



**PROBLEMS FACED BY STUDENT TEACHERS
IN CONSTRUCTING TESTS**

THESIS

*Submitted to the English Teaching Department of Tarbiyah
and Teacher Training Faculty of IAIN Batusangkar
As a Requirement to obtained bachelor Degree in
in English Teaching*

RIDWAN FADHIL

Reg. No.14 104 06

**ENGLISH TEACHING DEPARTMENT
TARBIYAH AND TEACHER TRAINING FACULTY
STATE INSTITUTE FOR ISLAMIC STUDIES (IAIN)
BATUSANGKAR**

2021

SURAT PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini:

Nama : Ridwan Fadhil

Nim : 14104065

Tempat/tanggal lahir : Batusangkar/ 29 April 1996

Fakultas : Tarbiyah dan Ilmu Keguruan

Jurusan : Tadris Bahasa Inggris

Menyatakan dengan sesungguhnya bahwa skripsi yang berjudul "PROBLEMS FACED BY STUDENT TEACHERS IN CONSTRUCTING TETS" adalah benar karya saya sendiri bukan plagiat.

Apabila di kemudian hari terbukti bahwa karya ilmiah ini plagiat, maka saya bersedia menerima sanksi sesuai dengan ketentuan perundang-undangan yang berlaku. Demikian pernyataan ini saya buat dengan sebenarnya untuk digunakan semestinya.

Batusangkar, 25 Februari 2021

Saya yang Menyatakan,



Ridwan Fadhil
NIM. 14104065

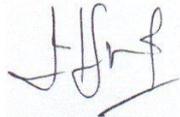
THESIS ADVISOR APPROVAL

The advisor of **RIDWAN FADHIL**, Reg No. 14 104 065, approves that the thesis entitled "**PROBLEMS FACED BY STUDENT TEACHERS IN CONSTRUCTING TESTS**" has fulfilled the requirements to thesis examination

This approval is granted and used appropriately.

Batusangkar, January 2021

Advisor

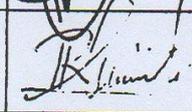


Nina Suzanne, M.Pd

NIP. 19790915 200604 2 006 003

THESIS EXAMINERS' APPROVAL

The thesis written by **RIDWAN FADHIL**, Reg. No. 14104065, judul "**PROBLEMS FACED BY STUDENT TEACHERS IN CONSTRUCTING TEST**", has been examined by board of examiners of English Teaching Department of Tarbiyah and Teacher Training Faculty of IAIN Batusangkar on Tuesday, February 09, 2021 and approved to accepted as requirement to obtain Bachelor Degree in Teaching English

NO.	NAMES	POSITIONS	SIGNATURES	DATE
1.	<u>Nina Suzane, M.Pd</u> NIP. 197909152006042006	Advisor		25-2-2021
2.	<u>Dr. Suswati Hendriani,</u> M.Pd, M.Pd NIP. 196609141992032003	Examiner		25-02-2021
3.	<u>Rahmawati, M.pd</u> NIP. 197009282005012009	Examiner		27/02/2021

Batusangkar, Februari 2021
Approved by,
Dekan of Tarbiyah and
Teacher Training Faculty
State Institute Islamic Studies (IAIN)
Batusangkar




Dr. Adripen, M.Pd.
NIP: 196505041993031003

ABSTRAK

RIDWAN FADHIL, NIM 14104065, Judul Skripsi: “**PROBLEM FACED BY STUDENT TEACHERS IN CONSTRUCTING TEST**” Jurusan Tadris Bahasa Inggris , Fakultas Tarbiyah dan Ilmu Keguruan, Institut Agama Islam Negeri Batusangkar.

Penelitian ini dilatarbelakangi oleh beberapa analisis soal yang dibuat oleh mahasiswa PPL Jurusan Tadris Bahasa Inggris, Fakultas Tarbiyah dan Ilmu Keguruan, Institut Agama Islam Negeri Batusangkar yang tidak memenuhi kriteria soal yang baik. Tujuan dari penelitian ini adalah untuk mengetahui masalah yang dihadapi oleh mahasiswa PPL dalam menyusun sebuah soal bahasa Inggris.

Jenis penelitian ini adalah Kualitatif, Informan dalam penelitian ini adalah 8 orang mahasiswa Jurusan Tadris Bahasa Inggris IAIN Batusangkar yang melaksanakan PPL di Sekolah Menengah di Tanah Datar tahun angkatan 2019/2020. Informan dipilih menggunakan *purposive sampling*. Pada penelitian ini, peneliti berperan sebagai instrumen kunci. Teknik pengambilan data pada penelitian ini menggunakan teknik wawancara dan menggunakan alat pendukung lainnya seperti, alat perekam, dan dokumentasi. Untuk menguji keabsahan data peneliti menggunakan *time triangulasi* untuk mendapatkan informasi secara lebih jelas terkait masalah yang diteliti. Selanjutnya penulis menggunakan teori Miles and Huberman (1994) yang meliputi reduksi data, penyajian data, serta verifikasi data dan kesimpulan.

Hasil penelitian menunjukkan bahwa masalah yang dihadapi mahasiswa PPL jurusan Tadris Bahasa Inggris dalam membuat soal, yaitu: 1) *Getting Difficulty Construct Valid Test*, 2) *Getting Difficulty to construct various and appropriate level of thinking in test*, 3) *Not sure about the usability of the test*, 4) *Lack of Preparation*, 5) *Getting difficulty to find suitable texts for testing reading*, 6) *Grammatical Problem*, 7) *Getting Difficulty to develop the test from table of specification*.

TABLE OF CONTENTS

COVER PAGE	i
SURAT PERNYATAAN KEASLIAN	ii
THESIS ADVISOR' APPROVAL	iii
THESIS EXAMINERS' APPROVAL	iv
ABSTRAK	v
TABLE OF CONTENTS.....	vi
LIST OF APPENDICES.....	Error! Bookmark not defined.
CHAPTER I INTRODUCTION.....	1
A. Background of the Problem.....	1
B. The focus of the Research	7
C. Research Question.....	7
D. Purpose of the Research	7
E. Definition of the Key Terms	7
F. Significance of Research.....	8
CHAPTER II REVIEW OF RELATED LITERATURE	9
A. Review of Related Theories	9
1. Language Testing	9
a. Definition of Language Testing.....	9
b. Types of Tests	9
c. Criteria of Good test	12
d. Teacher Competence in Construct test	23
2. Problem in Constructing Test.....	23
B. Review of Relevant Study.....	24
CHAPTER III RESEARCH METHODOLOGY	25
A. Research Design.....	25
B. Research Setting.....	25
C. Research Informants.....	25
D. Research Instruments	26

E. The technique of Data Collection.....	26
F. Checking Data Trustworthiness	27
G. The technique of Data Analysis	27
CHAPTER IV RESEARCH FINDING AND DISCUSSION	29
A. Research Findings	29
1. Description of the Research.....	29
2. Data Analysis.....	30
B. Discussion	35
CHAPTER V CONCLUSION AND SUGGESTION	37
A. Conclusion.....	37
B. Suggestion	37
BIBLIOGRAPHY	39

CHAPTER I

INTRODUCTION

A. Background of the Problem

Tests are series of questions or assignments that must be done by students whose results are used to measure students' abilities. As a measurement tool in evaluation, the test provides quantitative data. Meanwhile, the assessment of test is more qualitative by assessing students according to the quality of their learning outcomes. Besides, the test can compare things based on certain sizes or criteria. In the evaluation of learning, there are measurement and assessment activities.

Evaluation and assessment are two different terms in the scope of education. The difference between evaluation and assessment. Evaluation is the final process of assessment, while assessment is a process of measuring an object. This process will be reviewing challenged as a measure of the process that has done. All in all, something is wrong if the test taker wants to assess a particular object without preceded by previous measurements.

The purpose of the test itself is to see whether the student learning process is following the learning implementation. Examine student learning outcomes whether there are deficiencies or not in the learning process, look for solutions to natural student deficiencies, and infer how students mastered competencies are student teachers.

Terms and conditions for example, evaluations must use an appropriate measuring instrument (Valid) in accordance with their objectives, whether to measure or just assess. Then other requirements, namely the test equipment must be reliable or produce the same results (consistent), and the final evaluation requirements, namely the evaluation must be practical or easy to use not difficult for educators or students.

To prepare future English teacher candidates who are able to respond to needs in the world of work, the experience is certainly needed. For this

reason, the English Study Program also hold a teacher training program (PPL) where students are sent directly to the field to teach in several schools such as junior and senior high schools in Tanah Datar. At school, student teachers are asked to practice every skill they acquire during college, including measuring students' language skills. In this case, student teachers are required to make their own tests related to the material they have taught.

Based on previous research by Maryumah Hejji Alanazi "A Study of the Pre-Service Trainee Teachers Problems in Designing Lesson Plans"(Department of Curriculum and Educational Technology Northern Border University Arar, Saudi Arabia) The findings regarding the lesson plan designing conclude that the use of information and resources of the trainee teachers vary depending upon their knowledge. This current finding showed a clear picture of the pre-service trainee teachers' interest and competency level. This research finding may help the educators and trainer to think and execute the far-reaching teaching strategies to motivate the trainees to excel in their plan designing tasks.

Pre-service teachers used a broad selection of information channels and resources in their lesson plan design. Few trainees actively seek information for their lesson plans from external information sources, but the entire information taken from web sources. This is the foremost objectionable feature among the trainees.

These external sources cannot provide entire information regarding a particular lesson what they desire to teach. However, out of laziness or their own problems, trainees are interested to take the readymade materials and reluctant to prepare on their own. At the time of teaching, the trainee teachers perplexed and fail to execute the assigned task. Eventually, they are unable to deliver the teaching task effectively and lose their confidence. It is very important to mention that; the creative skills of the pre-service trainee teachers were at best as they designed the activities for their teaching practice session. Majority of the teacher trainees were good in preparing and arranging aids. This effort clearly showed their creative skills. As an observer, the

researcher found two contrast practices among the pre-service trainee teachers.

One of trainees were competent enough to attract the class because their engagement plan was supple and adept. Hence, it was easy for them to execute the plan and accelerate the class. Moreover, the trainees conducted the activities suitable to the standard and the interest of the children using the rubrics.

The researcher could notice and understand the material which was prepared and designed by them showed their communicative competence. It is important to give a optimistic note that these pre-service trainee teachers' lesson plan designing and execution were commendable. It was very obvious the trainees who were not successful in their teaching were using the off-the-rack materials without any rehearsal. Actually, they are all ignoring their literary skill and purely relying on the prefabricated materials.

During the first few days of training the trainees felt nervous and out of excitement became anxious. It was observed during the whole training period that the majority of the students were struggling in administering their plan. The researcher could feel there were some problems in their plan and its execution. But, the pre-service trainee teachers' perception regarding specific information cannot be indicated only in terms of their current lesson plan. They may instantly access well-known information resources to sufficiently complete the designing lesson plan task. The readymade teaching kit and jerry-build lesson planning materials are extremely simple and easy to use, but it impedes the students from thinking and writing on their own. Moreover, they felt that this practice , the training session is merely to score marks and complete the course.

This clearly indicates that the pre-service trainee teachers failed to recognize the importance of the training sessions. Another important obstacle is designing the lesson plan; it is a very new notion for them. In this context the main issue is time. So, just to teach them to manage the time is very important. In addition to that, making them understand the information

acquisition is a very important factor in the process of designing the lesson plan also which in turn will help them to design the task on their own. Predominantly, the findings of this research offered insights into the pre-service trainee teachers' point of view of designing lesson plans. The findings revealed that although most of the pre-service trainee teachers are aware of the importance of lesson planning, they found it difficult to develop a lesson plan. It is assumed that the pre-service trainee teachers are not spending considerable time to design the lesson plan. If they, pre-service trainee teachers spend more time on designing lesson plans, they will become meticulous in designing and excellent in executing it. So, the study concludes that the trainers must focus and give special attention to the trainees to design the lesson plan effectively. Also, insist the pre-service trainee teachers spend enough time to plan their lesson. From the results of previous studies above the researcher interested to continuing research on pre-service teachers or student teachers but with a different focus about problems faced by student teachers in constructing tests.

The reasons why the researcher chooses to research for highschool students is in the current curriculum (Kurikulum 2013), students learn English for the first time in junior high school. For some elementary school, they did not learn English intensively. When the students are in junior high school, they will face some problems, like lack of understanding the vocabulary; lack of students interest in learning English, and English as the foreign language, those are the reasons why English test are difficult for them (even it is the daily test or midterm test); etc.

Based on the analysis of some tests that the researcher get from student teachers of English Department Students at IAIN Batusangkar 2019/2020 academic year, the researcher found that the teacher students were asked to make an exercise and midterm test. Meanwhile, the final test was conducted by the English Teacher Association. In fact, many student teachers confused to conduct the test, because it was their first time conducting the test at the school.

Furthermore, when they conducted the test, they just took the some tests from the internet because they were difficult to conduct the test. Then, they improve the items and adjust it with the material that they teach. It happened several times. The reasons why did they do that because they did not have any experiences in teaching junior high school students. They also did not know the students' level of English proficiency. It made they did not know what kind of test that the would give to the students. They also did not have any efficient time for the learning process.

Another finding was the students choose an essay test as the companion of multiple choice test. But, they mention the test as an essay. In chase, they choose short answer test, not an essay. An essay test, should have explanation answer, it means that students should explain or describe their answer based on the question. Meanwhile, the student teachers just give the question based on the passage. Also, students just give an answer that they could find on the passage. The question of the test asked students to answer by find and found, not explain or describe.

In middle and high schools in Indonesia, there are several tests or examinations that students must take. They are a) daily exams. Daily exams are also known as daily tests (UH). Its function is to determine the level of mastery of student material on one or more concepts/themes / subject matter. Daily exam scores will be combined with assignments and other grades into the student's daily grades. This grade will be an assessment element on the student report card. b) Midterm Exams, Midterm Exams (UTS), held in the midterm. Serves to determine the level of mastery of students of several concepts/themes/subject matter. Midterm test results will determine the student report card grades. c) Semester examination. The semester exam is a test conducted at the end of the semester after students complete all material in the semester. The semester exams are divided into two semesters (odd semester) and the second-semester exam (even semester). The even semester exam is also called the grade promotion test. d) School exams. School End Exams (UAS) are held after students complete all learning material. Exams

conducted by students of the last class at the school education level serve to determine the graduation of students from a school. e) National Examination (UN) is a national exam held on certain subjects. The function of the National Examination as quality control of national education. In addition, it also determines the graduation of students from a school. Then it becomes an administrative requirement for students to continue their education to the next level.

As one of the tools used to measure artificial tests, Teachers play a very important role because the test is part of the learning process. As Madziyire (2010) argues, teacher-made tests help teachers to identify content (knowledge or skills) that have been mastered by students and the teacher knows through the results of his tests in areas where students experience difficulties and then find ways to overcome difficulties so that these students can do better. The results of teacher-made tests allow teachers to assess their strengths and weaknesses.

Then, the researcher looked at the tests that they made before. The first test was a daily test (PH). The topic of the test was about daily expression. The teacher students chose multiple choice short answer, and fill in the blank test. There were three groups of the test. One multiple-choice test, one short answer tests, and one fill in the blank tests. The choice of the test were only three options. While for junior or senior high school students, they can give four or five options. It seems like they did not know the next possible or another answer that almost relates to the right answer.

The second test was a daily test (PH) about daily expression and greeting cards. It was ten exercises of multiple-choice and five exercises of short answer. The student teachers state that the B test was an essay but it was a short answer test. It seems like the student teachers did not know which one is an essay and short answer test.

Based on preliminary interviews and analyses above, student teachers said that the teacher asked them to make their own tests and one of them was the Midterm test. In conducting this test, students claim that they have

difficulty understanding KD and KI and it is also difficult to determine the level of difficulty of the question. Some of them said that they tend to copy the questions from the internet rather than create their own. Furthermore, others try to make their own test.

Another problem is the teacher does not check the test from student teachers. Based on the explanation above, it can be concluded that making a test is not an easy thing especially for student teachers who do not have experience with it. That the reason why researchers are interested to conduct research about problems faced by student teachers English Department of IAIN Batusangkar in constructing English test for high school in Tanah Datar.

B. The focus of the Research

Based on the background of the problems above, there are several problems that can be studied. In this research, the researcher focuses on problems faced by student teachers in constructing test for highschool in Tanah Datar.

C. Research Question

Based on the focus of the research above, the research question for this research is "what are the problems faced by student teachers of IAIN Batusangkar in constructing English test for high school in Tanah Datar? "

D. Purpose of the Research

Based on the above research question, this study tries to find out what problems are faced by student teachers of IAIN Batusangkar in constructing English test for high school in Tanah Datar.

E. Definition of the Key Terms

To avoid misinterpretation of the terms used in this study, the definitions of key terms are:

1. Problem

Problems in this study is situation that need to overcome faced by student teachers of IAIN Batusangkar in constructing English test for high school students in Tanah Datar.

2. Midterm exam and daily test

The midterm exam and daily test in this study refers to the formative tests made and used by student teachers in assessing students' abilities for high school in Tanah Datar in 2019/2020 Academic Year.

3. Student teachers

Student teachers in this study mean that English students at IAIN Batusangkar who participated in the Field Training Program (PPL) in several high schools in Tanah Datar class of 2019/2020.

F. Significance of Research

The results of this study are expected to provide contributions and information to language assessors, students, teachers, and other researchers both theoretically and practically, as follows:

1. Lecturer

The lecturers' especially who teach language testing subjects. Hopefully, this research can provide information about the ability of student teachers in applying the principles of language assessment and their problems in designing tests.

2. Student teachers

The results of this study are hopefully capable of providing benefits for student teachers in developing knowledge and skills in designing tests early midterm for high school in Tanah Datar and find solutions to their problems in designing a test.

3. Other researchers

Hopefully, for other researchers who interested in the same field of study, this research will contribute as a source for further research.

CHAPTER II

REVIEW OF RELATED LITERATURE

A. Review of Related Theories

1. Language Testing

a. Definition of Language Testing

There are many definitions to define what is language testing. First, tests are a sub set of assessment, a genre of assessment techniques. They are prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated Brown (2010: 3). It is a common practice in schools to determine whether a previously taught knowledge has been learned by the students during the instruction in the class. Given this definition, testing is the systematic collection of data about what students know, understand and are able to do in relation to students' achievement of particular goals of learning.

In addition, Bachman (1990:18) states that a test must measure which may be defined as a process of quantifying a test-taker's performance according to explicit procedures or rules. It means that a test should measure what should be measured and the result of a test should quantify the level of language ability of test takers.

Based on the explanation above, it can be concluded that language testing is an active process of measuring students' ability in language. A test should measure what should be measured and the result of a test should quantify the level of language ability of test takers. At the end of the test the teachers will get the information related to the level of student's ability after following the learning process.

b. Types of Tests

The most common use of language tests is to identify strengths and weaknesses in students' abilities. For example, through testing we

might discover that a student has excellent oral language abilities but a relatively low level of reading comprehension. Information gleaned from tests also assists us in deciding who should be allowed to participate in a particular course or program area. Another common use of tests is to provide information about the effectiveness of programs of instruction.

1) Placement Tests

Placement tests assess students' level of language ability so they can be placed in an appropriate course or class. This type of test indicates the level at which a student will learn most effectively. The primary aim is to create groups of learners that are homogeneous in level. In designing a placement test, the test developer may base the test content either on a theory of general language proficiency or on learning objectives of the curriculum. Institutions may choose to use a well-established proficiency test such as the TOEFL®, IELTS™, or MELAB exam and link it to curricular benchmarks. Alternatively, some placement tests are based on aspects of the syllabus taught at the institution concerned (Alderson, Clapham, & Wall, 1995). At some institutions, students are placed according to their overall rank in the test results combined from all skills. At other schools and colleges, students are placed according to their level in each skill area. Additionally, placement test scores are used to determine if a student needs further instruction in the language or could matriculate directly into an academic program without taking preparatory language courses.

2) Aptitude Tests

An aptitude test measures capacity or general ability to learn a foreign or second language. Although not commonly used these days, two examples deserve mention: the Modern Language Aptitude Test (MLAT) developed by Carroll and Sapon in 1958 and the Pimsleur Language Aptitude Battery (PLAB) developed by

Pimsleur in 1966 (Brown, H.D., 2004). These are used primarily in deciding to sponsor a person for special training based on language aptitude. Diagnostic Tests Diagnostic tests identify language areas in which a student needs further help. Harris and McCann (1994) point out that where “other types of tests are based on success, diagnostic tests are based on failure” (p. 29). The information gained from diagnostic tests is crucial for further course activities and providing students with remediation. Because diagnostic tests are difficult to write, placement tests often serve a dual function of both placement and diagnosis (Harris & McCann, 1994; Davies et al., 1999).

3) Progress Tests

Progress tests measure the progress that students are making toward defined course or program goals. They are administered at various stages throughout a language course to determine what students have learned, usually after certain segments of instruction have been completed. Progress tests are generally teacher produced and narrower in focus than achievement tests because they cover less material and assess fewer objectives.

4) Achievement Tests

Achievement tests are similar to progress tests in that they determine what a student has learned with regard to stated course outcomes. They are usually administered at mid- and end-point of the semester or academic year. The content of achievement tests is generally based on the specific course content or on the course objectives. Achievement tests are often cumulative, covering material drawn from an entire course or semester.

5) Proficiency Tests

Proficiency tests, on the other hand, are not based on a particular curriculum or language program. They assess the overall language ability of students at varying levels. They may also tell us how capable a person is in a particular language skill area (e.g.,

reading). In other words, proficiency tests describe what students are capable of doing in a language. Proficiency tests are typically developed by external bodies such as examination boards like Educational Testing Services (ETS), the College Board, or Cambridge ESOL. Some proficiency tests have been standardized for international use, such as the TOEFL®, which measures the English language proficiency of foreign college students who wish to study in North American universities or the IELTS™, which is intended for those who wish to study in the United Kingdom or Australia (Davies et al., 1999). Increasingly, North American universities are accepting IELTS™ as a measure of English language proficiency.

Based on the explanation above there are five types of tests that can be used in construct tests. The first one is placement test assess students' level of language ability so they can be placed in an appropriate course or class, Aptitude test measures capacity or general ability to learn a foreign or second language, Progress tests measure the progress that students are making toward defined course or program goals, Achievement tests are similar to progress tests in that they determine what a student has learned with regard to stated course outcomes, Proficiency test assess the overall language ability of students at varying levels.

c. Criteria of Good test

A good test is constructed by considering the principles of language assessment. Principles of language assessment are used to know if the test is effective, appropriate and useful or in down-to-earth terms of a good test and also these are used to provide the valuable information for the students.

In Addition, Brown (2004) considers five principles of language assessment; they are validity, reliability, practicality, authenticity and washback. Besides, Bachman and Palmer (1997) propose six other

principles of language assessment; reliability, construct validity, authenticity, interactiveness, impact, and practicality. In this article the researcher only focus on the five principles of language assesment, such as validity, reliability, practicality, authenticity and washback.

1) Validity

When teachers come to assessment, they deal a lot with a question of how to measure students' abilities. The question word 'how' implies that teachers should be able to design a measurement to bring up students' potentials as they wish. It is validity. Validity links to accuracy. A good test should be valid or accurate. Some experts have defined the term of validity. For example, the validity of a test is the extent to which it measures what it is supposed to measure. The relationship between test performance and other types of performance in other contexts is considered. According to Brown (2010:22) validity is the degree to which inferences made from assessment results are appropriate, meaningful, and useful for the purpose of the assessment. The test should be appropriate to observe or the test identify for performance that suitable with the course or unit of study being tested. In addition, Bachman (1990) says that validity is about the extent to which meaningful inferences can be described from test scores. It can be said that validity refers to how an instrument can measure something. An assessment tool can be used, if the validity of the test measures what should be measure. For example, the teacher composes reading comprehension test to assess students' reading comprehension skill It can be inferred that when a test is valid, it can elicit students' certain abilities as it is intended to. The valid test can also measure what it is supposed to measure.

To gain valid inferences from test scores, a test should have some kinds of evidence. The evidence of validity includes face validity, content-related evidence, criterion-related evidence,

construct-related evidence, and consequential validity. In the following section, those kinds of evidence are explained in detail.

a) Content validity

content-related validity/content validity. According to Bachman (1990) content validity involves two crucial concepts. Those are content relevance and content coverage. Content relevance refers to the level in which aspects of ability to be assessed which is tested by the appropriate task, indicate the requirement to identify the ability and the test method facets. Content coverage refers to the degree to which the test sufficiently expresses the performance in the target language.

Another aspect of content validity is similar to the content representativeness, which also concerns to the extent which the test accurately samples the behavioral domain of interest. Sugiono (2011:182) says that to test the content validity can be done by comparing between the content of the test and the material that stated in syllabus and lesson plan. Moreover, Arikunto (2009: 67) states that a test has content validity if it measures a certain specific purpose which is related to the subject or material given.

It can be said that this describes about whether the content of the test is appropriate with the materials given to the students or indicators that should be achieved by the students as stated in syllabus and lesson plan. The test will represent what the students have learned and appropriate with the context and situation of target language.

b) Criterion-related Evidence

Comparison between test scores and a suitable external criterion of performance refers to criterion-related evidence. For example, the result of a teacher-made test about past tense is compared to the result of a test of the same topic in a textbook.

There are two types of criterion-related evidence based on time for collection of the external criterion, concurrent and predictive validity. Concurrent validity focuses on using results of a test to estimate current performance on some criterion collected at concurrent time. For example, a teacher-made test design is considered having concurrent validity when it has the same score with an existing valid test like TOEFL. If students have high scores in TOEFL and concurrently have good scores in doing the teacher-made test, it means that the teacher-made test has concurrent validity. On the other hand, predictive validity focuses on using results of a test to predict future performance on some other valued measure collected in the future time. For example, a teacher-made test is administered to some students and they get high scores. It, then, turns out that by the end of teaching and learning process the students still achieve high scores. It means that the teacher-made test has predictive validity. In addition, when a test taker does a particular test from which result he can be predicted to survive overseas, the test also has predictive validity. It can be found in performance test, admissions batteries, language aptitude test, and the like. To examine criterion-related evidence, correlation coefficient and expectancy table are utilized.

c) Construct-related Evidence

A construct-related evidence, so called construct validity, is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. Constructs may or may not be directly or empirically measured. Their verification often requires inferential data. The construction of a test starts from a theory about behavior or mental organization derived from prior research that suggests the ground plan for the test. Before an assessment is built, the creator must review some theories about content of it. He then will get new concept related

to the content of the items. In language assessment, test makers believe on existence of several characteristics related to language behavior and learning. When the test makers interpret the results of assessment on basis of psychological constructs, they deal with construct-related evidence.

For example, scoring analysis for the interview will need several factors: pronunciation, fluency, grammatical accuracy, vocabulary use, and sociolinguistic appropriateness. The justification of these factors lies in a theoretical construct that claims those factors to be major components of oral proficiency. When a teacher conducts an oral proficiency interview that evaluates only two of the factors, the teacher could be justifiably suspicious about the construct validity of the test.

This kind of validity is the broadest among the previous validity. In other words, it covers all kinds of evidence (face, content-related, criterion-related, and other relevant evidence). Although it is endless to obtain construct-related evidence, test makers should list from the most relevant ones.

Construct validity is a major issue in validating large-scale standardized tests of proficiency. Because such tests must adhere to the principle of practicality, and because they must sample a limited number of domains of language, they may not be able to contain all the content of a particular field or skill.

d) Consequential Validity

Consequential validity encompasses all the consequences of a test. This evidence as washback validity. It focuses on the effect of tests with regard to specific uses, e.g. its impact to preparation of test-takers, the effect on the learners (positive or adverse effects), or social consequences of test interpretation and use. For teachers, consequential evidence is important. They can

judge test scores and use the judgment to improve learning. For stakeholder, this evidence leads to development of curriculum.

e) Face Validity

The concept of face validity is that when a test item looks right to other testers, teachers, moderators, and test-takers. In addition, it appears to measure the knowledge or abilities it claims to measure. Heaton argues that if a test is examined by other people, some absurdities and ambiguities can be discovered.

Face validity is important in maintaining test takers' motivation and performance. If a test does not have face validity, it may not be acceptable to students or teachers. If students do not take the test as valid, they will show adverse reaction (poor study reaction, low motivation). In other words, they will not perform in a way which truly reflects their abilities.

The face validity will likely be high if learners encounter:

- (1). a well-constructed, expected format with familiar tasks,
- (2). a test that is clearly doable within the allotted time limit,
- (3). items that are clear and uncomplicated,
- (4). directions that are crystal clear,
- (5). tasks that relate to their course work (content validity), and
- (6). a difficulty level that presents a reasonable challenge.

To examine face validity, no statistical analysis is needed. Judgmental responses from experts, colleagues, or test takers may be involved. They can read thoroughly to the whole items or they can just see at glance the items. Then, they can relate to the ability that the test want to measure. If a speaking test appears in vocabulary items, it may not have face validity.

2) Reliability

Brown (2010:20) state that A reliable test is consistent and dependable. If the teachers give the same test to the same student or matched students on two different occasions, the test should has a

similar result. In addition, Schumacher and McMillan (2001:181) in Indriani says that reliability refers to the consistency of measurement, the extent to which the scores are similar over different forms of the same instrument or occasions of data collection. This means that reliability in language test refers to accuracy, consistency and stability of measurement by a test. Reliability consists of estimating the amount of variation in language test scores that is due to measurement error (Bachman, 1990 in Restrepo and Villa, 2003:67). It is crucial in language assessment because students' performance on a testing process may be influenced by some factors beside the ability that the teacher is measuring (illness, fatigue, poor test conditions, poor test design, and score inconsistencies).

Moreover, Gay and Airasian (2000:169) add that reliability is a degree to which a test consistently measure whatever its measuring. As a result, the result of the test will be stable even though the test will be tested anytime. So, reliability is refers to the consistency of measurement which a test will has the consistently measure whatever it's measuring or the result of the test will be stable whenever the test will be tested.

There are some types of reliability. Those are;

- a) Student related reliability. It is related to the students' physical or psychological factor such as illness, fatigue, bad day, anxiety, etc.
- b) Rater reliability. It has two categories; Inter-rater reliability occurs when two or more scorers have inconsistency possibly for lack of attention to scoring criteria, inexperience, inattention, or even preconceived biases. It comes from the two or more scorer field consistent scorers of the same test. Intra-rater reliability is a common occurrence for classroom teachers because of unclear scoring, criteria, fatigue, bias toward particular good and bad students, or simple carelessness. To measure this reliability the

teacher can read half of the tests before exposing the final scores or grades, then to recycle back through the whole set of tests to guarantee an unbiased judgment. However, in writing skills tests rater reliability is particularly hard to achieve because writing proficiency involves various characteristics that are difficult to be defined (J. D. Brown, 1991).

- c) Test administration reliability. This reliability is based on the condition around the students when and which the test is administered. Since the disturbances from the outside of the classroom, it makes the test unreliable. For example, when the teacher conduct a listening test by using a tape recorder that played items for comprehension, but because of the noise outside the classroom, students could not hear the tape accurately. This was a clear case of unreliability caused by the conditions of the test administration. Other factors of unreliability are such as photocopying variations, the amount of light in different classroom, temperature, and the condition of desks and chairs.

3) Practicality

Practicality refers to the need to ensure that the assessment requirements are appropriate to the intended learning outcomes of learning process and that in their operation they do not distort the learning process; and, that they do not make unreasonable demands on the time and resources available to learner or teacher. According to Brown (2010) Practicality refers to the logistical, down-to-earth, administrative issue involved in marking, giving and scoring an assessment. It is involve, stays with in budgetary limits, can be completed with in appropriate time constraint, has clear directions for administration thus done smoothly, appropriately utilizes available human resources, does not exceed available material resources, and considers the times and effort involved for both design and scoring.

Test that is prohibitively expensive is impractical. A test of language proficiency that takes a student five hours to complete is impractical-it consumes more time (and money) than necessary to accomplish its objective. A test that requires individual one-on-one proctoring is impractical for a group of several hundred test-takers and only a handful of examiners. A test that takes a few minutes for a student to take and several hours for an examiner to evaluate is impractical for most classroom situations. There are several characteristics of impractical tests are: a) these test are excessively expensive b) they are too long c) they require a handful of examiners to administer and score and d) it takes several hours to grade a test.

There are three types of consistency:

- a) over time (test-retest reliability),
- b) across items (internal consistency),
- c) across different researchers (inter-rater reliability).

Assessing test-retest reliability requires using the measure on a group of people at one time, using it again on the *same* group of people at a later time, and then looking at **test-retest correlation** between the two sets of scores. A second kind of reliability is **internal consistency**, which is the consistency of people's responses across the items on a multiple-item measure. In general, all the items on such measures are supposed to reflect the same underlying construct, so people's scores on those items should be correlated with each other. Like test-retest reliability, internal consistency can only be assessed by collecting and analyzing data. One approach is to look at a **split-half correlation**. This involves splitting the items into two sets, such as the first and second halves of the items or the even- and odd-numbered items. Then a score is computed for each set of items, and the relationship between the two sets of scores is examined. **Inter-rater reliability** is the extent to which different observers are consistent in their judgments.

4) Authenticity

Brown (2010) state that authenticity is the degree of correspondence of characteristics a given language test task to features of target language task. The following ways of authenticity are contains language that is as natural as possible, has items that are contextualized rather than isolated, includes meaningful, relevant , interesting topics, then provides thematic organization and the last offers task that replicate real world test. In other side, authenticity us Engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively. The tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field." said Wiggins (1993).

While, Tanner (1997) provides a good summary of the interrelationship between authenticity and learning experiences noting that Authentic assessment presumes that students will produce something that reflects not a narrow, compartmentalized repetition of what was presented to them, but an integrated scholarship which connects their learning housed in other disciplines and which is presented in a setting consistent with that in which the learning is likely to be most useful in the future. In conclusion, the goal of authentic assessment is to determine if student knowledge can be applied outside of the classroom.

An assessment can authentic if it is indicate the real-world situation. In authentic assessment, the teachers do not have to develop an authentic assessment. He or she may already be using authentic tasks in the classroom Because when the teacher already design an authentic assessment. They only need to pay attention more on the measurement and criteria in assessing student's language ability.

5) Washback

A test that provides beneficial washback if a test has positively influences what and how teacher teach, has positively influence what and how learners learn, a test offers learners a chance to adequately prepare. A test should give learners feedback that enhances their language development and it is more formative in nature than summative, a test provides conditioning for peak performance by the learners.

Brown (2010) state Washback is the effect of the testing on teaching and learning. The beneficial washback are positively influences what and how teacher teaches, positively influences what and how learners learn, offers learners a chance to adequately prepare, gives learner feedback that enhances their language development, it is more formative in nature than summative, the last provides conditions for peak performance by the learner.

Washback refers to is the effect of the testing on teaching and learning. The beneficial washback are positively influences what and how teacher teaches, positively influences what and how learners learn, offers learners a chance to adequately prepare, gives learner feedback that enhances their language development, it is more formative in nature than summative, the last provides conditions for peak performance by the learner. In designing a test with a good washback the teachers should give the feedback to the students after assessing their task. For example when the students' performance a speech as their speaking task, after they perform, the teacher should give some comments and suggestion to increase students language ability.

From the explanation above, it can conclude to construct test for students the must be valid or accurate, reliability, practicality of the test, authenticity of the test, and washback.

d. Teacher Competence in Construct test

According to Jhonson in Wina Sanjaya (2008), he state that competency as rational performance satisfactorily meets the objective for a desired condition. Law no.14 about Teachers and the Lecturer stated that competence is a set knowledge, skills and behavior that must be owned by teacher or lecturer according to type ,level and education unit formal at the place of assignment. So, competency includes ability someone in shape knowledge, skills and behavior in carrying out tasks and its authority to achieve aim.

One of the competencies of the teacher is to make a test, the teacher must be able to make a valid test. Suharsimi arikunto in Irhamsyah (2005), teacher made test is a test written and made by a teacher in the school, so the validity and reliability of the tetst is not like the standardiced test. The effectiveness of this type of the test depends on the skill and ability of the teacher in designing the test. This test is based on materials and spesific goals formulated by the teacher for his own class.

2. Problem in Constructing Test

There are several theories about the problem in constructing test. First, according to Osadebe (2015:51), he state that when a teacher constructs a test, it is said to be a teacher made test that is poorly prepared. It means that the teacher students should construct the test in a well prepared. Because,they are sued by their departement to be a a good real teacher in the future. Also, if the teachers have a good preparation in cinstucting test, the students will not confused in understanding the the instruction of the test, the question of the test (essay and multiple choices test), and the option of the test (if it is a multiple choice test). Not only make students understand the test, but also the students do not disturb another, like the students are quiet (because the do not ask another students). By do not construct the test recklessly, the teachers know the students skill and ability in their subject.

B. Review of Relevant Study

Related to this research, the researcher also finds the other researches which are relevant to this research. The previous researches talk about student's opinion.

Suharsimi Arkunto (2005). *The Analysis of The Teacher-Made Test For Senior High School at State Senior High School 1 Kutacane. Aceh*

He found that there are several problems of teacher-made test related to lecturer's corrective feedback. They are: negative and positive opinion.

The similarity between her theses with this research is both of the research talked about students opinion. While, the when Lucky's research talked about student's opinion related to Lecturer's corrective feedback in writing research proposal, this research will talked about Student teacherss of IAIN Batusangkar problems on designing mid term test for high school (SMP/MTs) at Tanah Datar.

Then, a thesis by Ita Widyawati in 2012, entitled "Analisis Kesulitan Belajar Siswa pada Mata Pelajaran Bahasa Inggris Kelas VIII di SMP N 1 Sawahan Madiun Semester Genap Tahun Ajaran 2010-2011. She found that there are several factors of the students difficulties in learning english, such us: cognitive aspect, affective aspect, and psyschomotoric aspect.

The similarity between her theses with this research is both of the research talked about difficulties of English test. While, the when Ita's research talked about student's difficulties related to learning English, this research will talked about Student teacherss of IAIN Batusangkar problems on designing mid term test for high school at Tanah Datar.

CHAPTER III

RESEARCH METHODOLOGY

A. Research Design

This research use qualitative research method. The case study to find out problems faced by student teachers in constructing the test. According to Moleong (2010: 6), qualitative research is research that intends to understand the phenomena experienced by research subjects such as motivation, perception, behavior, holistically, etc. Gay (2012:7) states that qualitative research attempts to investigate deeply into the setting of research to gain an understanding of what is happening, why they are like that, and how participants in the context perceive them.

Dealing with the theory above, this study describe the problems faced by student teachers English Department of IAIN Batusangkar in constructing a test for high schools in Tanah Datar.

B. Research Setting

The reasearch setting refers to the place where the data were collected. This study was conducted in IAIN Batusangkar. The interviewed started from Nov 21st until Dec 29th in IAIN Batusangkar.

C. Research Informants

Research Informant is defined as whom the data got. According to Gay (2000: 139), qualitative research chooses participants whom they judge to be thoughtful and who have information, perspective, and experience related to the topic of the research. Moleong (2006:327) states that research informants must meet several criteria, they are: (1) honest, (2) reliable, (3) speaking, (4) subjects not involved in certain groups, and (5) the subject has certain views about the case.

In order to find the informant, the researcher use a purposive sampling technique, because researcher want to know what is most students based on experience or knowledge. Gay (2000:138) propose that purposive sampling is

judgment sampling when researcher select informants based on the informants' experience or group's knowledge. Students who became informants in this study were student teachers who were included in the criteria of the informants as follows: (1) student teachers in the English Department IAIN Batusangkar in the academic year 2019/2020. (2) Students attend a teaching training program from beginning to end, (3) students make their own midterms test for high school students.

In this research, the student teachers of the English department who follow the teaching practice program at junior high school who had problems in constructing the test will use as informants.

D. Research Instruments

The research instrument be the researcher himself. Sugiyono (2007: 305), states that the main instrument of qualitative research is the researcher himself. In this study, researchers use interviews as a method of collecting data. The researcher conduct interviews with student teachers of IAIN Batusangkar about problems faced in constructing the midterm test for high school students in Tanah Datar to get the information from the informants.

E. The technique of Data Collection

The data collect by doing an interview. The interview be conducted to get the data related to problems faced by student teachers in constructing the test. As stated by Sugiyono (2007: 194) that interviews are used when researchers want to know more about informants. Researchers use unstructured interviews. According to Sugiyono (2007: 197), unstructured interviews are free interviews where researchers do not use interview guidelines that have been arranged systematically and completely for data collection.

The interview guidelines are only used in the form of an outline of the problem to be asked to the informant. Researcher use Indonesian and Minang languages to minimize missing informant information. besides, researcher also use a recorder to record all conversations in the interview.

F. Checking Data Trustworthiness

In analyzing the data, the researcher use triangulation to get good information from the informant. Triangulation is one of the technique used in checking the data trustworthiness in qualitative research. In data collection techniques, triangulation is defined as data collection techniques that are combining various data collection techniques and existing data sources.

In data collection techniques, triangulation is defined as data collection techniques that are combining various data collection techniques and existing data sources. Source triangulation means, to get data from different sources with the same technique. In this research, researcher use time triangulation to make sure that the researcher will not missing the informations The researcher did the interview in two times. For the first interview researcher was ask about their problems in constructing tests and for the second interview is for make sure again the data in the previous interview to check credibility of the data.

G. The technique of Data Analysis

Data analyze by using Miles and Huberman's method. Miles and Huberman (1994: 10) describe three steps for analyzing data, as follows: data reduction, data display, and conclusions and data verification.

1. Simplification of Data

Data reduction is interpreted as a process of choice, focused interest, simplification, abstraction, and transformation of existing data from written records in the field. Data reduction occurred during the interview. In other words, when data is collected, phases continue to occur such as making conclusions, codifying, investigating themes, grouping, and writing memorandums.

In reducing data, researcher conduct several activities. First, he made the interview transcription. Then, provide the code transcription for each informant as follows: I.1, I.2, I.3, I.4, and I.5 and related to R.1 for researcher 1, R.2 for researcher 2 , and so on.

2. Describe Data (Data Display)

Data display is a collection of information compiled and given the possibility to conclude. In this step, the researcher analyze the problems faced by student teachers in constructing the midterms test. The researcher display data in tabular form to find out and show students problems in constructing a midterm test.

3. Data Verification and Conclusion

After reducing and displaying data, the final step is to make conclusions and verify the data. In this activity, researcher make conclusions and verify data by concluding problems faced by student teachers in constructing a midterm test.

CHAPTER IV

RESEARCH FINDING AND DISCUSSION

A. Research Findings

1. Description of the Research

This research discuss about the problems faced by student teachers English Student Department of IAIN Batusangkar in constructing English test for high school in Tanah Datar. The documents of this research is interview recording from English Teaching Department (TBI) students who followed Pre-Service Teaching program. There are eight informants that the researcher choose from TBI A, TBI B, TBI C. In this research, the tests were midterm test and daily test. The researcher did the interview by calling them via WhatsApp voice call an record it.

In doing analysis, first ,the researcher did the interview, second transcript the recording, third, researcher reduce the data, and the researcher coded the data as follow: I.1 the first informant,I.2, the second informant and so forth. Then, the researcher coded the problems as follows: P.1 as the problem 1, P.2 as the Problem 2 and so on. The researcher used unstructured interview and did interview one on one interview and time triangulation to get the data, it useful to check validity of the data.

The findings of the problems are faced by student teachers English Student Department of IAIN Batusangkar in constructing English test for high school in Tanah Datar, based on the data, the researcher found several problems are:

Table 1.1
Problems Faced by Student Teachers

Student Teachers problem	Informants
Getting Difficulty Construct Valid Test	I.1,I.2,I.4,I.5,I.6,I.7,I.8
Getting Difficulty to construt various and appropriade level of thinking in test	I.2,I.4,I.5
Not sure about the usability of the test	I.1,I.3,I.5,I.7,I.8
Lack of Preparation	I.1,I.2,I.3,I.6,I.8
Getting difficulty to find suitable texts for testing reading	I.2,I.3,I.5,I.7
Grammatical Problem	I.1,I.6
Getting Difficulty to develop the test from table of spesification	I.6

2. Data Analysis

In this study, the researcher analyze eight recordings of the interview. For more specific information about the problems are faced by student teachers of IAIN Batusangkar in constructing English test for high school in Tanah Datar, from the quotation is :

a. Getting difficulty construct valid test

The student teachers getting difficulty in construct a valid test. As a next teacher, student teacher should give them a test to measure their ability and skill. So, they had to prepare a valid test .

Excerpts from interviews wit infromants show:

I.1 : *”a jadii a kan mawar kan ngajarnya di MAN,MAN 1 Tanah Datar nah untuk menentukan soal ini udah, udah pas atau belum.”* (So, I taught in MAN, MAN 1 Tanah Datar. The test that I choose already fit with the students or not.)

I.7: *“kalau dalam persiapan yang jadi kendala saya oo memperkirakan soal yang tepat untuk siswanya soalnya saya tidak*

kadang agak ragu kan mengenai kemampuan siswa soalnya kemampuan siswa itu berbeda-beda kadang kadang soalnya kadang saya sulit menemukan soal yang tepat yang bisa untuk semua kalangan begitu". (My problem in preparing test was about students' various ability. It doubted me. Sometimes, it hard for me to found the best test for all students.)

Based on the data above, the student teachers getting difficulty to construct valid test, it can be summarize that the student teachers were confuse in preparing test.

- b. Getting difficulty to construct various and appropriate level of thinking in test

A test had different level of difficulties for each level. So, student teacher should think and manage the test carefully. Then students would be easy to answer the question. But the student teachers getting difficulty base on the quotation is :

I.4: *"ho oh,terus aa yang kedua mungkin waktu ok selanjutnya yang kedua mungkin aa permasalahan yang ditemui dalam membuat soal uh juga ok jadi yang kedua itu kesulitannya lagi adalah dalam membuat soal objektif yang mana nantik sebenarnya kita ketahui bahwasanya itu ada yang namanya oo mudah atau dia optinnya A,B,C,D nya itu ada mungkin beberapa mendekati, yang tidak mendekati dan sebagainya dan itu juga merupakan salah satu masalah yang mungkin saya hadapi karena kenapa karena untuk ini juga merupakan pengalaman pertama untuk membuat soal juga apalagi o kita itu ngajar kelas 7 yang mana mereka belum menyentuh o bahasa inggris sama sekali gitu." (Second, the problem in constructing test was the multiple choice items. Like A, B, C, D. It might be had the closest answer. Another items which were incorrect also the problem. It was my first time to construct the test, especially for grade seven which never learn English.)*

I.5: *"oo jadi kan pasti dapat procedure teks kemaren kayak gitu bg, dibikin tu a ada yang A B C nya tu hampir-hampir sama tiap jadi padahal jawabannya itu A gitu bg mereka tu memang semua menjawabnya C gitu bg ada beberapa siswa 3 kalau ngga salah dari 36 siswa ituyang 3 yang betul seperti itu"*

(Well, they got procedure text. Their choice item almost same each A B C. the correct answer was C. Only three of 36 students were correct.)

Based on the interview, it can be concluded that the student teachers had few information about the level of difficulties of the test. Also their students learned English for the first time. They had to know students' thinking level. They also had to prepare the best test that suit to their students. Their students had various thinking level. Then, the items should content relevant material, which can decrease the reliability and the validity of the test score. Also, the question should be a question or a partial sentence. Because, it allows students to focus on answering the question rather than holding the partial sentence.

c. Not sure about the usability of the test

Student teachers need to ensure that the assessment requirements are appropriate to the intended learning outcomes of learning process and that in their operation they do not distort the learning process and, that they do not make unreasonable demands on the test and resources available to learner or teacher. According to Brown (2010) Practicality refers to the logistical, down-to-earth, administrative issue involved in marking, giving and scoring an assessment. For example:

I.1: *“kalau untuk pemikiran waktu soal itu juga kemaren agak agak ada kesulitan sih jadi aa ya untuk menentukan masing-masing apa namanya,kira-kira anak ini bisa ngga sih untuk aa 5 soal ini waktu nya cukup apa ngga.”* **(It was difficult to think about the time management. Can my students done these five items in a time?)**

I.3: *“ada karena oo kalau memberi soal jarang saya yang memberi durasi per soalnya adanya itu untuk satu soal durasinya berapa eh untuk semua soal durasinya berapa”* **(I rarely give the time of the test for each item than for one test.)**

Based on the data above, it can be concluded that the student teachers did not what the test for. Another part in practicality is the test should have specific time. It means that the time test is efficient or not.

d. Lack of Preparation

Student teachers had lack of sources. They choose internet to be alternative source. Internet can provide anything, it seems like it was very helpful for them. Let's see the result:

I.1: *“terbayang untuk membuat soal kan, jadi ada kendalanya gitu jadi kadang aa lihat di internet gitu tapi kata-katanya diganti sendiri a jadi bahasa sendiri gitu atau sama kayak ada story mungkin disitu nanti ada yang diganti gitu”.* **(I think in constructing a test, it has the obstacles. So, sometimes, I look at the internet. But, I change the kind of language by myself or same as the story of the test.)**

I.3: *“bisa jadi kurangnya kesiapan misalnya karena baru awal mengikuti pl jadi masalah yang saya temukan itu seperti kurangnya sumber untuk membuat soal jadi yang saya lakukan yaitu misalnya mencari soal di internet, di copy setelah itu saya edit atau langsung saya ambil yang ada jawabannya sekalian atau bisa juga mencari soal di internet nah nanti saya cari sendiri jawabannya”* **(It might because the lack of preparation in following pre service teaching. So, the problem was lack of sources to construct the test. I would find the test in the internet. After I copied the test, I edited or I also took the answer too. Also, I looked up the question, and I answered it by myself.)**

Based on the data above, it can be concluded that the teacher students did not have any sources. They construct their test from the internet, which it was not 100% correct. Also, some of them did not edit the question. They just copied and pasted it. But, some of them edited the question and make their key answer.

e. Get difficulty to find suitable texts for testing reading

Student teachers should find a text for reading tests can be understood by students. Fit it with the level of grade. If the students are in junior high school, the teacher students should give the familiar

vocabulary of the text. If the level is senior high school, teacher students should give them medium vocabulary variation of the text.

I.2 : *“ya vocabnya agar siswa tu bisa memahami dengan mudah ketika menjawab soal... ya yang tepat ya yang tepat untuk siswa supaya siswa tu mengerti apa yang diminta oleh soal dan ditanya oleh soal”* **(The vocabulary, so students would be easy to answer the question.)**

I.3 : *“ada yaitu seperti teks-teks yang setiap soal kan tentu memerlukan teksnya ya kan, nah dari berbagai macam teks itu ada yang teks yang bahasanya itu terlalu ilmiah atau mungkin terlalu tinggi untk dipahami oleh siswa”* **(The texts, sometimes, used scientific vocabulary. The vocabularies are hard to understand by the students.)**

f. Grammatical problems

Student teacher should master in grammar. It is a requirement to be an English teacher. Also, since they learn about grammar, the test also should consist of grammar. Student teacher should know which test that suit to measure students' ability in mastering grammar.

I.1 : *“o itu tadi kalau misalkan ee apa namanya oooo karna pertama kali tadi tu ya jadi kurang terlalu bisa gitu jadi liat di internet terus juga paling kalau misalkan tanpa melihat internet paling di gramar trus juga”* **(Because it was the first time, so I couldn't do it yet. I looked up to the internet. Like, without looked up to the internet, grammar.)**

I.6 : *“oo yang grammar tadi tu bg oo dalam pembuatan soalnya oo apalagi untuk pembuatan soalnya oo apalagi untuk pertanyaan pasti lebih harus sesuai oo jadi adek tu takut kalau grammar atau tata bahasanya salah saat oo dipertanyakan ujian itu bg atau latihan itu o iya jadi kalau grammar salah”* **(I am afraid if my grammar is error when I construct the test. The test should have correct grammar.)**

Based on the the data above, it can be concluded that the student teachers have problem in grammar. They was afraid if their grammar is incorrect.the occurence of an error in the grammer of course will cause the question instructions to be unclear,and can change the meaning.

g. Getting difficulty to develop the test from table of spesification

Before construct the test, student teachers should prepare the grid. Grid also can be called as table of spesification. It means that there are several spesification of the test that will be constructed by the teacher or student teachers. Grid is helpful because it is a guideline for student teachers to construct the test that fit with the goal.

I.6 : *“iya bg ada bg mungkin karena baru pertama kali membuat dan merancang soal apalagi soalnya harus sesuai dengan kisi-kisi yang di indikator Kd nya harus sesuai jadi dalam pembuatan soalnya harus teliti dan harus sesuai dengan kisi-kisinya disitunya agak terkendala bg. iya bg saat menyesuaikan soal kita dengan kisi-kisinya agak sulit bg”*
(Yes. Maybe it is the first time to construct the test. Especially, it is should fit with the grid in the indicator, the competency. So it will be easy to construct the test. But, the problem is in adjusting the test with the grid.)

Based on the data above, it can be concluded that teacher students were getting difficulty to develop the test from table of spesification. It is happen because they don't look up to the grid before construct the test. Table of spesification is helpful in constructing test.

B. Discussion

This research finding showed that there are several problems that found related to constructing test. There are seven problems that found by the researcher related to constructing the test. Accoding Lin and Gronlund (1995:47) state that a good test should be fulfills three characteristics, namely: Validity, reliability, and reusability. First is difficulty to construct valid test. Student teachers were confused in construct test. They did not know how to construct a valid test, while they had to know how to construct a valid tests. They also had to prepare the valid test that suit to their students.

Second is Getting Difficulty to construt various and appropriade level of thinking in test. As a test maker, teacher must understand and pay attention to the items and the test takers. (Rasyid.Mansur in Ahmad, 2008:239) It means that the level of difficulties in items of the test is suit with the students. Since

it was their first time in constructing test, they did not know how to construct the test that appropriate with the students level of thinking in test. Also, their students could not answer the multiple choice question correctly. According to Brown (2010) Practicality refers to the logistical, down-to-earth, administrative issue involved in marking, giving and scoring an assessment. Third is Not sure about the usability of the test. In this case, the teacher students did not pay their attention to time management. Time management is needed when they construct the test. The test will be practically if they can give the test with appropriate time.

Forth is lack of preparation. For lack of preparation they searched the example of the test from internet. Not all of the teacher students who edited the question. Some of them copied it and gave the question to the students without look up the right answer key. Fifth is get difficulty to finds suitable texts for students for reading tests. According to Nurgiantoro in Shofiah (2013:356) state that what must be considered in choosing a reading text is not only the level of proficiency of students in second language but also the difficulty level of reading. Student teachers did not know what the suitable text for student level. So that the student can understand the text in reading tes.

Sixth is grammatical error. As Palmer In Banjar, Siti A, Muhimmatul I, (1994) state that writting is difficult to learn because the writer should involve a procces that includes planning, Organizing, and revising to present meaning in word or sentence. Student teachers was afraid of they have grammatical error. They had construct test correctly. If the grammar was incorrect, it would be hard for students to answer the question. Getting Difficulty to develop the test from table of spesification. Before construct the test, student teachers should construct the grid. The grid is helpful in constructing the test. But, teacher students were constrained in matching the grid with the test. It is happen because they don't look up to the grid before construct the test.

CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

This chapter concludes the findings of the research problems faced by student teachers in constructing test. Based on the research findings, the researcher found seven problems that faced by student teachers of IAIN Batusangkar in constructing English test for high schoolin Tanah Datar. Those difficulties are: 1) Getting difficulty construct valid test , 2) Getting difficulty to construct various and appropriade level of thinking in test, 3) Not sure about the usability of the test, 4) Lack of Preparation, 5) Get difficulty to find suitable texts for testing reading, 6) Grammatical problems 7) Getting difficulty to develop the test from table of spesification.

B. Suggestion

Based on the results of this research, the researcher would like to give some suggestions to:

1. Lecturer

Based on the result of this research, hopefully give information to the English lecturers about research problems faced by student teachers in constructing test. Most of students' problems are getting difficulty construct valid test, get difficulty to construct various and appropriade level of thinking in test, not sure about the usability of the test, lack of preparation, getting difficulty to find suitable texts for testing reading, grammatical problems and getting difficulty to develop the test from table of spesification. Hopefully for the lecturer who teach language assessment can find the strategies how the student more comprehend in constructing test. Therefore they will not get problems in constructing the test.

2. Student Teachers

From the result of this research, the researcher gives some suggestion the students who will follow pre-service teaching. It is necessary for the students to consider facing problems in constructing test such as getting difficulty construct valid test, get difficulty to construct various and appropriate level of thinking in test, not sure about the usability of the test, lack of preparation, getting difficulty to find suitable texts for testing reading, grammatical problems and getting difficulty to develop the test from table of specification. Because of that the students must prefer themselves as well, having many vocabularies, understand about grammar and confidence with themselves, it will make them easy to construct the test.

3. Further researcher

Based on the research findings, the researcher found seven problems that faced by student teachers of IAIN Batusangkar in constructing English test for high school in Tanah Datar. Those difficulties are: 1) Getting difficulty construct valid test, 2) Getting difficulty to construct various and appropriate level of thinking, 3) Not sure about the usability of the test, 4) Lack of Preparation, 5) Getting difficulty to find suitable texts for testing reading, 6) Grammatical problems and 7) Getting difficulty to develop the test from table of specification. The researcher would like to give a suggestion to the further researchers to do a research more deeply about the coping strategies of the problem faced by student teachers in constructing test.

BIBLIOGRAPHY

- Bachman, Lyle. F. (1990). *Fundamental Consideration in Language Testing*. Oxford University Press.
- Brown, Douglas. (2004). *Language Assessment Principle and Classroom Practice*. San Fransisco : Longman.com
- Brown, Douglas. (2010). *Language Assessment Principle and Classroom Practice Second Edition*. San Fransisco : Longman.com
- Gay, L.R. (2020). *Educational Research Tenth Edition*. San Fransisco : Longman.com
- Gay, L.R. and Peter Airaisan. (2000). *Educational Research: Competencies and for Analysis and Application*. New Jersey. Practice Hall, Inc
- Jabarifar, Taghi. (2009). *The 2nd International Conference of Teaching and Learning (ICTL 2009.)* Malaysia: INTI University College
- Jhonson in Sanjaya, Wina. (2008). *Kurikulum Dan Pembelajaran Teori Dan Praktik Pengembangan KTSP*. Jakarta : Kencana Prenada Media Grup
- Linn, Robert L. And Gronlund, Norman E (1995). *Meassurement and Assesment in Teaching*. (Seventh Edition).Ohio:Prentice-Hall, Inc.
- MacMillan, J.H. and Schumacher, S. (2001). *Research Education. A Conceptual Introduction, 5th Edition*. Boston: Longman.
- McNamara, T. F. (1996). *Measuring Second Language Peformance*. Malaysia: Longman
- Milles, Matthew and Michael Hubberman. (1994). *Qualitative Data Analysis: An Expended Sourcebook*. Sage
- Moleong, Lexy. J. (2010). *Metodologi Penelitian Kualitatif*. Bandung: Rosdakarya.
- Nurgiantoro In Sofiah Nurul (2017). *Pertimbangan Pemilihan Teks Bacaan Dalam Pengajaran Dan Pembelajaran Membaca*. Malang.
- Osadebe, Patrick Uzo. (2015). *International Journal of Evaluation and Research in Education (IJERE)*.

- Palmer In Kumala P.B. Siti, A. Muhimatul, I. (2018). *An Analysis of Grammatical Errors on Student` Writing*.
- Palmer, S. Adrian and Lyle F. Bachman. (1996). *Language Testing in Practise: Designing and Developing Useful Language Tests*. Oxford University Press
- Rasyid, Harun & Mansur In R Ahmad N,K.(2013). *Analisis Tingkat Kesulitan (Difficulty Level) Soal Pada Buku Sejarah Kebudayaan Islam Kurikulum 2013*. Bumiaji Batu.
- Rivai, E., Ridwan, A., Supriyati, Y., & Rahmawati, Y. (2019). *Influence of Test Construction Knowledge, Teaching Material and Attitude on Sociological Subject to Quality of Objective Test in Public and Private Vocational Schools*. *International Journal of Instruction*.
- Sugiyono. (2007). *Metode Penelitian Kualitatif Kuantitatif dan R&D*. Bandung: Alfabeta
- Suharsimi Arkunto in Irhamsyah, Hakim Lukmanul. (2005). *The Analysis of The Teacher-Made Test For Senior High School at State Senior Hgh School 1 Kutacane. Aceh*
- Tanner, Howard. (1997). *Assessment: A Practical Guide for Secondary Teachers*. London: Continuum.