

**EXPLORING STUDENTS' ABILITY IN CONSTRUCTING
ENGLISH TEST THROUGH PEERWISE**

THESIS

Submitted in partial fulfillment of the requirement for the degree of
Sarjana Pendidikan (S.Pd) in Teaching English



**UIN SUNAN AMPEL
S U R A B A Y A**

By:

Mahmudah Asnaniyah
NIM D05215015

**ENGLISH TEACHER EDUCATION DEPARTMENT
FACULTY OF TEACHING AND TEACHER TRAINING
ISLAMIC STATE UNIVERSITY OF SUNAN AMPEL
SURABAYA**

2019

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : MAHMUDAH ASNANIYAH
NIM : D05215015
Jurusan/Program Studi : PENDIDIKAN BAHASA INGGRIS
Fakultas : TARBIYAH DAN KEGURUAN

Menyatakan dengan sebenarnya bahwa skripsi yang saya buat ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan saya sendiri. Segala materi yang diambil dari karya orang lain hanya digunakan sebagai acuan dengan mengikuti tata cara dan etika penulisan karya ilmiah yang ditetapkan oleh jurusan

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Apabila pernyataan tidak sesuai dengan fakta yang ada, maka saya selaku penulis bersedia dimintai pertanggungjawaban sesuai ketentuan peraturan perundang-undangan yang berlaku.

Surabaya, 8 Juli 2019

Pembuat Pernyataan,



Mahmudah Asnanayah
NIM.D05215015

ADVISOR APPROVAL SHEET

This thesis by Mahmudah Asnaniyah entitle "*Exploring Students' Ability in Constructing English Test Through PeerWise*" has been approved by the thesis advisors for further approval by the board examiners.

Surabaya, July 8th, 2019

Advisor I,



Fitriah, Ph.D

NIP. 197610042009122001

Advisor II,



Hilda Izzati Madiid, M.A

NIP. 198602102011012012

EXAMINER APPROVAL SHEET

This thesis by Mahmudah Asnaniyah entitled "*Exploring Students' Ability in Communicative English Test Through PeerWise*" has been examined on July, 18th 2019 and approved by the board of examiners.



Dean,

Ali Mas'ud, M.Ag, M.Pd.I

NIP. 196301231993031002

Examiner I

Rizka Safriyani, M. Pd

NIP. 198409142009122005

Examiner II

Drs. Muhtarom, Med.Grad Dip TESOL

NIP. 196512201992031005

Examiner III

Fitriah, Ph.D

NIP. 197610042009122001

Examiner IV

Rakhmawati, M. Pd

NIP. 197803172009122002



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Mahmudah Asnawiyah
NIM : 005215015
Fakultas/Jurusan : Tarbiyah dan Keguruan / Pendidikan Bahasa Inggris
E-mail address : Asnawiyah95@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

EXPLORING STUDENTS' ABILITY IN CONSTRUCTING
ENGLISH TEST THROUGH PEERWISE

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 05 Agustus 2019

Penulis

(Mahmudah Asnawiyah)
nama terang dan tanda tangan

ABSTRACT

Asnaniyah, Mahmudah. (2019). *Exploring Students' Ability In Constructing English Test Through Peerwise*. An Undergraduate Thesis, English Language Education Department, Faculty of Education and Teacher Training, UIN Sunan Ampel Surabaya. Advisors: Fitriah Ph.D and Hilda Izzati Madjid M.A.

Key Words: Students' Ability, English Test Construction, PeerWise.

Test becomes the main source in measuring and determining the assessment of students' ability related to student competence. However, to do the test is also not easy because it must make a test item as an instrument to measure the students' ability particularly in the preparation of the English test which requires several criteria. This needs to be observed that focusing on English test construction particularly in the form of multiple-choice test. This matter is supported by the peerwise application that is used by students in Language Testing and Assessment class as the same time to be the research subject. In addition, the students are from English department that also as the students teacher in the future. So, this research was conducted to explore the students' ability in constructing English test particularly in the form of multiple choice test through peerwise. This research purposed to know how further the students construct English test through peerwise. The data is document of multiple-choice made test by students in peerwise from 3 topics with total of 50 questions that was selected based on the students who got high rating. The researcher analyzed the data by documentation guideline using descriptive qualitative approach. The result of this study indicated that all of questions made by students met 2 aspect of standard criteria for constructing the test item correctly. Those are in terms of stem and alternatives. Although, one of criteria are not listed in each items that related to two aspects above.

ABSTRACT

Asnaniyah, Mahmudah. (2019). *Exploring Students' Ability in Constructing English Test Through PeerWise*. Skripsi. Pendidikan Bahasa Inggris, Fakultas Tarbiyah dan Keguruan, UIN Sunan Ampel Surabaya. Pembimbing: Fitriah Ph.D dan Hilda Izzati Madjid, M.A.

Kata Kunci: Kemampuan Siswa, Pembuatan Tes Bahasa Inggris, Peerwise

Test menjadi sumber utama dalam mengukur dan menentukan penilaian terhadap kemampuan siswa yang berhubungan dengan kompetensi siswa. Namun, melakukan test juga tidak mudah karena harus membuat test item sebagai alat untuk mengukur kemampuan siswa terutama dalam penyusunan tes Bahasa Inggris yang mana memerlukan beberapa kriteria. Sehingga hal ini butuh untuk diteliti yang focus pada pembuatan tes Bahasa Inggris khususnya dalam bentuk tes pilihan ganda. Hal ini didukung oleh aplikasi peerwise yang digunakan oleh siswa di kelas Language Testing and Assessment sekaligus menjadi subjek penelitian. Selain itu, siswa berasal dari jurusan Bahasa Inggris yang juga sebagai calon guru di masa depan. Jadi, penelitian ini dilakukan untuk mengeksplorasi kemampuan siswa dalam membangun tes Bahasa Inggris khususnya dalam bentuk pilihan ganda melalui peerwise. Data penelitian adalah dokumen tes pilihan ganda yang dibuat oleh siswa dalam peerwise dari 3 topik dengan total 50 pertanyaan yang dipilih berdasarkan siswa yang mendapat peringkat tinggi. Penelitian ini menganalisis data dengan pedoman dokumentasi yang menggunakan pendekatan kualitatif deskriptif. Hasil penelitian ini menunjukkan bahwa semua pertanyaan yang dibuat oleh siswa memenuhi 2 aspek standar kriteria untuk membuat item tes dengan benar. Aspek-aspek itu adalah dalam hal stem dan alternatives. Meskipun, salah satu kriteria tidak tercantum di setiap item yang terkait dengan 2 aspek diatas.

TABLE OF CONTENTS

COVER	
ADVISOR APPROVAL SHEET	ii
EXAMINER APPROVAL SHEET	iii
MOTTO	iv
DEDICATION	v
ABSTRACT	vi
ACKNOWLEDGEMENT	viii
PERNYATAAN KEASLIAN TULISAN	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xii
LIST OF FIGURES	xv
LIST OF ABBREVIATION	xvi
LIST OF APPENDICES	xvii
CHAPTER I: INTRODUCTION	1
A. Background of Study	1
B. Research Question	4
C. Objective of Study	4
D. Significance of the Study	5
E. Scope and Limitation	6
F. Definition of Key Term	6
1. Students' Ability	6
2. English Test	6
3. Constructing English Test	7
4. PeerWise	7
CHAPTER II: REVIEW OF RELATED LITERATURE	8
A. Theoretical Framework	8
1. Test	8
2. Multiple-Choice	9
3. PeerWise	12
B. Previous Studies	15
CHAPTER III: RESEARCH METHOD	19
A. Research Design	19
B. Research Setting	19
C. Data and Source of Data	20

D. Research Instruments	20
E. Data Collection Technique	21
F. Data Analysis Technique	21
G. Research Stages	23
CHAPTER IV: RESEARCH FINDING AND DISCUSSION	24
A. Research Findings	24
B. Research Discussion.....	57
CHAPTER V: CONCLUSION AND SUGGESTION	62
A. Conclusion	62
B. Suggestion.....	63
REFERENCES	64
APPENDICES	66

LIST OF TABLES

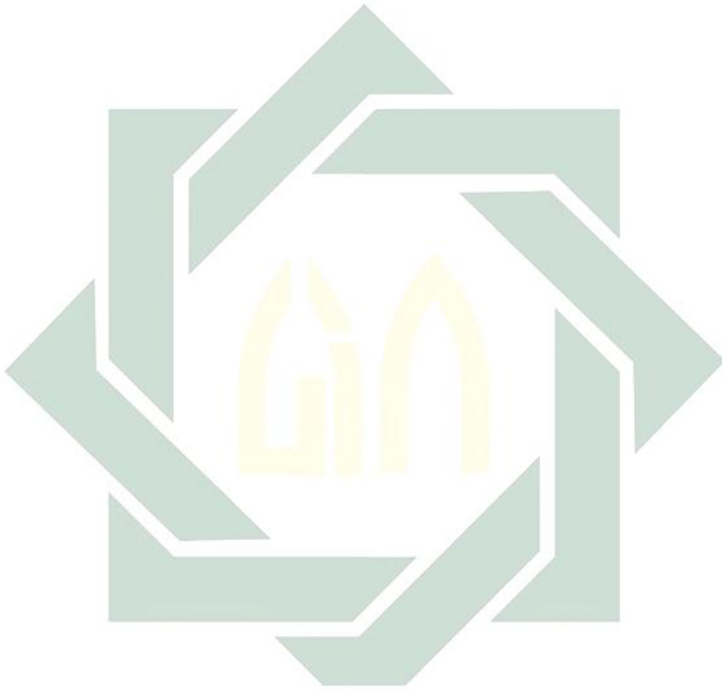
Table	Page
4.1 Example of Multiple Choice which Has Not Instruction (Topic 1)	25
4.2 Example of Multiple Choice which Has Not Instruction (Topic 2)	25
4.3 Example of Multiple Choice which Has Not Instruction (Topic 3)	26
4.4 Parts of Stem in Multiple Choice Construction	26
4.5 Example of Multiple Choice Stem which Presents A Single Problem (Topic 1)	27
4.6 Example of Multiple Choice Stem which Presents A Single Problem (Topic 2)	28
4.7 Example of Multiple Choice Stem which Presents A Single Problem (Topic 1)	29
4.8 Category of Positive Stem in Multiple Choice Construction	30
4.9 Example of Multiple Choice which Presents Positive Stem in Form of W-H Question (Topic 1)	30
4.10 Example of Multiple Choice which Presents Positive Stem in Form of W-H Question (Topic 2)	31
4.11 Example of Multiple Choice which Presents Positive Stem in Form of W-H Question (Topic 3)	32
4.12 Example of Multiple Choice which Presents Positive Stem in Completion Form (Topic 1)	33

4.13 Example of Multiple Choice which Presents Positive Stem in Completion Form (Topic 2)	34
4.14 Example of Multiple Choice which Presents Positive Stem in Completion Form (Topic 3)	35
4.15 Example of Multiple Choice which Presents Positive Stem of Improper Formatting	36
4.16 Example of Multiple Choice which Presents Positive Stem in Form of Instruction	37
4.17 Example of Multiple Choice which Presents the Stem in Negative Form (Topic 1)	38
4.18 Example of Multiple Choice which Presents the Stem in Negative Form (Topic 2)	39
4.19 Example of Multiple Choice which Presents the Stem in Negative Form (Topic 3)	40
4.20 Grammar Error Ctaegory in Constructing Stem	41
4.21 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Ctaegory (Topic 1)	42
4.22 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Ctaegory (Topic 2)	42
4.23 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Ctaegory (Topic 3)	43
4.24 Example of Multiple Choice Stem in Term of Article of Grammatical Error Ctaegory (Topic 1)	44
4.25 Example of Multiple Choice Stem in Term of Article of Grammatical Error Ctaegory (Topic 2)	45

4.26 Example of Multiple Choice Stem in Term of Article of Grammatical Error Ctaegory (Topic 3)	45
4.27 Example of Multiple Choice Stem in Term of Singular/Plural Noun of Grammatical Error Ctaegory (Topic 1)	46
4.28 Example of Multiple Choice Stem in Term of Singular/Plural Noun of Grammatical Error Ctaegory (Topic 2)	47
4.29 Example of Multiple Choice Stem in Term of Preposition of Grammatical Error Ctaegory (Topic 3)	48
4.30 The Aspects of Alternatives in Multiple Choice Construction	49
4.31 Example of Multiple Choice Alternatives in Terms of Key Answer and Distractors (Topic 1)	50
4.32 Example of Multiple Choice Alternatives in Terms of Key Answer and Distractors (Topic 2)	51
4.33 Example of Multiple Choice Alternatives in Terms of Key Answer and Distractors (Topic 3)	52
4.34 Example of Multiple Choice Alternatives in Terms of Verb Tense of Grammatical Error (Topic 1)	54
4.35 Example of Multiple Choice Alternatives in Terms of Verb Tense of Grammatical Error (Topic 2)	55
4.36 Example of Multiple Choice Alternatives in Terms of Verb Tense of Grammatical Error (Topic 3)	56

LIST OF FIGURE

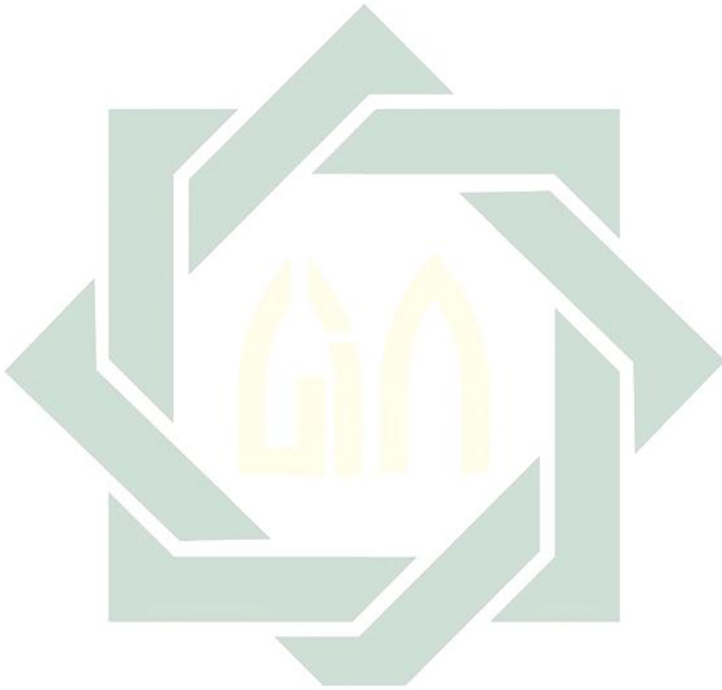
Figure		Page
2.1	Front Layout of PeerWise Application	13



LIST OF ABBREVIATION

MCQs

Multiple Choice Questions



LIST OF APPENDICES

- APPENDIX 1: Document of Students' Multiple Choice Test
- a. Topic 1 (Principle of Language Assessment)
 - b. Topic 2 (Standardized Test)
 - c. Topic 3 (Designing Classroom Language Test)
- APPENDIX 2: Documentation of Analysis Guideline in Constructing Good Multiple Choice Test
- APPENDIX 3: Result Analysis of Students' Multiple Choice Test
- a. Topic 1 (Principle of Language Assessment)
 - b. Topic 2 (Standardized Test)
 - c. Topic 3 (Designing Classroom Language Test)
- APPENDIX 4: Instrument Validation Letter
- APPENDIX 5: Advisors Approval Sheet
- APPENDIX 6: Task Letter of Thesis Guidance
- APPENDIX 7: Thesis Guidance Card

CHAPTER I INTRODUCTION

This chapter mentions background of study that describes important point of English test construction particularly in form of multiple-choice test so it should be observed and analyzed by the researcher. Then, the research question is formulated to the problem that along with the objective of the study. The benefit of this study relates to the importance of study to know the purpose of this study. This study provides the limitation and scope to prevent the expansion of the discussion. It is accompanied by definition of key terms to better understand of the research.

A. Background of Study

Testing has been used widely until today because it becomes the main source for the decision of restructuring of schools that related to students score. Standardized of testing claimed that testing is a valid and reliable indicator of students learning and also an effective system in ensuring the academic competencies are attained by all students¹. As described, test have a role to enhance students' emphasis in mastering the material of the course as one of the benefit for students². While, for the teachers can measure the students' ability in what area of their achievement related to the material through test. In contrast, it is also difficult for teacher to construct the test item.

Constructing test is one of the most challenging tasks especially for English test because it needs the elements to cover it. One of the important element is designing test specifications³. It should be consisting of some criteria. Those are reliability and validity. Reliability measures the consistency of the test while validity examines to what extent the test match with the set of

¹ Mehmet Dulger, et.al. "Assessing the Validity of Multiple-Choice Questions in Measuring Fourth Graders' Ability to Interpret Graphs About Motion and Temperature". *Nevada*, 2017

² Harold S. Madsen, *Teaching In Testing* (New York: Oxford University Press, 1983), 4.

³ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 71.

course goal⁴. All of them must be fulfilled when constructing English test. Therefore, every teacher should know and understand what criteria to construct English test in good way particularly in constructing multiple-choice item as one of form testing which usually used to measure the students' ability.

Multiple-choice question is the most popular test format that prevalently used across the Asian nations or regions⁵ for example National Examination, TOEFL and IELTS. On the other hands, Brown also argued that multiple choice which many appear to be the simplest kind of item to construct are extremely difficult to design correctly⁶. Teacher might spend even more time designing such items in multiple-choice than grading it.

Therefore, it needs to be observed that focusing on constructing English test especially in the field of multiple-choice test. Importantly, it is for students' teacher in English department to add their knowledge about constructing english test that could be benefit for the future. Fortunately, one of the course of English teacher education department at UIN Sunan Ampel Surabaya as the candidate of teacher implemented to construct English test through making multiple-choice test. It is for creating the good quality in teacher self. Teachers can make good qualified assessment in multiple-choice test. One of the way is through web-based application that is delivered as the independent study such as PeerWise. It provides collaborative learning for students in constructing multiple-choice questions that increase productivity of students during the learning process.⁷

⁴ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 20

⁵ Yeon Hee Choi, et.all. "Current Trends and Issues in English Language Education in Asia. The Journal of Asia TEFL". 2008.

⁶ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 55

⁷ Paul, Simon, and Melissa. "Evaluation of PeerWise as an Educational Tool for Bioengineers. American Society for Engineering Education". 2010.

Peerwise becomes the media in constructing the test because it provides different way between the traditional multiple-choice construction based paper with multiple choice construction in peerwise. In this case, the test construction stated explanation as the additional information of the options. On the other hands, it arranges the students assessment of their classmate test construction in Peerwise. So, it is a good chance to learn especially for teacher in qualified assessment.

In this case, there are some studies that discuss about PeerWise. For the first study by Alison Elizabeth Kay focused on the correlation between engaging with PeerWise and end of course exam score which the data analyzed by qualitative design to evaluate the correlation both of them. The result of this study is students who engage to a higher level with PeerWise achieve a higher exam score than their lower engaging peers⁸. The second study discussed about the correlation between students' participation and their examination score in PeerWise online learning by Fadhila Alfany. The data of this study analyzed by the quantitative correlational design which is taken from 105 students. The result of this study is the correlational between students participation and their achievement score as pearson correlation had positive correlation⁹. The third study examined that PeerWise provided additional support for students and encouraged active learning by Stephen McClean. The result of this study is the use of PeerWise is successful in improving examination performance¹⁰. The fourth of this research focused on measuring the validity of multiple-choice questions for testing the students' ability by Mehmet Dulger and Hasan Deniz. The finding of this study is students answer in multiple-choice test do not fully reflect

⁸ Alison Elizabeth Kay. "Student-generated content: Investigating Student Use of PeerWise". *The University of Edinburgh*, 2016.

⁹ Fadhila Alfany. "*The Correlation Between Students' Participation and Their Examination Score in PeerWise Online Learning at English Education Department UIN Sunan Ampel Surabaya*". (Surabaya: UINSA, 2016).

¹⁰ Stephen McClean. "Implementing PeerWise to Engage Students in Collaborative Learning". 2015.

students' true competency¹¹. The fifth study by Elyza Martiarini who identified the different result of students' reading comprehension using multiple-choice and cloze task. The result of this study is between multiple-choice and cloze test have significant differences which is the students' reading comprehension used multiple-choice was higher than cloze task¹².

From those of studies it can be concluded that PeerWise is only limited on students' examination score. However, PeerWise application can also influence students' ability in making multiple-choice questions. This matter related to the other study that discuss how the influence of multiple-choice test to assess the students' ability, whereas it is difficult to construct. Those studies can be combined and analyzed from students' multiple-choice made test ability in PeerWise. It can be used Nurman E Gronlund's theories in assessing of multiple-choice test to measure the students' ability. Therefore, this study is intended to analyze how the students' ability in constructing English test especially for multiple-choice test through PeerWise.

B. Research Questions

Based on the rationale stated in the background of the study above, the research question is formulated as follows: "How do the students construct English test through PeerWise?". In this matter, exploring the students construction of multiple choice test in Peerwise that related to the material that have been learned in the class before.

C. Objectives of Study

Based on the research questions above, the objectives of this study are follows:

¹¹ Mehmet Dulger-Hasan Deniz, "Assessing The Validity of Multiple-Choice Questions in Measuring Fourth Graders' Ability to Interpret Graphs about Motion and Temperature". *Academic Publisher*, 2017

¹² Elyza Martiarini, "Constructing Multiple-Choice and Cloze Task On Student's Reading Comprehension". *University of Indraprasta PGRI*, 2017

To know how further the students construct English test through PeerWise especially in form of multiple-choice test. This include identifying multiple-choice construction that focus on the instruction, stem, and alternatives based on the validated test construction particularly in the form of multiple-choice.

D. Significance of The Study

The findings of the study are expected to contribute both theoretically and practically which every reseach was done may have good information and result in for many parts:

1. Theoretically: this study about students' ability in constructing English test especially multiple-choice made test through PeerWise which is offer to the theory of constructing test by using online learning. The result of this study give information as the reference for who wants to establish a research on the same strategy as this research.
2. Practically
 - a. For teachers: they know how students' learning process in using PeerWise and also their ability in constructing English test especially multiple-choice questions. Besides that, to know how the measurement of students' multiple-choice made test.
 - b. For students: they know how to construct good multiple-choice questions that appropriate with expert's criteria in making multiple-choice test and also how the use of PeerWise as the online learning tool.
 - c. For further researcher: this study expects to give information for other researcher in constructing English test particularly good multiple-choice question through PeerWise.

E. Scope and Limitation

The scope of this study focuses on multiple-choice made test. This include identifying multiple-choice construction particularly in instruction, stem, and alternatives based on the validated test construction particularly in the form of multiple-choice. The form of test is writing assignment by all of students through constructing multiple-choice test in PeerWise as the media of online learning. This study analyzed the appropriateness of students' multiple-choice made test with the criteria of good multiple-choice based on Nurman E Gronlund theory.

It limits to the students of English Teacher Education Department as the student teacher in State Islamic University of Sunan Ampel Surabaya. In addition, this research conducted at language testing and assessment class. Equally, the research subject is all students of A class that limits on specific material consist of three topics included Principle Language Assessment, Standardized Test, and Designing Classroom Language Test.

F. Definition of Key Terms

To have some interpretation of these key terms, the researcher tries to list important terms in this study that to make the detail information about the keys which are related to the topic in this research. So, the information is explained below:

1. Students' ability: performance of students that involved to the score¹³. This study defined students' knowledge and skill in all of areas particularly in constructing English test so the assignment that is gotten can be reached.
2. English test: test is a set of techniques or items that requires performance on the part of test-taker to qualify general

¹³ Thomas M. Haladyna, *Developing and Validating Multiple-Choice Tests Item* (London: Lawrence Erlbaum Associates, 2004), 7.

ability¹⁴. In other words, English test in this study is a method to measure students' ability that focus on specific competencies in the field of english language. It includes in three topics about principle of language assessment, standardized test, and designing classroom language test.

3. Constructing English test: construct is identified or created highly prized abilities as four skills, critical thinking, and creative activity¹⁵. For this study defined designing the test particularly in the field of English language that focus on specific competencies and involves a set of procedures that conducts outside class.
4. PeerWise: a web-based provides a framework for students to work collaboratively with their classmates to create a large repository of multiple-choice questions¹⁶. This study concluded as free software of online learning tool to make multiple-choice questions that provides students' explanation related to the alternatives.

¹⁴ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 3.

¹⁵ Thomas M. Haladyna, *Developing and Validating Multiple-Choice Tests Item* (London: Lawrence Erlbaum Associates, 2004), 5.

¹⁶ Paul Denny, Beth Simon, et.al, "Evaluation of PeerWise as An Educational Tool for Bioengineers". *American Society for Engineering Education*, 2010.

CHAPTER II

REVIEW OF RELATED LITERATURE

In this chapter attends literature review containing test theories, types of test based on the scoring of test, multiple choice test, standard criteria in constructing multiple choice questions, Peerwise application and how the way to use the application. The next part is a review of previous research that illustrates similarities and differences between this study and also research of others.

A. Theoretical Framework

1. Test

A test is a special form of assessment. Brown said that test is “*a method of measuring a person’s ability, knowledge, or performance in a given domain*”¹⁷. It means that a technique to assess the test-taker capability in specific area or competency. This competency is determined by one or more criteria that related to the rule of scoring. It helps to create a test score based on the test item.

In this case, there are some types of test based on the scoring of tests¹⁸:

a. Subjective testing

Subjective tests establish somebody’s opinion about candidate performance such as oral test or written composition¹⁹. It contains the descriptions of language. It needs rating scales to assess that test.

According to Michael Harris and Paul McCann, 1994 said that to assess this kind of test, it is important to achieve intra-rater reliability. It makes sure the teachers to rate the test consistently²⁰. In addition, inter-rater reliability is used to ensure assessing of all raters in the

¹⁷ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 3.

¹⁸ J. B. Heaton, *Writing English Language Tests* (London and New York: Longman Group UK, 1988), 25.

¹⁹ Michael Harris and Paul McCann, *Assessment* (Macmillan Publishers Ltd, 1994), 56.

²⁰ Michael Harris and Paul McCann, *Assessment* (Macmillan Publishers Ltd, 1994), 57.

same way and all raters agree on the interpretation and meaning of the descriptions in rating scales.

b. Objective testing

The objective test is one of important testing in large numbers of candidates that can be scored mechanically by computer. It usually has one correct answer and the form of test more lead to the certain skills and areas of language that more effectively than another method²¹.

The areas of language in this test is wider sample of grammar, vocabulary and phonology. Most of item test that often used for this type is multiple-choice item which has detail criteria. So, it is not quite easy to construct and needed more time to design. But, it is benefit for learning process especially in teaching and testing to measure the ability or someone knowledge²².

2. Multiple-choice

Multiple-choice is a part of objective test that significant to more careful preparation. So, it is difficult to design in good way²³. It is caused by the weaknesses of test such as the technique of constructing tests particularly in the form of multiple-choice test. For that, to establish the format of multiple-choice test construction, it needs some criteria of good multiple-choice that include:

a. Instruction

Testing should consider to the instruction term in planning the test item construction. It plays a significant role to improve the effectiveness of information in deciding the outcome²⁴. This criteria gains the

²¹ J. B. Heaton, *Writing English Language Tests* (London and New York: Longman Group UK, 1988), 25.

²² J. B. Heaton, *Writing English Language Tests* (London and New York: Longman Group UK, 1988), 25.

²³ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (California: San Francisco, 2004), 55.

²⁴ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 2.

measurement of students achievement in understanding the direction²⁵. The item in this stage should construct clearly and directly to the learning outcome. It should be consists of aspects of knowledge that be measured and focus on the important point that illustrates the material outcome²⁶.

b. Stem

It is the stimulus to present the knowledge dealing with a problem or asks a question²⁷. It covers alternatives as possible solution to the problem. For example²⁸:

Poor: A table of specifications:

- A. indicates how a test will be used to improve learning.
- B. provides a more balanced sampling of content.
- C. arranges the instructional objectives in order of their important.
- D. specifies the methods of scoring to be used on a test.

Better: What is the main advantage of using a table of specifications when preparing an achievement test?

- A. It reduces the amount of time required.
- B. It improves the sampling of content.
- C. It makes the construction of test items easier.
- D. It increases the objectivity of the test.

²⁵ Thomas M. Haladyna, *Developing and Validating Multiple-Choice Tests Item* (London: Lawrence Erlbaum Associates, 2004), 67.

²⁶ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 39.

²⁷ Steven J. Osterlind, *Constructing Test Items* (London:1989), 175.

²⁸ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 39.

The poor one is the true-false statements with a common stem which is presented clear enough in the stem. While, the better one presented a single problem in the stem.

The problem in the stem of multiple-choice item should state some points. Those are presenting a single problem that states a single idea of item, the item in positive form so it should avoid the neagive phrase or form in constructing the stem, and grammatically consistent which should construct the stem in grammatically correct particularly in terms of verb tense or the use of article.

c. Alternatives

This defined the appropriate answer based on the stem. It usually followed by five response options that consist of one correct answer and distractors²⁹. The correct answer is only undeniably answer and distractor is unquestionably wrong answer. For example³⁰:

Poor: The recall of factual information can be measured best with a:

- A. matching item.
- B. multiple-choice item.
- C. short answer item.
- D. essay question.

Better: The recall of factual information can be measured best with:

- A. matching item.
- B. multiple-choice item.
- C. short answer item.
- D. essay question.

²⁹ Steven J. Osterlind, *Constructing Test Items* (London:1989), 175.

³⁰ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 45.

The first one is using indefinite article “a”. It makes the last distractor wrong. While, the second one is by changing the alternatives from singular to plural. It is possible to omit the article.

As described, there are some criteria in writing the alternatives including correct answer and distractors that consist of one key answer and three wrong answers. In addition, the correct answer and distractors should be unquestionable which have clear language, understanding, and reasonable. Importantly, the construction of distractors should have the same form of word as the correct answer, do not state the grammatical clues, and also have the same level of difficulty with the key answer³¹. On the other hands, the alternatives should make grammatically consistent particularly in terms of verb tense. The last part avoid the alternatives “all of the above” and “none of the above”³².

In addition, in constructing the alternatives of multiple-choice item should state the information for the aim of determining the correct decision related to the alternatives of questions³³. The correct decision has quality information dealing with the academic reason based on the material. It is useful for students who answer incorrectly.

3. PeerWise

PeerWise is web-based multiple-choice question repository built by student. The students develop new questions that consist of stem and options with associated explanations. In addition, it also answers the existing questions and rates

³¹ Harold S. Madsen, *Techniques in Testing* (New York:1983), 19.

³² Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 45.

³³ Lyle F. Bachman, *Fundamental Considerations in Language Testing* (Oxford University Press, New York: 1990), 22.

them for quality and difficulty which students take part in discussions that compete with other students to appear on leaderboards³⁴. It is as the media to do collaborative learning and also peer learning for students to another.

To apply this media in constructing multiple-choice can be through the steps below:

1. Creating the account

It must visit the address of PeerWise on <http://peerwise.cs.auckland.ac.nz/at/?uinsby.id>. Then, it will appear the front layout of PeerWise application.

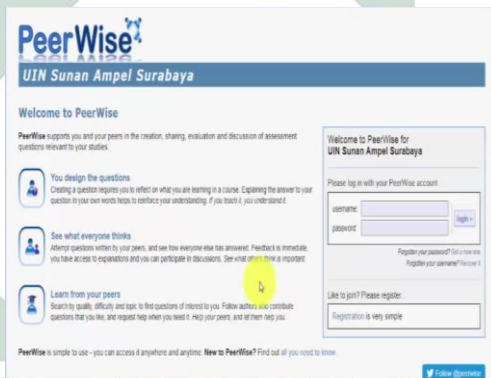


Figure 2.1 Front Layout of PeerWise Application

The table 2.1 shows the front layout of PeerWise that consist of username and password menu. The function of this menu is to login the account by filling the username and password in the box that has prepared. After that, it will appear the courses and select a course that you have joined.

After logging in and selecting the course, it will be shown the main menu for that course. The “Main Menu” is divided into the following areas: “Your Question” means that questions have been contributed, “Answered

³⁴ Paul Denny, *PeerWise* (New Zealand: The University of Auckland, 2010)

Questions” defined questions that have been answered, Unanswered Questions is all the questions that have not yet answered, Reputation Score means that the score of the multiple-choice construction that have been assessed from the other students.

2. Constructing the test

It contributes a new question by selecting the menu “View” in the box that located below the menu of “Your questions” then click menu “create new question”. After that, it will be shown some parts of writing a new question process:

a. Providing the question text:

It constructs the stem that should be clear and not likely to be ambiguous. It also uses professional language to avoid slang or spelling mistakes.

b. Providing the alternatives:

It is the answer of the question that usually consist of four options with indicating the correct answer that must be only one and exploring distractors. Ideally, each distractor should correspond to a misconception that other students may have regarding the question.

c. Providing an explanation:

This is very important and will serve to help anyone who answers the question incorrectly to understand what they have done wrong. The quality of the explanation is one of the criteria for how the question will be rated. It might like to provide an explanation corresponding to each one of the distractors for the question.

d. Associating “topics” or “tags”

The function of “tag” menu is to determine the certain topics related to the question. It can either select from a list of current topics or add new topics. This may help others to locate questions of interest.

3. Publication

If test-taker have created the question and ready to share the question, it should click the menu “show me a preview of this question” then PeerWise application will show all parts of the question. In addition, there are also some menu before writers decide to publish their questions:

- a. Menu to change or save your question. It is useful if there is a mistake with the question so directly can be changed it by selecting the menu “go back and make change”.
- b. Menu to share your question. If the questions have already to publish then click the menu “save question”. The last, it will be shown the form of the question that is followed by some aspects to assess the question.

B. Previous Studies

The first study, Fadhila Alfany discussed about how the correlation between students’ participation in PeerWise and also their achievement in the end of the course. This study involved 105 students of English education department who used PeerWise application. The data were collected by documentation and analyzed by the quantitative correlational design to find out the significant correlation between students’ participation and also their achievement in PeerWise. The result of this study is the correlation between students participation and their achievement score as pearson correlation had positive correlation³⁵. This study has same area in using PeerWise application. But, this study is different with researcher’s study which is focus on the multiple-choice construction in PeerWise than students’ participation and their achievement in PeerWise. The method of this study is also different that researcher focused on qualitative approach.

³⁵ Fadhila Alfany, “*The Correlation Between Students’ Participation and Their Examination Score in PeerWise Online Learning at English Education Department UIN Sunan Ampel Surabaya*” (Surabaya: UINSA, 2016).

The second study by Denis Duret et al who examined PeerWise affection in students' performance during the end of year 4 exam. This study used quantitative by distributing the questionnaire to identify how affected of PeerWise. In addition, it also used qualitative design by documentation analysis that focused on the students' comments to evaluate how students' contribution in PeerWise. The finding showed that PeerWise affection could improve students' course performance in the end of year 4 exam³⁶. This study focused on the influence of PeerWise on students' performance while the resercher study more focused on students' multiple-choice construction in PeerWise. Meanwhile, it is in the same domain particularly in using PeerWise.

The third study discussed by Farkad A. Kadir et al about the relationship between students' activity and their feedback in the use of PeerWise to enhance the academic performance. The data collection is from 120 students taking degrees in MBBS then analyzed by descriptive quantitative design which using questionnaire instrument to evaluate students' contribution and their perception on PeerWise. The result of this study has positive correlation between the students' activity and their feedback which shown by students posting and answering on peer questions³⁷. This study analyzed how Peewise could increase the students' academic performance based on their activity and feedback. While it has same focus in using PeerWise application, it is different with researcher study that focused on stdents' english test construction especially in multiple-choice made test through PeerWise.

The fourth study by Ni Putu Sintya Winata et.al that discussed about the appropriateness of grid in multiple-choice test item with the implementation of multiple-choice test in Indonesian Language subject. This study used qualitative descriptive to analyze the data which is taken from the teacher of subject. The

³⁶ Denis Duret et al. "Collaborative Learning With Peewise". *Association for Learning Technology*, 2018

³⁷ Farkad A. Kadir et al. "The Impact of Peewise Approach on The Academic Performance of Medical Students". *MOJET*, 2016

format of data is three teachers' item test construction for general examination in odd semester. The result of this study is the teachers implementation in making multiple-choice test have not been yet appropriate with the grid of multiple-choice test³⁸. This research has same concern with the researcher study particularly in multiple-choice test item but the theory and subject are different from the researcher study. This study focused on the quality of the test based on the grid which is measuring the level of item difficulty. While, the researcher study focuses on the aspects that should be fulfilled in constructing multiple-choice test.

The fifth of this research focused on measuring the validity of multiple-choice questions for testing the students' ability by Mehmet Dulger and Hasan Deniz. This study provided the students' explanation of their thinking in answering for each question that is done by 28 students through interview technique. The finding of this study is students answer in multiple-choice test do not fully reflect students' true competency that the test is claiming to measure³⁹. In this case, the study determined the influence of multiple-choice in measuring the students' competency which explore how valid of multiple-choice test. So, it is different with researcher study that explore how the students' construction in making multiple-choice test based on some criteria.

The sixth study by Elyza Martiarini who identified the different result of students' reading comprehension using multiple-choice and cloze task. The data were collected by using pre-test and post-test that involved 30 students of the fourth semester of English Department. Then, it was analyzed by comparative descriptive method. The result of this study is between multiple-choice and cloze test have significant differences which is the students' reading comprehension used multiple-choice was higher

³⁸ Ni Putu Sintya Winata, "Analisis Butir Soal Pilihan Ganda Mata Pelajaran Bahasa Indonesia SMK Negeri 3 Singaraja". 2014

³⁹ Mehmet Dulger-Hasan Deniz, "Assessing The Validity of Multiple-Choice Questions in Measuring Fourth Graders' Ability to Interpret Graphs about Motion and Temperature". *Academic Publisher*, 2017

than cloze task⁴⁰. It focused on the effectiveness of multiple-choice and cloze task in measuring the students' reading comprehension. It contrast with the resercher observation that will analyze the same term in form of multiple-choice test but it is more lead to the construction.

As a result, it can be known that most of research discussed about students' participation in using PeerWise application and also how the score that can be gotten in final examination if they use it. In this case, this application is supposed become effective tool in developing students' learning. On the other hands, some researchers also do observation how the measurement of multiple-choice to assess students' competence. From all of them, it can be summarized as the material to be observed with new creation by combining both of specific concern that is PeerWise application and multiple-choice. So, this study discusses how students construct English test that focus on multiple-choice made test through PeerWise.

⁴⁰ Elyza Martiarini, "Constructing Multiple-Choice and Cloze Task On Student's Reading Comprehension". *University of Indraprasta PGRI*, 2017

CHAPTER III RESEARCH METHOD

The part of study provides the method of research by exploring English test construction in form of multiple-choice test through PeerWise. This is followed by design of study, research arrangements, data form and data source, instruments, the technique to collect the data, technique to analyze data, and stages of research.

A. Research Design

This research uses qualitative approach. This approach explores the students' construction in english test through PeerWise. It is combined some procedures to get the result of data that appropriate with the research. The procedures include reviewing all of the data so it can be logic and arranging the data into specific category that influence to all sources.⁴¹

This study examines the students' english test construction especially in form of multiple-choice test. It is suitable to use descriptive design. It has three main purposes of research consist of describing, explaining, and validating the findings. This design is to provide an accurate and valid the representation of the variables that relevant to the research questions. It can be executed by the critical thinking to support the result finding. Therefore, it uses qualitative-descriptive research design to describe and analyze how students' ability in constructing multiple-choice questions through PeerWise.

B. Research Setting

This research conducts in State Islamic University of Sunan Ampel Surabaya which is located at Jl. Ahmad Yani 117 Surabaya. Equally, this university has some faculties, one of them is Faculty of Teaching and Teachers Training that will be became place of researcher's observation. Certainly, this observation is purposed for English Education Department.

The subject of this study is English Education Department students at State Islamic University of Sunan Ampel. This study

⁴¹ John W. Creswell, *Research Design* (USA: SAGE Publications.Inc, 2009), 175.

conducts in language testing and assessment subject which consists of thirty students. It is chosen because this class who do English test by constructing english test that focus on multiple-choice. The learning process of this subject is using online learning strategy. They used PeerWise as web-based application for their media learning. On the other hands, this study takes three topics from specific material of students made test to be observed which every topic take the multiple-choice question based on the result of students rating.

C. Data and Source of Data

The data for this study is the students' multiple-choice works. The students make the multiple-choice questions task as the daily assignment in every week after they got the material in language testing and assessment class. This study analyzes the multiple-choice questions that have been made by students in PeerWise.

The source of data can be gained from students works of English teacher education department students in State Islamic University of Sunan Ampel Surabaya. The data takes from language testing and assessment class who construct multiple-choice question in PeerWise. But, it needs teacher's helping to get data as the server of PeerWise application.

D. Research Instruments

This study uses a documentation as the research instrument. The content of documents is students' english test construction that focus on multiple-choice questions. In other words, this instrument to know how do students construct english test especially in multiple-choice test. It consists of instruction, stem, and alternatives.

It uses document analysis to identify the students' ability in constructing english test especially multiple-choice tasks. It is more suitable to assess students' ability in constructing the multiple-choice questions by using guideline analysis that is adapted from Norman E. Gronlund theory about the criteria in constructing good multiple-choice through PeerWise. There are three categories

to construct multiple-choice item that consists of instruction, stem, and alternatives which every category has some criteria to analyze the students' multiple-choice works. This category as described as in theoretical framework. This study uses three categories include instruction with two criteria, stem with three criteria, and alternatives with five criteria. However, there is addition category related to the PeerWise application that stated explanation. In this case, explanation is defined the information for the aim of determining the correct decision related to the alternatives of questions.

E. Data Collection Technique

The researcher collects the data by some steps. The first step, the researcher asks permission to the lecturer as the teacher of the subject in Language testing and assessment class. In this matter, the researcher explains the purpose of this study and what the researcher want to do in this study. The second step, after getting permission from the lecturer, the reseracher makes appointment with the lecturer to discuss about the researcher's study and get the data of students' multiple-choice test that exist in PeerWise application. In addition, PeerWise application is learning tool that automatically able to save or keep students' multiple-choice questions. So, the data can be collected by the server. It is the lecturer of language testing and assessment subject. It makes easy for researcher to get the document of students' multiple-choice questions that is taken from PeerWise its self through the lecturer's transcript.

F. Data Analysis Technique

From the discussion above, this study uses descriptive qualitative method. So, there are some steps to analyze the data depends on the qualitative research. The data of this study applies as the following steps based on Creswell's theory⁴²:

⁴² John W. Creswell, *Research Design* (USA: SAGE Publications.Inc, 2009), 185.

1. Organize and prepare the data

The researcher prepares the data that is gotten from the lecturer of subject as the center server of PeerWise application. The data is students' English test construction particularly in form of multiple-choice questions. It is arranging and collecting the data depending on the source of the information.

2. Read through all the data

The researcher reads the students multiple-choice works that have been made in PeerWise. It focuses on the main idea in form of multiple-choice construction to obtain the information and overall meaning of the data. While reading the data, the researcher writes some notes to record the general thought about students english test construction especially in form of multiple-choice question.

3. Coding the data

The researcher gives label the categories based on the form of multiple-choice construction that related to the theory. This stage analyzes 50 multiple choice questions that distinguish which one the category of instruction, stem, and alternatives that consist of correct answer and distractor. In addition, both are added with the explanation based on the students' multiple-choice works. This labelling is marked with different color for every category. It relates to the Norman E Gronlund theory about the criteria to construct good multiple-choice question.

4. Interpreting the meaning of the data

The researcher interprets the idea of data that is captured in students multiple-choice works in stage of coding data. The essence of the idea which is to be the reseracher's personal interpretation that relates to the expert's theory in understanding the students English test construction in form of multiple-choice question. In addition, it concludes the finding information of the whole research that focus on how do

students construct english test through PeerWise especially in multiple-choice question which conducts with some categories including instruction, stem, and alternatives with adding the explanation.

G. Research Stages

The process of this study explains as these following stages:

Before doing the research, the researcher have used PeerWise application and automatically knew how to use and operate it. From this matter, the researcher found problem that usually appear when students made test in PeerWise until the researcher designed the title and research question based on the lecturer's suggestion. To make the research completed, the researcher designed along with all of the content in research depend on the correct guideline.

Then, the researcher asked permission to the lecturer of Language Testing and Assessment class to collect the data because the center server that can open the application is the lecturer. After getting the data, the researcher conducted the study by analyzing document of students made test in PeerWise. The researcher of this study analyzed the data by using documentation analysis guideline. In this term, the researcher coding the data to classify the items based on the validated criteria in constructing test particularly in form of multiple-choice. Then, interpreting the meaning of coding data through make description and explain what the mentioned in coding of data. The last step decided the result of the study by giving conclusion and report of the research based on the analysis data.

CHAPTER IV RESEARCH FINDING AND DISCUSSION

This section discusses the analysis result that related to exploring students' ability in constructing English test through PeerWise. It mentions result analysis in finding then consider to students' ability in constructing english test particularly in form of multiple-choice test whether the data meet the standard criteria in constructing good Multiple Choice questions guidelines or not in terms of instruction, stem, alternatives.

A. Research Findings

To find out about students' ability in constructing English test through PeerWise particularly the test construction was in form of multiple choice questions, this research arranged based on the research question: How do students construct English test through PeerWise. The result of this study explained descriptively that was supported by the document of students works.

The researcher gained the data from English department student in language testing and assessment class who used PeerWise application in constructing English test particularly in form of multiple-choice question. The data divided into 3 topics which the first topic about *The Principle of Language Assessment* contained 20 questions, the second topic about *Standardized Test* contained 15 questions, and the third topic about *Designing Classroom Language Test* contained 15 questions. Totally, there were 50 questions of students made test in PeerWise.

The following is the finding of the students' ability in constructing English test particularly in form of multiple-choice question through PeerWise which categorized in terms of instruction, stem, and alternatives:

1. Instruction

The result of analysis of the items indicated that in general from 3 topics that consist of 50 questions totally did not have instruction. All of the items from 3 topics are directly to the question as the following example:

Table 4.1 Example of Multiple Choice which has not Instruction (Topic 1)

Field	Value
NO.	1
Question	Which one the disadvantage of use authentic material in language testing?
A	Topic are meaningful
B	Real-word task
C	Provide authentic cultural information
D	Need more teacher time and creativity to prepare
Explanation	a,b,c are the advantage of use authentic material (pg. 28)

As displayed in table 4.1 above, it showed that this question did not has instruction which this items directly stated the stem. The stem of question stated about the disadvantage of use authentic material in language testing.

Table 4.2 Example of Multiple Choice which has not Instruction (Topic 2)

Field	Value
NO.	1
Question	What kind of standardized test does refer to gate-keeping essay test to take graduate level course?
A	GET
B	TOEFL
C	ESLPT
D	ETS
Explanation	the answer is "A" because GET (The Graduate Essay Test) includes standardized test that refers to gate-keeping essay test to take graduate level course (see Brown's book p.69)

Based on the table 4.2 above, it described that this question stated the stem directly without included the instruction. The stem of question stated about the kind of standardized test that refer to gate-keeping essay.

Table 4.3 Example of Multiple Choice which has not Instruction (Topic 3)

Field	Value
NO.	1
Question	There are three parts of the english as a second language placement test at San Francisco State University, except. ...
A	Students writes a short articles and reads a summary essay
B	Students writes a response to an article
C	Students reads an essay and analyses grammar errors (multiple choice)
D	Students reads a short articles and writes a summary essay
Explanation	The answer is A because the first part is the students reads a short articles and writes a summary essay

As shown in table 4.3 above, it can be seen that this example directly related to the question and did not mentioned the instruction clearly. This question focused on the parts of the English as a second language of placement test.

2. The Stem Construction

The analysis of this study found that all of the questions from 3 topics have stem. In addition, the stem of those questions divided into 4 parts of the stem that covered as following table based on the students made test in PeerWise:

Table 4.4 Parts of Stem in Multiple-Choice Construction

No.	The Parts of Stem
1.	Present a single problem
2.	Positive phrase/form

3.	Negative phrase/form
4.	Grammatical Error

Based on the table 4.4 above, there were four parts of stem in multiple-choice question which have been constructed by the students. In this case, all of those parts explained based on the explanation below:

a. Present a single problem

All of the multiple-choice items from 3 topics presented a single problem that directly to the idea in each stem. The following is One of the example of items that presented a single problem in term of stem based on 3 topics:

Table 4.5 Example of Multiple Choice Stem which presents a single problem (Topic 1)

Field	Value
NO.	10
Question	Which statement does refer to <u>characteristic of practical test?</u> STEM
A	Test takers put the great expense
B	Examiner has constriction in scoring the test
C	The amount of question fatigues test takers
D	The test appropriates consecutive procedure and time consideration
Explanation	The answer is D because efficient time and organized procedure includes the practicality of the test

As displayed in table 4.5 above, it can be seen that the item has stem (as underlined by the green line) which presented a single problem particularly focus on a single idea. In this matter, the stem stated to find one of

statement in the options that appropriate with the characteristic of practical test.

Table 4.6 Example of Multiple Choice Stem which presents a single problem (Topic 2)

Field	Value
NO.	2
Question	<u>What is the type of decision that use to determine standardized test for the students?:</u> STEM
A	Diagnosis test
B	Placement test
C	Progress test
D	Achievement test
Explanation	Because placement test is one of test that use to determine the standardized of the student's test while diagnosis, progress, and achievement test is type of classroom test.

Based on the table 4.6 above, it described that this question stated the stem (as underlined by the green line). It was directly to the single idea. The stem of question stated about the type of decision that use to determine standardized test for students. It asked to find what kind of test based on the statement in the stem.

Table 4.7 Example of Multiple Choice Stem which presents a single problem (Topic 3)

Field	Value
NO.	2
Question	<u>What tests have a direct relationship to the curriculum of learning?</u> STEM
A	Diagnostic test
B	Achievement test
C	Placement test
D	Proficiency test
Explanation	it is because achievement test is a test related directly to classroom lessons, units, or even a total curriculum. (p 47)

As shown in table 4.7 above, it can be seen that the stem of this example directly related to single idea (as underlined by the green line). This question focused on the kind of test depend on the statement in the stem. It stated “the tests have a direct relationship to the curriculum of learning”.

b. Positive phrase/form

The finding of this study found that the stem of all of items have positive phrase/form. Based on this aspect, this study got the result that there are 41 multiple-choice questions in positive form from 3 topics. In addition, those multiple-choice items stated in positive form but in different category of stem as described in the table below based on students works in peerwise:

Table 4.8 Category of Positive Stem in Multiple-Choice Construction

No.	The Category of Positive Stem
1.	W-H Question
2.	Completion form
3.	Other formatting

As shown in the table 4.8 above, there were 3 categories of stem in multiple-choice construction that have been made by students in PeerWise. The following is the explanation of each point based on the category above:

1) W-H Question

Based on the finding that there were 31 items from 3 topics which have positive stem in form of W-H Question. All of those items divided into 10 multiple-choice items from topic 1, 10 multiple-choice items from topic 2, and 11 multiple-choice items from topic 3. The following are the examples of each topic that relates to the W-H Question stem:

Table 4.9 Example of Multiple Choice which presents Positive Stem in Form of W-H Question (Topic 1)

Field	Value
NO.	17
Question	What aspect can be removed in principle language assessment?
A	wash back
B	practicality
C	validity
D	reliability
Explanation	the correct answer is A, see brown pg 19
Tags	Chapter 2

Stem in form of question

Based on the table 4.9 above, it showed that the item stated the positive phrase of stem. In addition, it was in the form of questions (as underlined by green line) that can be indicated through W-H Question that is shown in the word “what” (as circled in orange) and question mark (as circled in yellow). The stem asked about one of unimportant aspect in principle language assessment.

Table 4.10 Example of Multiple Choice which presents Positive Stem in Form of W-H Question (Topic 2)

Field	Value
NO.	8
Question	<u>How many forms of test bias that involve the standardized tests?</u>
A	2
B	3
C	5
D	6
Explanation	The answer is C because there are five forms of test bias, among them are language, culture, race, gender, and learning styles
Tags	chapter 5, Standards based assessment

Stem in form of question

As displayed in the table 4.10 above, it showed that the item stated the positive phrase of stem. In addition, it was in the form of questions (as underlined by green line) that can be indicated through W-H Question that is shown in the word “How” (as circled in orange) and added by the question mark (as circled in yellow). This stem asked about the total number of test bias forms particularly in the standardized test.

Table 4.11 Example of Multiple Choice which presents Positive Stem in Form of W-H Question (Topic 3)

Field	Value
NO.	3
Question	<u>What kinds of test that use to test the global competence in a language?</u>
A	Summative and formative
B	Formative and informal
C	Informal and norm-referenced
D	Norm-referenced and summative
Explanation	The answer is D because in brown's book the proficiency test are almost always norm-referenced and summative.

Stem in form of question

As described in the table 4.11 above, it showed that the item stated the positive form of stem. In addition, it has questions form (as underlined by green line) that can be indicated through W-H Question that is shown in the word “what” (as circled in orange) and added by the question mark (as circled in yellow). Actually, the stem of item stated the kind of test to measure the global competence in a language.

2) Completion Form

The result of this study indicated that there were 8 items from 3 topics which have positive stem in completion form. All of those items divided into 4 multiple-choice items from topic 1, 2 multiple-choice items from topic 2, and 2 multiple-choice items from topic 3. The following are the examples of each topic that relates to the completion form:

Table 4.12 Example of Multiple Choice which presents Positive Stem in Completion Form (Topic 1)

Field	Value
NO.	8
Question	<u>If the student can complete the test reasonably within the set time frame, it means they do the Test.</u>
A	Practicality
B	Valid
C	Authentic
D	A,B,C are correct
Explanation	The answer is A, you can read the test procedures practical for make it clear, the practical test need students can complete the test reasonably. (Page 31)
Tags	Principle of language assesment

Stem in completion form

As displayed in table 4.12 that the example of this item was the positive phrase of stem in the completion form. It showed as underlined by the green line and can be indicated from the missing word (as circled in orange color) in the middle of sentence stem. So, it should be completed to become the complete sentence. The test taker should fill the blank word by choosing one of the correct answer in the options.

Table 4.13 Example of Multiple Choice which presents Positive Stem in Completion Form (Topic 2)

Field	Value
NO.	12
Question	<u>A main issue in validating large-scale standardized tests of proficiency is...</u>
A	Content Validity
B	Construct Validity
C	Consequential Validity
D	Criterion Validity
Explanation	page 25. B is the correct answer because such test must for economic reasons, and it's related with the explanation of construct validity. The explanation of construct validity according to the book is a major issue in validating large-scale standardized tests of proficiency
Tags	chapter 2

Stem in completion form

From the table 4.13 above explained that the example of this multiple-choice item was the positive phrase of stem which stated as completion format. It showed as underlined by the green line and can be indicated from the periods in the end of sentence as circled in orange color. It needed the answer to make complete the sentence of stem through selecting one of options that appropriate.

Table 4.14 Example of Multiple Choice which presents Positive Stem in Completion Form (Topic 3)

Field	Value
NO.	7
Question	<p><u>The students will write a paragraph story about the experience in the past.</u></p> <p><u>This sentence is example of ...</u></p>
A	The purpose of the test
B	The objective of the test
C	The specifications of the test
D	Selecting and separating items arrangements
Explanation	The correct answer is B because in designing the objective of the test should consist of performance and linguistic domain. In that sentence the performance is writing and the linguistic domain is grammar.
Tags	Designing classroom language

The table 4.14 above described that this multiple-choice item was the positive phrase of stem which gathered by the passage before the stem (as underlined by the purple line). In addition, the formatting of the positive stem stated as completion format. It showed as underlined by the green line and can be indicated from the periods in the end of sentence as circled in orange color. It asked to know the meaning of the passage of stem above that related to the example of what kind of test.

3) Other Formatting

The finding of the study found that there were 2 multiple-choice items from topic 1 in positive stem but have different format sentence which are not found in other topic. The format sentence of the stem is

improper formatting and form of instruction. As described in the following example below:

Table 4.15 Example of Multiple Choice which Presents Positive Stem of Improper Formatting

Field	Value
NO.	13
Question	<u>Temporary illness, anxiety, fatigue, and psychological factors are part of reliability. It includes which part of reliability?</u>
A	Test reliability
B	Rater reliability
C	Student-related reliability
D	Test administration reliability
Explanation	The correct answer is C. it is influenced by students' emotional and psychological factors when students are doing the questions provided. for the example temporary illness, anxiety, fatigue, and etc
Tags	Principle of language assessment

The table 4.15 described that the example of item above constructed the positive stem which the stem has passage (as underlined by the purple line), but the word arrangement of the stem is more directed at Indonesian so it made improper formatting (as underlined by the green line). Besides that, it was added the questions mark (as circled in orange) while the stem did not in form of question. This stem asked what kind of reliability that appropriate with the passage but the structure in constructing the stem intended to the questions form. It stated as like as Indonesian item construction.

Table 4.16 Example of Multiple Choice which presents Positive Stem in Form of Instruction

Field	Value
NO.	3
Question	<u>Choose the best examples of summative test and formative test !</u> Stem in form of instruction
A	Quiz (formative) and periodic review test (Formative)
B	Final test (Summative) and Middle test (Formative)
C	Daily task (Summative) and quiz (Formative)
D	Final exam (Summative) and Student's practice speaking (Formative)
Explanation	The answer is D, because Summative is one kind of assessment that assess the student in the end (bad, good, medium) and formative assessment is the assessment that evaluate a process learning of students.
Tags	Summative, principle of language assessment, week2

Based on the table 4.16 showed that the stem of item used positive phrase (as underlined by the green line) while the format sentence of the stem used imperative sentence as same as the instruction. It can be indicated from the bare infinitive “choose” as the verb in the beginning of sentence (as circled in orange) and also the exclamation mark (as circled in yellow). However, it still called as the stem because it stated the single problem or idea while the form of the stem in instruction formatting.

c. Negative Phrase/Form

In this matter, the result of the finding indicated from 3 topics, there were 9 items of questions in PeerWise which have been made by students in form of negative

phrase. The following example stated one of the item from each topic that related to negative stem:

Table 4.17 Example of Multiple Choice which presents The Stem in Negative Phrase (Topic 1)

Field	Value
NO.	9
Question	<p><u>These are characteristics that show a test have authenticity</u> except.</p> <p>Stem in negative form</p>
A	Have some thematic organization , such as story line or episode
B	Items are isolated rather than contextualized
C	Task represent, or closely approximate, real world task
D	The Language is as natural as possible
Explanation	Items should be contextualized rather than isolated
Tags	Principle of language <u>assesment</u> . week3

As displayed in the table 4.17 above, the example of item showed that the stem of item focused on the characteristics of the authenticity test (as underlined by the green line). But, the stem formatting used negative phrase. This was evidenced by the word “except” in the end of the sentence of stem (as circled in orange).

Table 4.18 Example of Multiple Choice which presents The Stem in Negative Phrase (Topic 2)

Field	Value
NO.	4
Question	<u>This is the characteristic of standardized test</u> , except
A	Give score to students
B	The score report strategies is give answer to students
C	Score interpretation is percentile
D	Level of generality is global
Explanation	The correct answer is <u>B</u> . Because give answers to students include in classroom test characteristic
Tags	Test characteristic

In the table 4.18 above, the example of multiple-choice item has the negative stem (as underlined by the green line) which can be indicated from the content of the stem. In this case, the stem stated to look for the characteristics that does not include in standardized test. Besides that, in the end of the sentence of stem was proved by the word “except” (as circled in orange).

Table 4.19 Example of Multiple Choice which presents The Stem in Negative Phrase (Topic 3)

Field	Value
NO.	1
Question	<p><u>There are three parts of the English as a second language placement test at San Francisco State University, except . . .</u></p>
A	Students writes a short articles and reads a summary essay
B	Students writes a response to an article
C	Students reads an essay and analyses grammar errors (multiple choice)
D	Students reads a short articles and writes a summary essay
Explanation	The answer is A because the first part is the students reads a short articles and writes a summary essay
Tags	Week5, designing classroom language test

The example of multiple-choice item based on the table 4.19 above stated the negative stem (as underlined by the green line) which can be indicated from the the word “except” (as circled in orange). Indirectly, the stem of this item stated the negative meaning which can be seen from the content. It asked to look for the example that does not include in parts of English as second language placement test.

d. Grammatical Error

This study got the result how the students constructed the stem that related to the grammar. In this matter, the students often did not pay attention to the structure of the sentence when constructing the stem of multiple choice especially in terms of grammar. So, in this

study found some mistakes from 3 topics which divided into four categories that are shown in the table below:

Table 4.20 Grammar Error Category in Constructing Stem

No.	Grammar Error Category
1.	Verb Tenses
2.	Article
3.	Singular/Plural Noun
4.	Preposition

As shown above, the table 4.20 described the grammar error category that have been made by the students in constructing the stem of multiple-choice item. It categorized into four categories of grammar error that focused on verb tenses, article, plural/singular noun, and preposition based on the students made test in peerwise. The following example of item showed each parts of the grammar error category from 3 topics:

1) Verb Tenses

There are 6 questions of multiple choice in term of stem did not have grammatical consistency particularly in verb tenses category. It can be seen in the following example:

Table 4.21 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Category (Topic 1)

Field	Value
NO.	1 Verb Tense
Question	<u>Which one the disadvantage of use authentic material in language testing?</u> STEM
A	Topic are meaningful
B	Real-word task
C	Provide authentic cultural information
D	Need more teacher time and creativity to prepare
Explanation	a,b,c are the advantage of use authentic material (pg. 28)
Tags	week2, Chapter 2, Principle of language assessment

As displayed in the table 4.21 above, the students constructed uncomplete sentence in the stem of this item (as underlined by the green line). In this matter, the stem was less of the verb tense (as shown by a red circle) based on the grammar error category that appeared. So, it should be completed by adding tobe “is” as the verb to make the proper sentence.

Table 4.22 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Category (Topic 2)

Field	Value
NO.	2 Verb Tense
Question	<u>What are the type of decision that use to determine standardized test for the students?</u> STEM
A	Diagnosis test
B	Placement test
C	Progress test
D	Achievement test
Explanation	Because placement test is one of test that use to determine the standardized of the student's test while diagnosis, progress, and achievement test is type of classroom test.
Tags	Classroom vs standardized

As shown in the table 4.22 above, the stem of this item stated the complete sentence (as underlined by the green line). But it became uncomplete because there was grammar error in terms of verb tense (as shown by a red circled). In this matter, the plural verb “are” does not appear to agree with the singular subject “type”. It considered to change the verb form for subject agreement which verb “are” was changed by verb “is”.

Table 4.23 Example of Multiple Choice Stem in Term of Verb Tenses of Grammatical Error Category (Topic 3)

Field	Value
NO.	9 Verb Tense STEM
Question	<u>What is the tests you need to create?</u>
A	LAT Test, placement test, and archive achievement test
B	Placement test, diagnostic test, and achievement test
C	Placement test, diagnostic test, and Lang aptitude test
D	Placement test, diagnostic test, and LAT test
Explanation	Because the placement test, diagnostic tests, and achievement test role in regulating learning in language classes
Tags	Designing classroom language test, Week5

In the table 4.23 displayed that the stem of multiple-choice item has grammatical error in terms of verb tense as circled in red color. This term directed to the main focus of subject verb agreement error so it considered to be changed by the plural verb “are”. In this case, a plural subject of a sentence or clause requires a plural verb.

2) Article

There are 5 questions of multiple choice in term of stem which consist of 1 questions in topic 1, 2 questions in topic 2, and 2 questions in topic 3 have grammatical error. It appeared particularly in Article category. It can be seen in the following example:

Table 4.24 Example of Multiple Choice Stem in Term of Article of Grammatical Error Category (Topic 1)

Field	Value
NO.	4 Stem has passage
Question	It is used to demonstrate the accuracy of a measure of <u>procedure</u> by comparing it with another measure or procedure which has been demonstrated to be valid. That statement is constitute the aim of...
A	Content-related evidence Stem in Form of Fill in the Blank
B	Criterion-related evidence
C	Construct-related evidence
D	Consequential validity
Explanation	The right answer is B, the function of criterion-related evidence is used to demonstrate the accuracy of a measure of procedure by comparing it with another measure or procedure which has been demonstrated to be valid.

Based on the table 4.24 above, this item has problem with the grammar particularly in term of article (as shown by the red circle). It should add an article which the noun phrase “procedure” seems to be missing a determiner before it. So, it considers adding an article “the” to make a complete sentence as the stem.

Table 4.25 Example of Multiple Choice Stem in Term of Article of Grammatical Error Category (Topic 2)

Field	Value
NO.	6
Question	Which one of these is included in <u>standardized</u> test?
A	TOEFL STEM
B	Quiz
C	Journal
D	Interview
Explanation	The answer is A because TOEFL is produced by the Educational Testing Service (one of the expert, and is equal to all school and test taker)
Tags	Types of Testing, Assessment

The table 4.25 claimed that the multiple-choice stem included the grammatical error in terms of article as same as in the previous example (as shown by the red circle). Consider to this problem, it should add the article “a” before the word “standardized”. In addition, the article “a” intended to the singular nouns.

Table 4.26 Example of Multiple Choice Stem in Term of Article of Grammatical Error Category (Topic 3)

Field	Value
NO.	15 STEM
Question	Which one is good in writing the objectives of <u>test</u> ?
A	Students will have fun through a relaxed activity and thus enjoy their learning
B	Students will learn tag questions
C	Students will produce yes/no questions with final rising intonation
D	Tag questions
Explanation	A. It is bad because the learning purpose is unclear and there are not language point and performance, B. Bad because it is too general and there is no performance, C. It is good because there are two features about that included language point: yes/no questions and the performance: speaking lead to rising intonation, D. It is bad because there are not language point and performance
Tags	week 5-stages of test development

According to the table 4.26 indicated that the item of multiple choice test has stem but in uncomplete structure that missing of article. It was intended to the word “test” as circled by the red color. It appeared that an article is missing before the word “test”. So, it considered adding the article “the” before it. In addition, the article “the” can be used with singular or plural nouns to indicate something specific.

3) Singular/Plural Noun

Based on the result of finding, there are 2 questions of multiple choice in term of stem which have problem with grammar particularly in singular/plural noun category. The grammatical error of mutiple-choice found in topic 1 and topic 2 where each topics have 1 questions in grammatical error. It can be seen in the following example:

Table 4.27 Example of Multiple Choice Stem in Term of Singular/Plural Noun of Grammatical Error Category (Topic 1)

Field	Value
NO.	11
Question	What is <u>Principles</u> of language assessment that consistent and dependable? STEM
A	Practicality
B	Reliability
C	Validity
D	Washback
Explanation	The answer is B, because if we give the same test to the same student or matched on two different occasions, the test should yield similar results.
Tags	Principle of language assessment, Chapter 1

Based on the table 4.27 above, this item has problem with the grammar particularly in term of singular/plural noun (as shown by the red circle). It appears that there is to be in the form of verb “is” so the noun should be singular. Therefore, the word “principles” should become “principle”.

Table 4.28 Example of Multiple Choice Stem in Term of Singular/Plural Noun of Grammatical Error Category (Topic 2)

Field	Value
NO.	3
Question	<u>what is the test⁶ prepared well by an expert and the procedures together?</u>
A	Unstandardized test
B	Standardized Test
C	Objective test
D	Subjective test
Explanation	because the definition of standardizes a test is prepared by an expert or specialist. This type of test samples behavior under uniform procedures.
Tags	Types of Tests

According to the example above (see the table 4.28), this item indicated singular/plural noun error in terms of grammar. As circled by the red color examined the error of the item. In this case, the word “tests” shown a plural subject while the verb form was in singular verb “is”. So, it should consider to replace the form of subject become the singular subject by omitting the letter “s” in the end of the word “tests”.

4) Preposition

The finding found only 1 question of multiple-choice that have grammar error in terms of preposition. From 3 topics indicated only the third topic stated this error as displayed in the following example:

Table 4.29 Example of Multiple Choice Stem in Term of Preposition of Grammatical Error Category (Topic 3)

Field	Value
NO.	4
Question	What is the differences <u>of</u> diagnostic test and achievements test? STEM
A	Diagnostic test more concern in what students needs to continue to work in the future and achievements test analyzes the students that have acquired language features which already thought
B	Diagnostic test analyzes the students that have acquired language features which already thought and achievements test more concern in what students needs to continue to work in the future
C	Diagnostic test more concern in what students do not need to continue to work in the future and achievements test analyzes the teachers that have acquired language features which already thought
D	Diagnostic test analyzes the teachers that have acquired language features which already thought and achievements test more concern in what teachers needs to continue to work in the future
Explanation	The answer is A because actually there is similarity between diagnostic test and achievement test, it is achievement test can also serve the diagnostic role of indicating what students needs to continue to work in the future. But there is also found the difference.
Tags	designing classroom language test, Week5

The item above from the table 4.29 explained that the stem of this item has problem in terms of preposition error. It appeared that the preposition “of”

was incorrect in this context. So, it considered to be changed with “between in”. This preposition has function that indicates how a noun or noun phrase relates to the rest of sentence.

3. The Alternatives Construction

This term described the options of multiple choice construction. The analysis of this study found that all of the questions from 3 topics that totally 50 questions have alternatives. Based on the students made test in PeerWise, there are some aspects of alternatives which included in students’ multiple-choice item construction that covered as following table:

Table 4.30 The Aspects of Alternatives in Multiple-Choice Construction

No.	The Aspects of Alternatives
1.	It contains one key answer and three distractors
2.	The key answer and distractors states unquestionable formatting
3.	Grammatical Error
4.	Keep the alternatives independent

Based on the table 4.30 above, there are four aspects of alternatives that have been constructed by students through PeerWise application. All of them are in terms of key answer and distractors, unquestionable formatting, grmmatical error, and independet alternatives. In this case, all of those aspects have been described as the following examples of question in each topic below:

- a. It contains one key answer and three distractors

This study got the result that all of the multiple choice questions which totally consist of 50 questions from 3 topics have the alternatives with one correct answer and also three distractors. It is showed in the following example:

Table 4.31 Example of Multiple Choice-Alternatives in terms of Key Answer and Distractors (Topic 1)

Field	Value
NO.	19
Question	Consistent and dependent test means if there is same student do the test on two different occasions and the test yield similar result. What principle of language assessment does relate to that characteristic?
A	Practicality → Distractor
B	Reliability → Key Answer
C	Validity → Distractor
D	Authenticity → Distractor
Explanation	the answer is "b" because reliability is related to consistency of the test. However, other answers are wrong. Practicality is determined by time, cost, and administrative details. Authenticity refers to whether the test is natural and understandable for students. Validity talks about direction, structure of test, and appropriate difficulty level. (For further information, you can see Brown's book pg.20)
Total ratings	8

As shown in the table 4.31 above that the item has one correct answer (as shown by a green arrow). The correct answer is stated in "B" option namely reliability. While the distractors have three options as the wrong answer (as

shown by the blue arrow). Three distractors stated in “A” options namely practicality, “C” option namely validity, and “D” option namely authenticity. This item used word form as the options.

Table 4.32 Example of Multiple Choice-Alternatives in terms of Key Answer and Distractors (Topic 2)

Field	Value
NO.	10
Question	What is the major issue in validating large - scale standardized test of proficiency?
A	Criterion related evidence → Distractor
B	Content validity → Distractor
C	Construct validity → Key Answer
D	Content related evidence → Distractor
Explanation	You can look 25 to check your answer. Construct validity is major issue in validating large-scale standardized test of proficiency.
Tags	Chapter 2

Based on the table 4.32 above that the item has one correct answer as the key listed in “C” option that says construct validity (as shown by a green arrow). Besides that, three distractors were indicated by the blue arrow. For the first distractor wrote criterion related evidence in “A” option. The second distractor stated content validity as “B” option. And the last distractor in “D” option was content related evidence. Those were in form of noun phrase options.

Table 4.33 Example of Multiple Choice-Alternatives in terms of Key Answer and Distractors (Topic 3)

Field	Value
NO.	8
Question	What is the criteria of classroom test?
A	Consist of general knowledge → Distractor
B	Consist of international information → Distractor
C	Consist of classroom syllabus → Key Answer
D	Consist of TOEFL test → Distractor
Explanation	As what mentioned and explained by Mr. <u>Albard</u> that classroom test is a test that designing especially for the member of the class only, so the materials and the knowledge that tested must be suitable with the syllabus of the subject matter. So, it indicates that point C is the correct answer.
Tags	designing classroom language

As displayed in the table 4.33 above that the alternatives of multiple-choice item stated in form of noun phrase that contained one correct answer (as shown by a green arrow) which attached classroom syllabus as the “C” option. For other options stated as three distractors (as shown by the blue arrow). In this matter, the three distractors listed in A, B, and D option that wrote consist of general knowledge, international information, and TOEFL test.

- b. The key answer and distractors stated unquestionable formatting

The finding of this study explained the correct answer and wrong answer from 3 topics that totally consist of 50 questions have unquestionable formatting. Those can be shown in the following example from each topics:

1) First Topic

Based on the result of analysis described that the alternatives presented unquestionable key answer and distractors as shown by the green and blue arrow (see the table 4.31). This item can be unquestionable because each points that consist of key and distractors used clear words and language. In addition, the form of options did not show the key answer through grammatical clue. To know how it can be measured through seeing the explanation of the answer which it provided a description of particular problem that contained more detail response of the correct answer (as shown in the orange box).

2) Second Topic

In this topic, all of alternatives stated unquestionable form that totally consist of 15 questions of multiple-choice item. One of the example of this item mentioned in the table 4.32. It described clear key answer as the only one correct answer and three distractors but there was different point in this topic that the distractors formed plausible. This matter can make the test taker fooled which between the correct answer in C option “construct validity” and the distractor in B option “content validity” showed the same form of word as the correct answer and vice versa in A “criterion related evidence” and D “content related evidence” options.

3) Third Topic

This topic displayed table 4.33 as one of the example of item which has unquestionable key answer and distractors. Both of them stated clear language and have the same level of difficulty. In this matter, the sentence context of key answer and all of the distractors formed easily understood moreover each options

beginning with the same word “consist of”. So, it can simplify the test-taker in deciding the right choice.

c. Grammatical Error

Based on the result of this study, there are 46 questions that have consistency in grammatical and 4 questions did not have the consistency of grammar that particularly in term of alternatives construction. In this matter, the alternatives have grammatical error in terms of verb tense that is shown in the example below:

Table 4.34 Example of Multiple Choice-Alternatives in terms of Verb Tense of Grammatical Error (Topic 1)

Field	Value
NO.	6
Question	Which statement does enhance reliable test administration?
A	There is street noise beside the building while doing the test
B	Students hear the tape inaccurately
C	The amount of desk and chair is not suitable
D	Lighting in different parts of room is equally well
Explanation	The answer is “d” because good lighting is criteria of test administration reliability
Tags	Principle of language assessment, Chapter 2, Reliability

As shown in the table 4.34 above, this item did not have the consistency of grammar which lead to the grammar error in terms of verb tenses category. It appeared in C option that stated “the amount of desk and

chair is not suitable”. In this matter, to be “is” as singular verb should be changed by to be “are” because the subject was plural. So, it considered to change the verb form become the plural verb.

Table 4.35 Example of Multiple Choice-Alternatives in terms of Verb Tense of Grammatical Error (Topic 2)

Field	Value
NO.	4
Question	This is the characteristic of standardized test , except
A	Give score to students Verb Tense
B	The score report strategies is give answer to students
C	Score interpretation is percentile
D	Level of generality is global
Explanation	The correct answer is B. Because give answers to students include in classroom test characteristic
Tags	Test characteristic

According to the table 4.35 described the grammatical error of alternatives in terms of verb tense as circled by the red one. It claimed that the form of the verb “give” does not work with “is” in this sentence of alternatives. In this matter, to include multiple verbs in one sentence is incorrect except in form of active voice. So, it considered to change the verb become “give” and omit the singular verb “is”.

Table 4.36 Example of Multiple Choice-Alternatives in terms of Verb Tense of Grammatical Error (Topic 3)

Field	Value
NO.	1
Question	There are three parts of the English as a second language placement test at San Francisco State University, except...
A	Students writes a short articles and reads a summary essay
B	Students writes a response to an article
C	Students reads an essay and analyses grammar errors (multiple choice)
D	Students reads a short articles and writes a summary essay
Explanation	The answer is A because the first part is the students reads a short articles and writes a summary essay
Tags	Week5, designing classroom language test

As displayed in the table 4.36, the item has alternatives but all of the alternatives in each point of option have grammatical error particularly in terms of verb tense as circled by the red one. In this matter, the subject was in form of plural subject “students” while the verb form was written in form of singular “is”. So, it should be changed become the plural verb by omitting the letter “s” in the end of word “writes” and become the verb form “write” tha has plural meaning.

d. Keep the alternatives independent

This study got the result that all of the alternatives in each items from 3 topics that totally contained 50 questions arranged independently. It can be shown in the table 4.31 (topic 1), 4.32 (topic 2), 4.33 (topic 3) as one of the example of item which has alternatives independently.

For more detail information, it was signed by the green arrow as the key answer while the blue arrow as the distractors. All of the alternatives do not construct in overlapping form.

B. Research Discussion

In this discussion section, this study discusses the result of analysis in constructing English test particularly in multiple-choice formatting that relates to the theories in the chapter 2. The discussion explains whether students have constructed English test in form of multiple-choice item through PeerWise according to the drafting criteria or not. In this case, there are several aspects to construct multiple-choice item based on the expert theory that relates to students' ability:

1. Students' ability in constructing instruction

For this aspect, all of the students' multiple-choice item construction do not have instruction (see table 4.1, 4.2, and 4.3). The students' items direct to the stem construction without include the instruction as the direction of item. Actually, in designing test, the test designer should include instruction to measure the validity of the test. How the test could be administered if there is no instruction because it is the guideline of test taker to answer the question. If there is no instruction, it will appear misunderstanding between students. This statement is supported by the previous study of Michael C. Rodriguez that has the same explanation regarding to the construction of instruction⁴³, but it has different result with the researcher study. It stated that the relation between instruction and testing is very important and both are mutually sustainable. It concluded that stating the instruction in test item as a set of guideline can influence to the use of test item and also the score of test taker. So, the instruction should be clear stated when making the items test. Likewise, instruction is a guideline to achieve the goal

⁴³Michael C. Rodriguez. "Item and Test Design considering instructional sensitivity".
National Council, 2017

of test which is in constructing the good test should have clear instruction and directly to the learning outcome⁴⁴. It means that the instruction directly asks what the students need to do and lead to the outcome of the test which can make test taker easy to understand the purpose of the test direction.

2. Students' ability in constructing stem of multiple-choice test

In the way test designer, the finding indicated that some students still have difficulty in constructing the stem particularly in terms of grammatical. While, to construct the stem of test item should include some criteria based on the standard criteria that related the theory of Norman E Gronlund⁴⁵: (1) present a single problem (2) have positive formatting (3) grammatically consistent. Based on the criteria of good stem, only criterion number 1 was found in all of multiple-choice items which the stem states a single idea in each item (see table 4.5, 4.6, and 4.7).

Based on the second criterion, the items of students' work have 42 items of multiple-choice that in positive stem while in different formatting as described in the table 4.8. In that explanation states one of the postive stem formatting that there were 8 questions from 3 topics in completion format. In this case, the stem construction uses tobe in the end of stem as describe in the table 4.13. It is supported by J. B. Heaton theory that there are some formats of stem⁴⁶, one of them is an incomplete statement or completion format. When constructing the stem in this form, it must pay attention to the use of tobe at the end of stem because it includes a weak stem which students can easily guess the correct answer. While the opposite, there were 9 questions in negative form (see the table 4.17, 4.18, and 4.19).

⁴⁴ Harold S. Madsen, *Techniques in Testing* (New York:1983), 21.

⁴⁵ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 39-43.

⁴⁶ J. B. Heaton, *Writing English Language Tests* (London and New York: Longman Group UK, 1988), 29.

Actually, the negative form uses the word “not” or “except” as the sign but this should be avoided in constructing the good stem⁴⁷. If this negative stem can not be avoided, the negative word “not” or “except” should be highlighted by underlining or capitalization.

The other finding based on third criterion is grammatically consistent which is difficult for students in constructing the stem of multiple-choice item using good grammar. The result of this study found 14 questions have grammar error. The error that is often made by students are in terms of verb of tense and article. It can be happened because the students often do not pay attention to grammar when constructing the items⁴⁸. From that, it will lead to choose the correct answer easily. So, to reduce the errors in grammar, this should avoid using the article at the end of stem and pay attention to the using of verb tense correctly particularly in terms of subject verb agreement. The explanation of this main finding has similarities with previous study by Agnes Siwi in terms of grammatical structure⁴⁹. But, it has different way and subject where this study collected the data through using questionnaire and the subject is the English teacher. The result of analysis stated this research designed a diagnostic test to measure the ability of teachers in grammatical structures which it designed to help teachers review and improve their competence in grammatical structures.

3. Students’ ability in constructing alternatives of multiple-choice test

In designing the alternatives or the options should consist of two parts included the correct answer and

⁴⁷ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 43.

⁴⁸ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 45.

⁴⁹ Agnes Siwi Purwaning Tyas. “A Diagnostic Test On Grammatical Structures For English Teachers”. *Indonesian journal of ELS*, 2015

distractors. The result analysis of this study that related to students' ability in constructing the option have both of those parts. This is identical to multiple-choice test which there are some criteria to construct the alternatives of multiple-choice test item. These are the criteria of alternatives construction⁵⁰: (1) it contains one correct answer and three distractors (2) the key answer and distractors should be unquestionable (3) grammatically consistent (4) keep the alternatives independent (5) avoid using the alternatives "all of the above" and "none of the above".

Based on the criteria above, all of the multiple-choice items in three topics matched the criteria of number 1 and 4 in terms of constructing the alternatives. Likewise, the criteria of number 2 matched with the result of analysis that the key answer and distractors should be unquestionable. It means that both of points stated clearly and arranged well construction. While, it has different condition and form in each topic. As mentioned that in constructing the alternatives forbid to mention the grammatical clues in deciding the correct answer and have the same level of difficulty in each options. The multiple choice item should have the same level in constructing the sentence context of each options to make students easy in reading the item⁵¹. Moreover, the key answer and distractors have the same form of word. In addition, when the correct answer is not the same form as the distractors, the test-taker might answer the multiple-choice question correctly for wrong reason⁵². In this matter, the test designer should pay attention to both of parts particularly in constructing the distractors because it will decide how the test taker answer the test. This explanation has the same result as the other analysis of

50 Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 44-52.

51 Harold S. Madsen, *Techniques in Testing* (New York:1983), 20.

52 Harold S. Madsen, *Techniques in Testing* (New York:1983), 19.

study by Artur Mkrtchyan. It mentioned that the distractors have an important role for multiple-choice testing process⁵³. If the test item has good quality of distractors, it will provide a more objective and credible picture of the knowledge of the examinees involved where it can be seen from the results of students test. So, to make the test item has high quality, it also regards to the distractor construction.

On the other hands, the problem that often appears when constructing options is grammar error. This found in all topics that totally 4 items of multiple-choice. Those mentioned in the first topic that has 2 items error while the second and third topic have 1 item error in each topic. The grammar error that always arises is dealing with the verb tense. It can happen because of the lack of accuracy in constructing the alternatives particularly in terms of subject verb agreement so it can appear the grammar error⁵⁴. This result can make the alternatives having inconsistently grammar with the stem.

For the last criteria includes to avoid the using of alternative “all of the above” and “none of the above”. According to this criteria, the finding of this study states that all of multiple-choice items do not use this aspect because it can clearly show the correct answer and will automatically leave the other options⁵⁵. It can also identify the wrong answer which is among the options include one wrong answer. So that, it can simplify the test-taker in determining the correct answer. However, when using “all of the above” and “none of the above” as the correct answer choice, it only needs to eliminate one answer choice as implausible.

⁵³ Artur Mkrtchyan. “Distractor Quality Analyze In Multiple Choice Questions Based On Information Retrieval Model”. *State Engineering University of Armenia*.

⁵⁴ Norman E. Gronlund, *Constructing Achievement Tests* (United State of America: 2002), 45.

⁵⁵ Harold S. Madsen, *Techniques in Testing* (New York:1983), 51.

CHAPTER V

CONCLUSION AND SUGGESTION

It mentions conclusion from students' ability in constructing English test through PeerWise. This matter provides recommendation that purposed to teachers, students, as well as researchers in the future as explained below:

A. Conclusion

In this section, the study states the conclusion of students' ability in constructing English test through PeerWise. Actually, in designing the good English test construction particularly in the form of multiple choice test should include 3 aspects consist of instruction, stem and alternatives. But, the research finding and discussion of these 50 multiple choice items just met 2 aspects in constructing good multiple-choice test, namely in terms of stem and alternatives. In terms of stem, all of items that totally 50 questions present a single problem that focus on the idea of knowledge and 41 items have positive form which include in different formatting. It states W-H Question formatting that indicated with question word and question mark, completion form that highlighted by the missing word or the periods in the end of stem, for others are improper formatting and instruction formatting. For the opposite, 9 stems of multiple-choice questions state in negative form that identified with the word "except". However, students have problem in constructing the stem in term of grammatical error particularly about verb tense, article, singular/plural noun, and preposition. For the second aspect is alternatives construction which students' multiple-choice items provide one key answer and three distractors that state unquestionable form. It means that both of points have clear language, understanding the test-taker, and reasonable. The opposite, the result of study mentions the grammar error of students work in constructing the alternatives particularly in verb tense.

In conclusion, the students' ability in constructing English test particularly in the form of multiple choice test already met the standard criteria of constructing good multiple choice item. While, the test items have lack in terms of instruction and also the

grammatical error that often arises that related to the subject verb agreement. So, those criteria should more give attention to reduce the students' error in constructing good English test based on the standard criteria.

B. Suggestion

According to the analysis result of the construction of 50 multiple-choice items from 3 topics through PeerWise, the significant suggestions from the researcher are as the following:

1. For English department students

The finding showed that English department students still facing errors when constructing the multiple-choice item in PeerWise particularly in terms of instruction, stem and alternatives. The suggestion for students is to pay attention to the validated criteria in constructing the multiple-choice item before publish it. So, it can make easy for students to know the mistake of their items construction.

2. For teacher

Based on the result, the teacher should give the students further training before constructing the multiple-choice item in PeerWise. In order to help the students in constructing good multiple-choice item based on the validated criteria. So, they will have deep knowledge about constructing test item in form of multiple-choice.

3. For future researcher

The researcher gives suggestion to the future researcher that can examine the same domain as the researcher did which in terms of multiple-choice in PeerWise but in a different aspect. That is more lead to the students' assessment or comments on multiple-choice items that have been made by their classmates through PeerWise.

REFERENCES

- Alfany, Fadhila. "The Correlation Between Students' Participation and Their Examinataion Score in PeerWise Online Learnit at English Education Department UIN Sunan Ampel Surabaya". Surabaya: UINSA, 2016.
- Bachman, Lyle F. *Fundamental Considerations in Language Testing*. New York: Oxford University Press, 1990.
- Brown, Douglas. *Language Assessment: Principles and Classroom Practices*. San Fransisco: Pearson Education, 2004.
- Creswell, John W. *Research Design*. USA: SAGE Publications.Inc, 2009.
- Choi, Yeon Hee et.al. 2008. "Current Trends and Issues in English Language Education in Asia". *The Journal of Asia TEFL*.
- Denny, Paul et.al. *Evaluation of PeerWise as An Educational Tool for Bioengineers*. New Zealand: American Society for Engineering Education, 2010.
- Dulger, Mehmet et.al. 2017. "Assessing the Validity of Multiple-Choice Questions in Measuring Fourth Graders' Ability to Interpret Graphs Abut Motion and Temperature". *Nevada*.
- Elizabeth, Alison. "*Student-generated content: Investigating Student Use of PeerWise*". United Kingdom: The University of Edinburgh, 2016.
- Gronlund, Norman E. *Constructing Achievement Tests*. United States of America: Library Congress Cataloging in Publication Data, 2002.
- Haladyna, Thomas M. *Developing and Validating Multiple-Choice Tests Item*. London: Lawrence Erlbaum Associates, 2004.

- J. B. Heaton. *Writing English Language Tests*. New York: Longman Group UK, 1988.
- Madsen, Harold. *Teaching In Testing*. New York: Oxford University Press, 1983.
- Martiarini, Elyza. 2017. "Constructing Multiple-Choice and Cloze Task On Student's Reading Comprehension". *University of Indraprasta PGRI*.
- McClellan, Stephen. 2015. "Implementing PeerWise to Engage Students in Collaborative Learning".
- Michael Harris and Paul McCann. *Assessment*. Macmillan Publishers Ltd, 1994.
- Mkrtchyan, Artur. "Distractor Quality Analyze in Multiple-Choice Questions Based on Information Retrieval Model". *State Engineering University of Armenia*.
- Osterlind, Steven J. *Constructing Test Items*. London: Kluwer Academic Publishers, 1989.
- Paul, et al. 2010. "Evaluation of PeerWise as an Educational Tool for Bioengineers". *American Society for Engineering Education*.
- Putu, Ni et.al. 2014. "Analisis Butir Soal Pilihan Ganda Mata Pelajaran Bahasa Indonesia SMK Negeri 3 Singaraja". *Universitas Pendidikan Ganesha*. Vol.2 No.1, 2014.
- Rodriguez, Michael. 2017. "Item and Test Design Considering Instructional Sensitivity". *National Council of Measurement in Education*.
- Siwi, Agnes. 2015. "A Diagnostic Test on Grammatical Structures for English Teachers". *Indonesian Journal of English Language Studies*. Vol.1 No.1, March 2015.