

TEACHERS' COMPETENCE IN A READING TEST CONSTRUCTION

Luthfiyatun Thoyyibah

*English Education Department, Faculty of Teacher Training and Education
Universitas Galuh, Indonesia
Email: luthfiyatun20@gmail.com*

APA Citation: Thoyyibah, L. (2022). Teachers' competence in a reading test construction. *Indonesian EFL Journal*, 8(2), 205-214. <https://doi.org/10.25134/ieflj.v8i2.6457>

Received: 07-03-2022

Accepted: 12-05-2022

Published: 31-07-2022

Abstract: Having a certain reading ability is one of the prerequisites for becoming a professional teacher. The focus on language literacy, particularly reading abilities, has been upgraded as well. The focus of this research is to look into the process of developing an English teachers' reading test in vocational schools, and at the same time to investigate the quality of teachers' reading test reflects their ability in test construction. This study employs a case study approach, focusing on three English teachers who work in three distinct vocational schools. This study employs interviews to better understand the process of creating reading tests, as well as thorough expert standards to assess the quality of teacher-created tests. The study discovered that the process of creating a reading test for teachers includes the basic ability to find materials, select the correct text, determine the type of question, and determine the number of digits in the test. Read and rate. As far as the quality of the reading test conducted by the teacher is concerned, the results of the scoring scale show that the teacher's first reading test is considered poor and good because the passing scores of the test are 69 and 87 respectively. With scores of 51 and 56, reading test number two created by the teacher of English was deemed inadequate. The third teacher's reading test score was 81, so he was considered very good. Other teachers' abilities in test creation have not been effectively utilized in terms of test design, relevance, balance, efficiency, validity, reliability, adequacy of test items, and technical voice of reading test. For reading test, English teachers should use authentic materials. The teacher is also suggested to use real text to create high-level comprehension problems. It is preferable to use a more comprehensive list as one of the study techniques in the future and to involve more participants.

Keywords: *Reading test; competence; criteria; process of constructing.*

INTRODUCTION

Assessment has major consequences for both students and teachers. In assessing students' learning, teachers use tests. Test becomes the most possible tool that teacher can use to see students' learning (Hartell & Strimel, 2019). That relates to teachers' competence in test construction, including developing reading test.

Teachers need to be testing literate. That is clearly mentioned in the government regulation number 19 in 2005 about the national standard of teachers. Hence, there were four competencies that the teachers should accomplish; namely professional competence, pedagogical competence, self-competence, and social competence (Maolida & Anjaniputra, 2017). Likewise, Ma'rifatullah, Ampa, & Azis (2019) mentioned that constructing test is one of a teacher's pedagogical competences besides the ability to plan, to teach, and to develop students' reflection on learning. In line with Ma'rifatullah et al., (2019), Saraceno (2019) stated

that there are two kinds of content knowledge that teachers need to master, which are disciplinary and pedagogical content, in which assessment is included to one of them, the pedagogical content.

Based on the aforementioned things of the teacher made test, there are some research studies about the quality of reading test made by the teacher. A study from Elleman and Olsund (2019) explicitly mention that most teachers believe that they need strong measurement skills in developing the reading test. The teacher did not have sufficient knowledge of constructing appropriate reading test.

A study conducted by Razali and Jannah (2015) elaborated the comparison of the quality of teacher made-test on try-out with the national test items. According to the findings, more than half of the try out test items are irrelevant to the national test items. Meanwhile, the final national examination test items were created with a higher cognitive domain than teacher-created test items. It demonstrates that the teacher-created test items

(tryout) are more superficial than the final national test items. It could be one of the reasons why students who pass the teacher's tryout are unable to pass the national examination.

On the contrary, Hakim & Irhamsyah (2020) pointed out that the questions posed by the English teacher at Kutacane State Senior High School were mostly valid and in line with the curriculum. As a result, the English teacher at Senior High School 1 Kutacane created valid questions, and the level of validity created by the teacher demonstrates teachers' ability to design questions in English.

Furthermore, Saefurrohman and Balinas (2016) conducted a study to Philippines' senior high school teachers and Indonesian's junior high school teachers in constructing a test which showed that Philippines teachers tend to use their own made test, meanwhile Indonesian teachers prefer to use test available from textbooks.

As designing reading test includes to pedagogical competence, it is an essential part of evaluating students' learning engagement as well as their level of skills in applying what they have learned. Thus, they need to be tested by a valid and reliable test. However, research in Indonesia regarding teachers' competence in reading test construction, to researchers-based knowledge, are still limited. Therefore, it is worthwhile to conduct a study under the topic of teachers' ability to construct a reading test.

Teachers as professional educators have their main role in language teaching and testing. They perform their role in language teaching through some competencies. The definition of competence is determined as an entire concept of mixture of knowledge, understanding and skill (Caena, 2020).

Due to the importance of teacher's competencies, some educational experts put the competencies into some categories. One of the most crucial teacher competencies is disciplinary literacy pedagogical content knowledge (Saraceno, 2019).

Constructing a test is included to one of the pedagogical contents. One of specific competences in assessing students' ability in learning is creating test. Creating a test is crucial for the teachers of English. Through test, teachers are able to see the development of students' achievement.

Reading becomes one of required skills in 21st century. Language literacy has received renewed attention and emphasis in recent years, as twenty-

first-century global citizens, we must not only speak English, but also read and write it.

There are many definitions of reading. M. Wallace & Wray (2015) determined reading as a unitary and selective process. This means a reader poses some actions and willingness in gaining knowledge from a text. Another expert, Harmer as cited in Pazilah (2019) mentioned that reading is an activity that engages both the eyes and the brain; the eyes pick up on signals., and the brain must decipher the meaning of these messages. Based on what Pazilah (ibid) said, it can be assumed that reading makes the readers not only use their eyes but also become active to use their brain in comprehending what they read. To comprehend the meaning of the text, prior knowledge of the subject is required. The background knowledge is also useful in shaping situational or contextual knowledge.

Reading is also beneficial for language acquisition, according to Zulmaini (2021). Assuming that students comprehend the text, greater they read, the better they can be at it. Thus, reading relates to the routine activity of an individual. The routine activities combine between products and meaning because reading is getting meaning from the text.

In relation to reading strategies, there are many different views on it. Susanti (2020) defined reading strategies as "generally purposeful, fun actions performed by productive students, several times to remedy levels of cognitive failure." Furthermore, reading strategies, according to Firdaus (2017), are the emotional processes or knowledge processes that readers can choose and adapt in make meaning of what those who read. The term 'reading strategy' refers to a particular tactics used by the reader to fully understand the actual meaning. Thus, reading strategies are the methods applied by EFL students in order to gain goals in comprehension skills of reading.

On the other hand, Alderson (2000) pointed out reading strategies that are beneficial for students into some categories, those are: predicting, skimming, scanning, inferring, guessing the meaning of new words, self-monitoring, and summarizing.

The notion of test is measuring learning outcome among students. Therefore, the test is one of the most powerful tools for measuring a student's ability and improving their attitude towards

learning. This concept is supported by Hughes as cited in Rahmatun & Helmanda (2020), who stated that the test is a tool for measuring a student's language proficiency. In addition, Shohamy, Or, & May (2017) stated that testing is a method of measuring an individual's ability to know and perform a particular area. Similarly, Chen, Halilah, & Shauqiah (2017) defined assessment as a process performed to consider learners performance in a specific area within a particular time limit with a particular goal.

Test appears in certain intervals. There is dichotomy of time in administering the test. In most cases, the test will appear after the education process, but it may also appear to be placed before the education process. It depends on its purpose. It depends on its purpose.

First, one most important thing about creating a reading test is choosing the right text. Choosing the right text should be in line with the purpose of the course objectives. Second, the selection of the text should be in line with a number of students that will be tested. It is supported by Hughes (ibid) that reading test should be clear whether it is high stakes test or not. Third, is about the authenticity of the text. Clifford and Parry (2014) dismissed the fact that many of the genuine texts used in reading tests are not used in the facsimile format, thus denying reader clues such as fonts and formats. I'm emphasizing. Many reading texts come from more than one page of text.

In relation to designing a good reading test, Alderson (2000) broke down test specifications that should be covered the following points, those are test purpose, the learner taking the test (age, sex, level of language proficiency, first language, cultural background, country of origin, educational level, and nature of educational reason for taking test, likely personal and professional interests and levels of background knowledge etc.), test level (in terms of test takers ability), test construct, description of suitable language course or textbook, number of sections to the test, time for each section, weighting for each section, target language situation(s), text types, text length, text complexity/difficulty, language skills to be tested, language elements (structures/lexis/notions/functions), task types, number and weight of items, test methods, rubrics; example; explicit assessment criteria, criteria for scoring, description of typical performance at each

level, description of what candidates at each level can do in the real world, sample papers, and sample of students' performances on tasks.

METHOD

Three different English teachers from three distinct vocational schools participated in this study. Case study design, as the name implies, is concerned with the entire process. The set was picked with care because it deals with the process, relying on purposive sampling approaches.

This study involved three different participants. The first participant was identified as a crucial informant by the researcher (Cohen, Manion, & Morrison, 2017). The researcher discovered that she, too, is a teacher at the similar home school based as the first participant. He has almost twenty-five years of experience as a teacher. The second person is a twenty-eight-year-old woman. For the past five years, she has been working as a teacher at one of Banjar's public vocational schools. In addition to her vocational school teaching duties, she also led and taught a private English course. The age of the third participant is twenty-nine. Since 2014, she has worked as a teacher at a private vocational school. She attended the similar university as the second participant. All of the people who took part in this study were members of the Banjar forum of vocational English teachers.

Additionally, this study used multiple data sources. This included sources of both document analysis and data from interviews. Numerous data sources have helped researchers explore and recognize the study's focus. This is the teacher's ability to elevate the grade of the test designed by the teacher and the process of writing a reading test.

This research focuses on a particular case: instructor competency in creating a reading test. This study allows for an in-depth examination of specific details of teachers' abilities to build a reading test by focusing on a single scenario (Cohen et al., 2017).

When developing a reading exam, this study also used a small scale to maintain a comprehensive approach to understanding the context of teacher competence. This study needed extensive and thorough detailed from the teacher's conversation because of the tiny scale.

RESULTS AND DISCUSSION

The process of reading test construction

The primary data of answering the way teachers construct a reading test for their students taken from the teachers' interview. Each teacher had their own way of constructing the reading test. The elaboration of constructing a reading test is explained in the following result and discussion.

Making a plan became the most crucial thing in constructing a valid and reliable test. It covered everything in the test. Planning a test determined the numbers of reading test, the type of question and level of difficulties. But one thing that should be taken into account is the word "*sometimes*". It indicated that the teacher is frequently made the plan for the reading test. It was being admitted that the teacher did that only to fulfill their obligation as a teacher. It's just a routine.

The following step was preparing the material. It should be in line with the course objective. The course objective can be seen through the basic competence in the syllabus. Some teachers revealed that they sometimes combine more than one basic competences into one for one reading test. It depends on the materials' complexity.

The first reading test from the first participant was for the tenth grade of vocational school. The materials in the reading test were taken from one of the basic the basic competences for knowledge and practice from the syllabus. Next, the teacher decided on how many numbers in the test items because it will affect the time allotment for the test. Then, the teacher made the answer key. The teacher revealed that preparing the formula of examining the test was beneficial for him. The teacher stated that he got easier in examining and released the score of the students.

The finding from other teachers were quite similar. Explicitly, the teacher mentioned that one reading test could cover more than one basic competences. The next step was deciding the type of questions. It was followed by deciding the number of the questions. Considering the time allotment, the teacher made a decision on the numbers in the test after deciding the type of questions. The last step conducted by the teacher was about scoring.

In deciding the materials, it involved choosing a text. The authenticity of the text becomes an issue in reading test (Jahan & Ashraf, 2020). They further point out that many reading text come from more than pages of text. Based on that fact, students need to be prepared for the real situation.

Based on the finding on reading tests, three out of five reading tests contained authentic texts. Choosing the right text became the following stages in creating a reading test after planning the test. Planning the test included determining the basic competence and materials covered in the reading test. The choice on type of the questions and the numbers of questions affected aspects of validity and reliability of test. According to the result of this study, two out of five reading tests were categorized into less valid as it did not require the students or test takers to read the text in answering the questions. It also used many exact words in the questions. It made the students cut and paste the answer from the text presented. It did not involve reading comprehension. Reading involves using the brain in comprehending what they read. The two reading tests mentioned were less reliable as they did not give the information on what skill is being tested. The type of questions from the reading tests mostly consisted from multiple choice questions, matching, and short answer questions. The finding on types of questions was in line with the study conducted by Santy, Dewi, & Paramartha (2020). The finding of the study revealed that most secondary school teacher made multiple choice questions followed by essays.

Actually, based on the scoring, teacher could see on the strengths and weaknesses of the students' ability in answering the reading test. One at a time, teacher could also investigate the quality of questions they made. The result of that finding could be as a valuable input for teachers in teaching reading and constructing good questions for reading test.

Based on the procedures from Burke (1999), the teachers have implemented several steps in developing reading test, such as make sure of the correlation on the test with the course objective, arranging various type of questions, make sure the appropriate reading, making various level of questions, and allocate sufficient time for all students to finish the test.

Relevant to teacher competences, especially the pedagogical competence, this study found that all teachers involved in this study aware of the ability in constructing a test, especially reading test. They have taken some steps toward developing a reading test. They had an understanding on the appropriate use of language test. They knew what should be covered in one test as one of requirements for

competence in language testing. Through constructing and administering a test, teachers can see students' achievement. Students' achievement is one of the tools that teacher can use for evaluating their teaching and testing.

The quality of reading test on criteria of a good reading test

The checklist comprised of eight categories. The findings of the criteria checklists on a good reading test are presented as follow.

Table 1. *The finding from the first teacher*

Reading test 1				
Category	Rater 1	Rater 2	Rater 3	Average
Test plan	25%	25%	25%	25%
Relevance	50%	50%	50%	50%
Balance	100%	50%	50%	67%
Efficiency	100%	100%	100%	100%
Validity	100%	67%	67%	78%
Reliability	100%	67%	67%	78%
Adequacy of the test items	86%	86%	86%	86%
Technical	67%	67%	67%	67%
Rating scale				69%

The finding from the first reading test by the first participant reached 69%. It means that the reading test made by the teacher failed met several criteria in the instrument of a good reading test's criteria checklist. Accordingly, based on the rating scale used in this study, this reading test is considered 'poor'.

Besides, according to table above, it is shown that only one criterion, efficiency, reached 100%. From the descriptors of efficiency criterion, this percentage reveal that the texts used in the reading test is authentic and the questions are well sequenced. Therefore, it does not make test taker more difficult to answer the question.

The lowest number of the criteria checklist was on the test plan category. The teacher only fulfilled 25% of test plan criterion. It indicates that the teacher did not prepare the test very well. It corresponds perfectly with the finding coming from the interview. From the interview, the teacher admitted that he did not spend much time to construct the reading test. After test plan category, it is followed by relevance category which gained 50% of the criteria. For the adequacy of test item, it reached 86%. In the category of adequacy of test

items, only one criterion did not meet by the first reading test. The overall quality of this reading test can be categorized 'fairly good' as it reached more than 77% of the criteria.

Table 1. *The finding from the first teacher*

Reading test 2				
Category	Rater 1	Rater 2	Rater 3	Average
Test plan	25%	25%	25%	25%
Relevance	100%	50%	100%	83%
Balance	100%	100%	100%	100%
Efficiency	100%	100%	100%	100%
Validity	100%	100%	100%	100%
Reliability	100%	100%	100%	100%
Adequacy of the test items	86%	86%	86%	86%
Technical	100%	100%	100%	100%
Rating scale				87%

The finding of the second reading test from the first participant was rather different to the previous reading test. The test met 87% of the criteria checklists. The 87% is considered a "good" test. This indicated that the teacher has implemented several criteria of a good reading test.

The highest percentage laid on the category of relevance, balance, efficiency, validity, reliability, and the technical of test. Those criteria met perfectly the instrument. This reflects the teacher's ability in reading test construction was quite sufficient.

In spite of high percentage, the table above indicated that teacher lack of ability in planning the reading test. It implies that the teacher needs to allocate the time for preparing the reading test.

Table 2. *The finding from the second teacher*

Reading test 1				
Category	Rater 1	Rater 2	Rater 3	Average
Test plan	25%	25%	25%	25%
Relevance	50%	50%	50%	50%
Balance	50%	50%	50%	50%
Efficiency	50%	50%	50%	50%
Validity	33%	33%	33%	33%
Reliability	67%	67%	67%	67%
Adequacy of the test items	29%	29%	29%	29%
Technical	100%	100%	100%	100%
Rating scale				51%

Luthfiyatun Thoyyibah

Teachers' competence in a reading test construction

Table 3 shows that the reading test 51% met the criteria checklists. The rating scale is considered "very poor" in the rating scale proposed by Mahoney, Powell, & Finger (1986).

This finding was different with the two previous reading tests which gained higher percentage. Only one category reached 100%. It was the technical of the test. It implies that the test was free mistyping. Then, the instructions were given clearly and complete. Furthermore, the exam copy was legible. It indicates that this reading test fulfilled all the criteria of technical.

The rest of categories reached around 50%. The lowest percentage were on the test plan and adequacy of the test items. It only met 25% and 29% of the criteria checklists. Another category which gained not so much percentage was validity. Validity is one of the important points in testing. That is why, the table above indicated that this reading test was considered as 'very poor' reading test. It needs improvement in many aspects. One at a time, it indicates that teacher lack of competence in constructing a valid reading test.

Table 3. *The finding from the second teacher*

Reading test 2				
Category	Rater 1	Rater 2	Rater 3	Average
Test plan	25%	25%	25%	25%
Relevance	50%	50%	50%	50%
Balance	50%	50%	50%	50%
Efficiency	50%	50%	50%	50%
Validity	33%	33%	33%	33%
Reliability	67%	67%	67%	67%
Adequacy of the test items	71%	71%	71%	71%
Technical	100%	100%	100%	100%
Rating scale	56%			

Even though the fourth reading test was constructed by the same participant, the fourth reading test was slightly better than the first reading test by the second participant. Table 4 shows that 56% of category met the criteria checklists. Through that number, the reading test was similar to the previous one. It was categorized as 'very poor'.

The highest percentage was in the criteria of technical. It met 100% criteria checklists. The teacher paid attention to some aspects that probably teacher assumes it did not affect the reading test quality.

The percentage assumes that teacher needs for some improvement. The improvement can be done in the aspect of test plan and validity. A well-planned test is the starting point to the success of any tests, includes a reading test. Besides test plan, the test can be categorized into less valid. That was less valid because among three descriptors of validity category, the test failed to meet the two of them.

Table 4. *The finding from the third teacher*

Reading test				
Category	Rater 1	Rater 2	Rater 3	Average
Test plan	25%	50%	50%	42%
Relevance	50%	50%	50%	50%
Balance	100%	100%	100%	100%
Efficiency	100%	50%	100%	83%
Validity	100%	100%	100%	100%
Reliability	100%	100%	100%	100%
Adequacy of the test items	71%	71%	71%	71%
Technical	100%	100%	100%	100%
Rating scale	81%			

The last reading test by the third participant gained 81%. Based on the instrument used in this study, the quality of the reading test is considered 'good'. Even though it did not show excellent, the teacher has already known some basic competence in constructing a reading test. From all the reading tests, it can be seen that all the participants were lack of ability in planning the test. From the relevance category, most teachers posed the reading test without mentioning or giving information of the types of text in the test.

The elaboration of all tables above is presented in the following paragraphs on each category. The elaboration of tables is based on criteria checklists by justifying with the relevant theories posed in chapter two.

Test plan

The first category in the criteria checklists is test plan. Test plan category comprises of five criteria. Those four criteria were related to whether the reading tests were high stakes or not, required test takers' identity, time allotment of the tests, and the scoring is clearly mention or not. In total, from the three participants, 25% of those criteria were met teachers' reading test. Except for the third teacher,

each teacher in each reading test received the same percentage of the test plan category by chance.

Relevance

The text in reading test should be determined by the course objective. The texts in the second reading test by the first participant were relevant to the syllabus design. It was mentioned that advertisement and announcement were given at the beginning of the semester.

There have been proven before administering the test, all the topics of the text were based on specific basic competence. Actually, these reading tests were included to achievement test, especially the formative achievement test. That condition fulfilled the principles of achievement test (Hughes, 2003). Hughes (2003) asserted that achievement tests should be based on course content. Hopkins, Stanley, & Hopkins (1990) went on to say that all good achievement tests should be based on either explicit or implicit objectives or topics reflected in the syllabus. Every course has its own objective. The course objective lies on the syllabus and breaks down into several indicators. Those that become the consideration of teachers in constructing or taking the text for the reading test.

Balance

Similar to "Relevance" category which covering two criteria, the third category comprises of two criteria. The "Balance" category comprises of the existence of sections in the reading test and the balance order in the text and questions. The separation numbers on sections became one of the requirements in the criteria checklists used as the instrument of this study. In one of the most familiar English proficiency test, the TOEFL test, there are sections. The function of section is to divide from one type of question to another type of question. It is also to discriminate among the four language skills. The last reading test by the third teacher showed that sections had function on separating the multiple-choice question to the short answer question.

The first participant preferred to use authentic text. The teacher further explained that creating a text was time consuming. The teacher must devote the time for constructing the text first, the set of questions. The teacher also did not know the validity of the text they made. Therefore, the teacher needed to take the authentic one. Moreover,

the third participant argued that an authentic text provided authentic cultural information and gave the students exposure to real language. It indicated that the teacher already known about the knowledge of text authenticity.

Validity

Each participant had different result on validity. The highest number was gained by the first participant and followed by the third and the second participant. The highest result reached the number of 100%. The lowest number performed by the teacher was 33%. It indicated that from 3 categories, one only met the criteria checklists of a good reading test. The elaboration of each criterion of validity is explored below.

Reliability

The two reading tests gained 100% of reliability category. The first criteria in reliability category were about the easy questions to be understood by the students. From all the reading tests constructed by the teachers, all the reading tests met the first criteria. All reading tests contained easy questions to be answered by the students. All reading tests used a simple utterance in questioning, like "How long did Nova work in an electronics store?" Through the simple questioning, the students did not need to think in a complicated way of understanding what is being asked. If it used the word "How long", for sure, it related to time or frequency. Another sample of question was asking in the form of yes/no question; "Does Intan want to be Alia's friend?" That question was actually having a drawback of allowing students guess the answer without reading and comprehending the text. The students might answer by Yes or No. Therefore, it is recommended that teacher need to add follow up question like "Why?" instead of only asking yes/no question. By the follow up question, reading and comprehending the text was needed. There was a need for students to find out the reason of a more specific question.

Adequacy of the test items

Different from the first criteria, not all reading tests contained this kind of question. Accidentally, only reading tests with the multiple-choice question consisted this kind of question. The way teacher assessed "inferences" can be seen from one of the examples; "Which of the following is not a duty of

the advertised job?” Actually, the text mentioned explicitly the job that was being advertised. But, rather gave the explicit word, the teacher intended to choose another word with the similar meaning that was “*advertised job*”. The teacher was indicated aware of making inference question from the context by referring the topic rather explicitly mentioned the word. The context of the text was about job advertisement. This inference question was seen having lower level of thinking than the two-previous type of question. But this skill is closely related to literal meaning as proposed by Day and Park (2005). Questions asked literal meaning also marked the validity of the test as it asking the explicit information on the text. This criterion was characterized in several questions mentioned. First question is “*Which of the following is not one of the major divisions of horse breeds?*” The teacher asked the major divisions of horse breeds which was displayed in the text. Because that was a multiple-choice question, students had already given clues to answer it. Students were required to scan for a specific explicit major division of horse breeds. This kind of questions implicitly measures the teacher ability to ask the students capacity in implementing one of the reading strategies. The following criterion was about the question of weaving together ideas in the context. All the reading tests made by the teachers did not meet this criteria. It indicated that the teachers did not have the ability in creating this kind of question.

The last criterion is that the test item contains level of difficulties. This criterion was a summary of all the criteria of adequacy of test items. It can be inferred that most of the reading tests contain different level of difficulties as it consists of several characteristics of questioning, like drawing literal meaning, making inference, and etc. the reading test which consist of the essay which including short answer question did not show level of difficulties. Those questions were identified too easy for the grade intended. It needs variation in making short answer question.

Technical sound of the test

This category comprises of three criteria, those are “test was free of typing errors”, “instructions are clear and complete”, and “exam copy was legible – attractive”. All the criteria in the technical sound of

the test were met the criteria checklists of a good reading test.

One of the instructions states “*Complete the blanks using the words in the list!*” That instruction asked the students to fill the blanks with the available words on the list. That instructions led the test takers on how to do with the following questions. Another example was taken from one of the reading tests; “*Read the following text to answer questions number 9-10*” Instruction needs to be understandable and not too wordy. It dealt with efficiency in doing the test. Understanding the instruction should not be more difficult than understanding the questions. Therefore, the students as the test takers would not spend much time on comprehending the instruction. From the examples of test’s instructions, it was indicated that all teachers have the ability in constructing the clear and complete instruction. Instructions they made were clear and straight to the point. The instructions did not use unfamiliar words that made the students as the test takers difficult to comprehend and answer the questions.

These facts indicated that teacher have already aware of the concept of technical things in administering a test, especially a reading test. It related to the concept of practicality proposed by Nation (2008). Nation (ibid) points out that the reading test that test constructor should make the test easy to recognize in order to answer it.

CONCLUSION

The study sought to investigate the process for developing reading tests by three English teachers at a vocational school, as well as the performance of the teachers' reading tests.

All the teachers of English participated in the research accomplished a few stages in the development of a reading test, including defining the basic mastery of the materials, selecting the most appropriate text, defining the kind of questions, trying to decide how many numbers to include in the reading test, and grading. These exemplified that the teacher of English at vocational school are capable of creating a reading test.

The initial thought was to ascertain the materials on basic competence in a syllabus. The teacher of English determines one or more fundamental competencies for every assessment. The complexity of the materials covered dictated the choice of basic competence. As a result, the purpose of the test was

based on the learning goal as stated on the syllabus. The next step was to select text. Selecting the appropriate text has become one of the most important aspects of a reading test because it influences the reliability and validity of the test (Nation, 2008). The kinds and quantity of questions in the test were then decided upon by teachers. The selection of text, the kinds, and the quantity of the questions were all circular. They were related to one another. The number of questions was affected by the kinds of questions and the length of the text determined the quantity of questions. The final step was to assign a score. That had something to do with the students' grades. The students' grades were used as one of considerations in teaching. They might have noticed both the accomplishment and failure of teaching. As a result, the teacher of English there thought that it was necessary to create the scoring rubrics before administering the test.

According to the findings and discussions, some teachers are capable of developing reading tests at the vocational school level, while others are not. The first teacher created "good" and "poor" reading tests, the second performed two "very poor" reading tests, and the third proposed one "good" reading test.

The reading tests administered by the teachers revealed some strengths and weaknesses. The reading test was created by all the teacher of English based on the basic competence found on the syllabus. Each fundamental competence has its own goal with its own set of materials. This is consistent with Hakim & Irhamsyah (2020), who state that a good achievement test should be designed in accordance with the syllabus design in the teaching materials column. Then, using the following questions, it is demonstrated that the majority of teachers have adopted the text's information flow. One of the requirements for a good reading comprehension test has been met.

Furthermore, the teachers have already demonstrated decent skill to ask literal questions, questions that make inferences about a word's meaning based on context, questions to find an answer explicitly, and questions in paraphrase. The ability to plan, create, assess, and use the language test in ways that are appropriate for a given aim, context, and group of test takers has proven teachers' capability. According to Bachman and Palmer in Bachman and Palmer (2015), these were

the most important requirements for teachers' competence in language testing.

Nevertheless, some flaws were related to the reading tests. Foremost, some teachers were unaware of the text's authenticity. Authentic text, on the other hand, becomes one of the most important factors in selecting the right text for a reading test. Some teachers wrote their own texts that were too simple and easy for the vocational level. As a result, it did not provide students with information about the actual situation in the target language. Second, the questions in some reading test used the identical words from the text. The use of it classified into non-comprehension question. It also reduced the validity of the reading test (Primadani, 2019). It only requires the students to cut and paste the answers to the questions.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading* (1st ed.). Cambridge University Press.
- Bachman, L. F. (2015). Justifying the use of language assessments: Linking test performance with consequences. *JLTA Journal*, 18(0), 3–22. https://doi.org/10.20622/jltajournal.18.0_3
- Burke, J. (1999). *I hear america reading: Why we read, what we read*. Portsmouth, NH: Heinemann.
- Caena, F. (2020). 'Professional development of teachers' literature review quality in teachers' continuing professional development. *January 2011*.
- Chen, A.H., Halilah, N. B., & Shauqiah, J. (2017). The Development of SAH reading passage compendium: a tool for the assessment of reading performance related to visual function. *International Education Studies*, 10(12), 30. <https://doi.org/10.5539/ies.v10n12p30>
- Clifford, H., & Parry, K. (2014). *From testing to assessment*. Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. Taylor and Francis.
- Day, R. R., & Park, J. (2005). Developing reading comprehension questions. *Reading in Foreign Language*, 17(1), 60–73.
- Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 3–11. <https://doi.org/10.1177/2372732218816339>
- Firdaus, A. (2017). Looking at the link between emotional intelligence and reading comprehension among senior high school students. *Edukasi: Jurnal Pendidikan Dan Pengajaran*, 4(2), 18–28.
- Hakim, L., & Irhamsyah, I. (2020). The analysis of the teacher-made test for senior high school at State

Luthfiyatun Thoyyibah

Teachers' competence in a reading test construction

- Senior High School 1 Kutacane, Aceh Tenggara. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan Dan Pengajaran*, 21(1), 10. <https://doi.org/10.22373/jid.v21i1.4120>
- Hartell, E., & Strimel, G. J. (2019). What is it called and how does it work: examining content validity and item design of teacher-made tests. *International Journal of Technology and Design Education*, 29(4), 781–802. <https://doi.org/10.1007/s10798-018-9463-2>
- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation* (7th ed.). Prentice Hall 1990.
- Hughes, A. (2003). *Testing for Language Teachers* (3rd ed.). Cambridge University Press.
- Jahan, K., & Ashraf, S. (2020). The process of developing a reading test: A review article. August. *IJPR*, 24(7). <https://doi.org/10.37200/IJPR/V24I7/PR271078>
- Ma'rifatullah, M., Ampa, A. T., & Azis, A. (2019). Teachers' pedagogic competence in teaching English At SMAN 1 Sanggar in Bima. *Exposure: Jurnal Pendidikan Bahasa Inggris*, 8(1), 90–100. <https://doi.org/10.26618/exposure.v8i1.2087>
- Mahoney, G., Powell, A., & Finger, I. (1986). The maternal behavior rating scale. *Topics in Early Childhood Special Education*, 6(2), 44–56. <https://doi.org/10.1177/027112148600600205>
- Maolida, E. H., & Anjaniputra, A. G. (2017). Mengenai prinsip dan teknik mengajar bahasa Inggris pada anak bagi para guru bahasa Inggris. *Journal of Empowerment*, 1(2), 153–166.
- Nation, K. (2008). Learning to read words. *Eric*, 1. <https://doi.org/https://doi.org/10.1080/17470210802034603>
- Pazilah, F. N. P., Hashim, H., & Yunus, M. M. (2019). Using technology in ESL classroom: Highlights and challenges. *Creative Education*, 10(12), 3205–3212. <https://doi.org/10.4236/ce.2019.1012244>
- Primadani, P. (2019). *Investigating the authenticity of English try-out reading test items a case of ninth graders of SMP Negeri 29 Semarang in the academic year of 2018 / 2019* (Unpublished thesis, Unnes).
- Rahmatun, N., & Helmanda, C. M. (2020). Analysis of reading comprehension final test At English Department of Muhammadiyah Aceh University. *Getsempena English Education Journal*, 7(1), 72–85. <https://doi.org/10.46244/geej.v7i1.987>
- Razali, K., & Jannah, M. (2015). The Comparison between national final examination test items and English teacher made-test items of 2010 and 2011. *Al-Ta Lim Journal*, 22(1), 10–22. <https://doi.org/10.15548/jt.v22i1.116>
- Saefurrohman, S., & Balinas, E. S. (2016). English teachers classroom assessment practices. *International Journal of Evaluation and Research in Education (IJERE)*, 5(1), 82. <https://doi.org/10.11591/ijere.v5i1.4526>
- Santy, N. P. L., Dewi, N. L. P. E. S., & Paramartha, A. A. G. Y. (2020). The quality of teacher-made multiple-choice test used as summative assessment for English subject. *Prasi*, 15(02), 57. <https://doi.org/10.23887/prasi.v15i02.25560>
- Saraceno, L. M. (2019). *Disciplinary literacy pedagogical content knowledge (DLPCCK) today : An exploration of disciplinary literacy pedagogical content knowledge of middle and high school science, social studies, and English language arts* (Theses and Dissertation, Rowan University) . 1–242.
- Shohamy, E., Or, L. G., & May, S. (2017). *Language testing and assessment* (3rd ed.). Springer Netherlands.
- Susanti, E. (2020). A study on English department students' reading barriers at English department UNP. *Journal of English Language Teaching*, 9(2), 391–399.
- Wallace, M., & Wray, A. (2015). *Critical reading and writing for postgraduates* (J. Bowen (ed.); Vol. 7, Issue 1). Sage Publication Ltd. https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civilwars_12December2010.pdf%0Ahttps://think-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625
- Zulmaini, E. A. (2021). Teaching and learning process of test-taking strategies in answering reading comprehension section. *ELT Forum: Journal of English Language Teaching*, 10(2), 113–124. <https://doi.org/10.15294/elt.v10i2.43281>