texts are generally divided into two types: reading for pleasure and reading for knowledge. Reading for pleasure is suitable for readers who read a particular text to simply enjoy leisure times. It is reading for entertainment or enjoyment. In contrast, reading for knowledge is a more serious reading done by people who are always eager to find new ideas, thoughts, and information. It is reading for education (Thompson, 2002).

3.2 Choosing the Item Types

Having selected the two texts, the writers made 24 test items. Each text has 12 questions. The writers decided to have two test types. The first type is multiple choice covering

three tasks; choose the best answer of four possible options, match the statements from column A with the words in column B, and select True or False. The second type is open-ended questions. The open-ended questions cannot be answered with a simple ‘yes’ or ‘no’, and instead, require the students to elaborate on their points. These types of questions were aimed to help the writers or the test makers see things from the students’ perspectives as the answers were given with their own words.

Regarding the multiple-choice type, the writers think it is still reasonable to use multiple choice test items as they are commonly used to measure students’ reading skills. Multiple choice questions require fine distinctions between correct and nearly correct statements. Students can learn that these distinctions are not merely from smart guessing, but from essential reading skills, such as: skimming, scanning, surveying, and so forth. These question types sometimes make the content of the questions unrecognisable since students often read these types of thinking questions carelessly. Therefore, it is to the students' advantage to learn about the reading skills required to answer multiple choice questions and to learn how to read the questions carefully (Powell, 2003). From this point on, the writers choose multiple choice test items to measure the test takers’ ability to perform skimming, scanning, and vocabulary skills. The questions on each text are divided into four sections, namely section A (Choose the best option), section B (open-ended question), section C (True/False) and section D (Matching). All of the items in the sections mainly cover three skills, namely Skimming, Scanning, and Vocabulary Skills. The test items and skills are elaborated as follows:

a. Skimming

Skimming involves searching for the main ideas by reading the first and last paragraphs, noting other organizational cues, such as summaries, used by the author. Skimming for the main ideas by using the author's organization cues (Topic headings, italics, summaries, etc.) is a vital preliminary step to more intensive reading and maximum retention. It will provide a logical framework in which to fit the details (Brown and Yamashita, 1995). The examples of test items which measure skimming skill are shown below.

TEXT 1 Section A (Choose the best option) no 1

1. What does the passage mainly discuss?
 - a. Life is too short to work only
 - b. Don't do today what you can do tomorrow
 - c. Enjoy your life before it is too late
 - d. How to manage your time creatively

The question ‘What does the passage mainly discuss?’ is actually the same as “What is the topic of the passage?”. In other words, the question asks the test takers to activate their skimming skill, since it is principally used to search for main ideas of the passage. It is, therefore, suitable with the purpose of the test.

b. Scanning

Scanning involves running your eyes down the page looking for specific facts or key words and phrases. It is an aid in locating new terms to follow the author's reasoning without dictionary or glossary. Thus, scanning alerts readers to the new terms and concepts and their sequences (Brown and Yamashita, 1995). There are many test items which measure this reading skill, in particular open-ended questions and True or False questions. Some of them are as follow:

TEXT 1 Section B (open-ended question) no 2 and 3

2. What is the difference between *Real Life* and real life in the following sentence:
“We look back and realise that all those years waiting for *Real Life* to come along were in fact real life (paragraph 5)”.
3. Why does the author consider the man in Kent to be one of the luckiest chaps (par. 10)?

TEXT 1 Section C (True/False) no 5 and 6

Put T (True) or F (False) into the circles of the following statements.

5. Real Life begins when you have a home and family
6. The English people like saying ‘Have a nice day’.

TEXT 2 Section A (Choose the best option) no 1

1. Why would workers prefer investment to superannuation? Because an investment...
 - a. allows a huge payment at the end of retirement
 - b. is like a pension
 - c. is a reward
 - d. meets the expenses in old age

TEXT 2 Section B (open-ended question) no 4

4. What reasons are given by the author that ageing is not a disease?

All open-ended questions and True/False statements above contain specific information or keywords located in certain paragraphs. This activity involves the test takers’ ability to find or locate specific terms and concepts and their sequences. Thus, it is aimed to measure test takers’ ability to use their scanning skill.

In the case of open-ended questions, some researchers argue that they are the key to measure a deep understanding of test takers toward specific ideas as well as main ideas in particular paragraphs. The open-ended questions provide an opportunity for test takers to explore and reflect upon their ideas within a narrative of their own choosing. The context provided by answers to open-ended questions provide teachers with valuable clues about how far their students master scanning and skimming skills (Harmer, 2015).

c. Vocabulary Skills

In English reading classes, teachers often emphasise that vocabulary skill is very crucial to identify whether someone is a competent reader or not. Harmer (2015) argues that vocabulary skills are essential for students to comprehend reading passages. It involves not only the ability to know what words mean, but also how they are used. Whilst this obviously involves giving them the names for things (e.g. ‘table’, ‘chair’, etc.), it also involves showing them how words are stretched and twisted (e.g. ‘to table a motion’, ‘to chair a meeting’). In this case, students should be aware of the vocabulary items and its position in a sentence, so that they can do ‘a clever guess’ of meanings when they find new words or phrases in a sentence.

In this test multiple choice questions, such as: the four options of a, b, c, d and matching activities, are the best to measure students’ vocabulary skill. Some of the test items are:

TEXT 1 Section D (Matching) no 10-12

Match the underlined word in Column A with its synonym in Column B

| Column A | Column B |
|--|--|
| 10. It is, I believe, a <u>delusion</u> I share with the great hopeful majority. 11. In English they have a saying much <u>ridiculed</u> by the English. 12. I am not <u>advocating</u> that one should live for the minute in any hedonistic sense. | a. embarrassed b. practice c. suggesting d. story e. fantasy f. ordering g. mocked |

TEXT 2 Section A (Choose the best option) no 1

1. The image we have of older people is very prejudiced.
 What does the underlined word mean?
- a. tolerant c. bias
 b. fair d. broad minded

TEXT 2 Section D (Matching) no 10-12

Match the underlined word in Column A with its synonym in Column B

| Column A | Column B |
|---|--|
| 10. Workers may get a <u>lump sum payment</u> when they are retired. 11. Valuing the traditional role of older people in the family can help <u>struggling</u> families in poor countries. 12. Older people <u>pass on</u> the culture to younger generations | a. continue b. carefree c. constantly d. investment e. stressed f. progressively g. explain h. superannuation |

The two types of the multiple-choice format really measure the vocabulary skill of test takers. Both types contain specific terms that can be new vocabulary items for test takers. To some extent, these activities may involve scanning skills because all statements in the two activities are taken precisely from particular paragraphs of the passage that can be scanned by test takers. However, the multiple-choice items tend to measure vocabulary skills rather than scanning, since all key terms in all stems are underlined and the options provided are in the form of single vocabulary item.

4. Testing, Evaluation, and Revision

The writers' next steps are piloting the model test, distributing the questionnaire to the test takers after they completed the test, and evaluating the model test by reviewing the results of the questionnaire and the writers' self-evaluation on the test. Then, the writers revised the model test.

4.1 Test Trial

After making the final draft of the test, the writers began to look for the test takers. It took seven days to get the ten people volunteering for taking the test. The test takers were between 18 and 29 years old. They are a mixed group of male and female English students from different cultural backgrounds in Indonesia. They were willing to take the test because

they wanted to help us who were formerly their English teachers. In addition, they also want to know how far they have mastered reading skills after completing a General English course in their universities.

The important step in test trial is deciding the time for testing. According to Bachman and Palmer (2014), the time chosen for testing is influenced primarily by considerations of reliability and practicality. Based on this point, the writers administered the test together at the same time and place at the language centre of a top university in Malang, East Java, Indonesia nearing the end of 2021 when the Covid-19 Pandemic is relatively manageable. Based on the initial agreement, the writers administered the test during the day and gave the test takers 30 minutes to do the test. More specifically, the time allotment for the Text I was 15 minutes, while the time allocation for the Text II was 15 minutes. The test was done in a classroom that the writers have set beforehand.

Before the test takers started to do the test, the writers explained all instructions of the test as clearly as possible. For example, what they must do when facing the instruction ‘Choose the best answer of four possible options’, ‘Match the statements from column A with the words in column B, and so forth. During the test, the ten test takers did not have any problems with the test instructions.

The test takers who were intermediate-level students finished doing the test on time. Then, the writers gave them a ten-minute break before fulfilling the questionnaire that the writers had prepared beforehand. During the break the writers asked their opinions about the test. They said that the test was moderate in terms of difficulty. Some questions were difficult, and some were manageable. Having had the break, they fulfilled the questionnaire and returned it to the writers together with the test answer sheets.

4.2 Feedback Gathered

Two days after the trialling process, the writers had a meeting to discuss the test results as well as feedbacks from the test takers. Regarding the feedback, the writers found a variety of opinions from test takers. The questionnaire contained four questions as follow:

1. Which test item is the most difficult? Why?
2. Which test item is the easiest? Why?
3. Which test item is ambiguous/confusing? Why?
4. What do you think about the overall test?

For the first question of the questionnaire, three test takers stated that open-ended questions (Part B) both in Text 1 and Text 2 were the most difficult, while the other seven claimed that matching activities (Part D) of the test was the most difficult. In this case, the writers’ prediction is the same with most test takers. Matching activities (Part D) are difficult because there are many vocabulary items that are possibly unfamiliar for the Indonesian test takers, such as: delusion, procrastinating, exhortation (Text I) and lump sum payment (Text II).

Concerning the second question of the questionnaire, all the test takers agreed that True-False questions were the easiest. This makes sense because all statements in True-False questions were taken directly from the texts. They could directly scan the statements from certain paragraphs to get information whether they are true or false.

The answers for the third question of the questionnaire were various. Seven test takers had the same opinion that the second question of Part B, Text I was very confusing. It is probably because the text does not explicitly discuss the difference between the two terms (*Real Life* and *real life*) and therefore, test takers have to make inferences. Another test taker said that finding main idea (the first question, Part A, Text I) was confusing, whilst the last test taker

claimed that matching vocabulary (Part D, Text II) was ambiguous. This is possibly due to the fact that some options (distracters) were almost similar in meaning.

In responding the last question of the questionnaire, nine test takers stated that the test given was neither too easy nor too difficult. Only one test taker said that the test was a bit difficult for him and he claimed that it was because there were too many new words in the texts. In addition, there was an interesting comment from one test taker about the organization of the test. She said that various types of the test items made the test looked interesting.

4.3 Revision of the Model Test

Having developed and tried the test, the writers have some important points to revise. The revisions the writers made are based on the feedback given by the test takers and the writers' self-evaluation on the test. First, concerning level of difficulty of the texts, it would be better if Text II *Ageing around the World* is used as Text I, vice versa Text I *It Isn't A Rehearsal, You Know*, becomes Text II. The writers prefer this text position because Text II *Ageing around the World* is easier to understand than Text I. Psychologically, test takers would be more encouraged to do this kind of test if the first text is easy to understand.

Secondly, the test organization needs to be refined due to the issue of parallelism. Text I which contains only one multiple choice of four options format (a, b, c, d) is not parallel with Text II which has two multiple choice of four options format (a, b, c, d). The writers suggest that the number of test items in Part A should be the same for both texts.

Thirdly, an issue of skimming skill needs to be addressed fairly. There is one test item in Text I ('What does the passage mainly discuss?' – no. 1, Part A) which focuses on measuring test takers' ability in skimming. On the contrary, there is no test item which clearly measures skimming skill in Part A of Text II. Thus, the writers recommend Text II should also contain a test item, 'What does the passage mainly discuss?'

Lastly, the test item in Text II ('What does the underlined word mean?' – no. 1, Part A) which measures test takers' vocabulary skill should be located in part D (matching activities). Some vocabulary items in Part D as well as some True or False statements in Part C also need improving. The following Table 1 and Table 2 display the revisions made in the original reading comprehension test.

Table 1. Text Title: *It Isn't A Rehearsal, You Know*

| ORIGINAL TEST | | | | REVISED TEST | | | |
|----------------------|-----------------------------|-------|------------|----------------------|-----------------------------|------------|----------|
| Test Type | Section | No | Skills | Test Type | Section | No | Skills |
| Multiple Choice | A. Choosing the best answer | 1 | Skimming | multiple choice | A. Choosing the best answer | 1 | Skimming |
| | | | | | | 2 | Scanning |
| | C. True or False | 5-9 | Scanning | | C. True or False | 5-8 | Scanning |
| | D. Matching | 10-12 | Vocabulary | D. Matching | 9-12 | Vocabulary | |
| Open-Ended Questions | B. Open-ended | 3-4 | Scanning | open-ended questions | B. Open-ended | 3-4 | Scanning |

Table 2. Text Title: Ageing around the World

| ORIGINAL TEST | | | | REVISED TEST | | | |
|----------------------|-----------------------------|-------|-----------------------|----------------------|-----------------------------|------|-----------------|
| Test Type | Section | No | Skills | Test Type | Section | No | Skills |
| Multiple Choice | A. Choosing the best answer | 1 | Scanning | multiple choice | A. Choosing the best answer | 1 | <i>Skimming</i> |
| | | 2 | Scanning (vocabulary) | | | 2 | <i>Scanning</i> |
| | C. True or False | 5-9 | Scanning | | C. True or False | 5-8 | Scanning |
| | D. Matching | 10-12 | Vocabulary | | D. Matching | 9-12 | Vocabulary |
| Open-Ended Questions | B. Open-ended | 3-4 | Scanning | open-ended questions | B. Open-ended | 3-4 | Scanning |

The important lesson the writers learned is that designing, administering, and evaluating a test are not as easy as the writers thought. The test development is proven to involve some interrelated procedures, such as:

1. Test design which covers:
 - Title of test specification
 - A description of the purpose of the test
 - A description of text and item types
 - A description of the test takers or whom the test is intended
 - A definition of the construct to be measured
2. Test operation which includes:
 - Writing and developing test task/item
 - Writing instruction
 - Making a formula for scoring the test
3. Test administration which involves:
 - Preparing the place for the testing
 - Deciding the time for testing
 - Giving the test to the test takers
 - Communicating the test instruction with the test takers
 - Collecting the test from the test takers
 - Collecting feedback from the test takers

Reflecting on what have been done above, the writers relate the development of the reading comprehension test with five main considerations in designing an assessment proposed by Brown & Abeywickrama (2010). The first is *practicality* meaning that the assessment is not too expensive, relatively easy to administer, does not consume too much time, and has time-efficient and specific scoring procedure. The second principle is *reliability* which refers to

consistent and dependable. Weigle (2002) and Gronlund and Waugh (2009) define practicality as consistency of measurement across different characteristics of a testing situation such as a different prompt or raters. Brown & Abeywickrama (2010) argue that there are several factors that might affect the reliability, such as issues related to students or test-takers (e.g. illness, fatigue, anxiety), rater (e.g. lack of attention, inexperience, inattention), test administration (e.g. technical problem, condition of room, noise), and the test itself (e.g. too long, poorly written test). In writing assessment, Weigle (2002) states that the reliability of assessment might be influenced by the writing task itself (e.g. the topic, the number of writing samples a test-taker need to provide) and the scoring process (e.g. background and experience of raters, training give to raters, the nature of rating scale used).

The third principle is *validity* which mainly concerns with the interpretation and the use of assessment result (Bachman & Palmer, 2014; Brown & Abeywickrama, 2010; Gronlund & Waugh, 2009). In developing a valid assessment, there are some aspects to be considered. They classify the aspects into content representatives (determining the adequacy and representativeness of test content), criterion relationships (determining the extent to which criteria of the test are reached), construct evidence (determining whether a particular test is actually measuring what it is intended to do), consequences in using (determining the consequences in using a particular assessment), and face validity (determining to what extent a test looks right and appears to measure what it is intended to measure).

The fourth principle is *authenticity*. Brown & Abeywickrama (2010) suggest that authenticity can be achieved if the test uses of natural language and thematic organization, presents test items in context rather than in isolation, covers topic which is relevant and interesting for targeted learners, and it represents or closely represents real-world tasks. In writing assessment, authenticity means that a writing task represents type of writing that the test-takers will need in the real world beyond the test (Weigle, 2002).

The last principle is *washback*. Brown & Abeywickrama (2010) explain that washback includes the effects of assessment on teaching and learning during the preparation for the assessment. In summary, those five principles are important and should be taken into consideration by English teachers in designing assessment for a particular course or program.

CONCLUSION

In this paper, the writers describe the processes of designing, administering, evaluating, and revising a reading comprehension test for intermediate-level students who have taken a General English course at the language centre of an Indonesian university. Following the Plomp's model (1997), the writers employed the first four steps (Investigation, Designing, Realization/Construction, Testing, Evaluation, and Revision). The writer decided not to include the final step which is Implementation since the implementation of the revised reading comprehension test will be held in the next similar program.

Having done the processes of developing the reading test, the writers have some important points to revise including level of difficulty of the texts, the test organisation, and the reading skills tested. The writers recommend that understanding and practicing appropriate procedures to develop a reading comprehension test is not only crucial for English teachers to really measure their students' reading skills, but also necessary for their professional development. English teachers who are capable of designing and developing a valid and reliable test are important for their schools as they can take parts in the school assessment program and curriculum development.

REFERENCES

- Bachman, L.F. & Palmer, A.S. (2014). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Borg, W.R. & Gall, M.D. (1983). *Educational research: An introduction*. New York: Longman.
- Borg, W.R. and Gall, M.D. (2003). *Educational research: An introduction* (4th Ed). London: Longman Inc.
- Brown, H. D. & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd Ed). London: Pearson Education.
- Brown, J. D. & Yamashita, S.O. (1995). *Language testing in Japan*. Tokyo: Japan Association for Language Teaching.
- Butler, S. M., & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: understanding and using assessment to improve student learning*. San Francisco, CA: Jossey-Bass.
- Cahyono, B. Y., & Widiati, U. (2006). The teaching of EFL reading in the Indonesian context: The state of the art. *TEFLIN Journal*, 17(1), 23-46.
- Department of Education and Training. (2021, May 7). Assessment of student achievement and progress foundation to 10. Retrieved from <https://www2.education.vic.gov.au/pal/assessment-student-achievement/policy>.
- Fotis, D., & Mentzas, G. (2006). Competency based management: A review of systems and approaches. *Information Management & Computer Security*, 14(1), 51-64.
- Griffin, P. E., & Nix, P. (1991). *Educational assessment and reporting: A new approach*. Sydney: Harcourt Brace Jovanovich.
- Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement* (Vol. 9th). Upper Saddle River, N.J: Pearson.
- Harmer, J. (2015). *The practice of English language teaching* (5th ed.). London: Pearson Education.
- Hoge, A., Tondora, J., & Marelli, A. F. (2005). Strategies for developing competency models. *Administration and Policy in Mental Health*. Vol. 32 (5), 533-561.
- Lamb, M., & Coleman, H. (2008). Literacy in English and the Transformation of Self and Society in Post-Soeharto Indonesia. *International Journal of Bilingual Education and Bilingualism*, 11(2), 189-205. doi: 10.2167/beb493.0
- Masduqi, H. (2014). EFL reading in Indonesian universities: Perspectives and challenges in cultural contexts. *Journal of teaching and education*, 3(3), 385–397. <http://www.universitypublications.net/jte/0303/pdf/V4G339.pdf>.
- Masduqi, H., Fatimah, F., & Subiyanto, A. (2021). Qualitative research into reading in English as a foreign language in Indonesia – Quo Vadis? *Indonesian TESOL Journal*, 3(2), 76-87. <http://ejournal.iainpalopo.ac.id/index.php/ITJ/article/view/1805>.
- Plomp, T. (1997). *Development research on/in educational development*. Netherlands: Twente University
- Powell, R. (2003). *Taking multiple choice exam*. Melbourne: University of Victoria Press.
- Scriven, M. (1967). The methodology of evaluation. In R.W.Tyler (Ed.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Setiono, S. (2004). Competency-based learning: The dreams and realities, *The Jakarta Post*. June 29, 2004.
- Shohamy, E. (1996). *Open ended questions vs. closed ended questions*. New York: Allyn and Bacon.
- Thompson, C. (2002). *Types of reading*. Bloomington, Indiana, USA: ERIC Clearinghouse on Reading & Communication Skills.

- Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.
doi:10.1023/A:1023967026413.