



**REPUBLIK INDONESIA**  
**KEMENTERIAN HUKUM DAN HAK ASASI MANUSIA**

**SURAT PENCATATAN CIPTAAN**

Menteri Hukum dan Hak Asasi Manusia Republik Indonesia, berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta yaitu Undang-Undang tentang perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra (tidak melindungi hak kekayaan intelektual lainnya), dengan ini menerangkan bahwa hal-hal tersebut di bawah ini telah tercatat dalam Daftar Umum Ciptaan:

- I. Nomor dan tanggal permohonan : EC00201705885, 27 November 2017
- II. Pencipta  
Nama : **Slamet Asari, S.Pd., M.Pd.**  
Alamat : Perum Graha Indah Blok D/74 RT 002 RW 009  
Tambakrigadung, Kec. Tikung , Lamongan, JAWA TIMUR,  
62281  
Kewarganegaraan : Indonesia
- III. Pemegang Hak Cipta  
Nama : **Universitas Muhammadiyah Gresik**  
Alamat : Jl. Sumatera No. 101 , Gresik, JAWA TIMUR, 61121  
Kewarganegaraan : Indonesia
- IV. Jenis Ciptaan : Buku
- V. Judul Ciptaan : **English Language Testing and Assessment**
- VI. Tanggal dan tempat diumumkan : 27 November 2017, di Gresik  
untuk pertama kali di wilayah  
Indonesia atau di luar wilayah  
Indonesia
- VII. Jangka waktu perlindungan : Berlaku selama 50 (lima puluh) tahun sejak Ciptaan tersebut  
pertama kali dilakukan Pengumuman.
- VIII. Nomor pencatatan : 05168

Pencatatan Ciptaan atau produk Hak Terkait dalam Daftar Umum Ciptaan bukan merupakan pengesahan atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang dicatat. Menteri tidak bertanggung jawab atas isi, arti, maksud, atau bentuk dari Ciptaan atau produk Hak Terkait yang terdaftar. (Pasal 72 dan Penjelasan Pasal 72 Undang-undang Nomor 28 Tahun 2014 Tentang Hak Cipta)

a.n. MENTERI HUKUM DAN HAK ASASI MANUSIA  
REPUBLIK INDONESIA  
DIREKTUR JENDERAL KEKAYAAN INTELEKTUAL  
u.b.  
DIREKTUR HAK CIPTA DAN DESAIN INDUSTRI

Dr. Dra. Erni Widhyastari, Apt., M.Si.  
NIP. 196003181991032001

## ABOUT THE AUTHOR



**Slamet Asari** was born on August 20, 1970 in Pasuruan, East Java Indonesia. He is presently studying English Language Teaching for his Doctoral Degree at State University of Surabaya funded by Directorate Higher Education. He accomplished his master degree (S-2) in English Education particularly on the area of English Language Teaching in 2003 at State University of Malang. Long before achieving his master degree, he has achieved his strata 1 (S-1) in 1993 at English Department University of Muhammadiyah Malang.

His long experience in teaching English at University of Muhammadiyah Gresik brought him into several national or international academic events. Participating actively as speaker for International Conference on Lesson Study in several universities and cities in Indonesia is among things he has concerned with. He also had been given an opportunity to visit some countries to improve his teaching expertise.

Currently, he teaches some courses such as Language Testing, Authentic or Alternative Assessment, Academic Writing, Microteaching, and English for Specific Purposes (ESP) for undergraduate students at English Department of University of Muhammadiyah Gresik.

This book is his fourth book he has written for students of undergraduate level. Books such as 'Reading for Meaningful Passages', 'English for Law 1 and 2', and 'English for Psychology' are his previous publications.



English Language Testing And Assessment

# English Language Testing and Assessment



Slamet Asari

# **English Language Testing and Assessment**

**Slamet Asari**



**Universitas Muhammadiyah Gresik**

*Perpustakaan Nasional: katalog dalam terbitan (KDT)*

**Penulis:**

**Slamet Asari**

**“ENGLISH LANGUAGE TESTING  
AND ASSESSMENT”**

Gresik: UMG Press, 2017

15,5x23 cm viii, 85 hlm

ISBN: 978-602-50707-2-3

**Editor:**

Dr. Khoirul Anwar, M.Pd.

**Penyunting:**

Ainur Rofiq

**Desain sampul dan Tata letak**

Moch. Indah Hastomo Nugroho, S.Kom.

**Penerbit:**

UMG Press

**Redaksi:**

Jln. Sumatera 101 GKB

Gresik 61121

Telp +6231 3951414

Fax +6231 3952585

Email: [press@umg.ac.id](mailto:press@umg.ac.id)

Anggota IKAPI No. 189 dan APPTI No. 002.021

Cetakan pertama, Oktober 2017

*Hak cipta dilindungi undang-undang*

*Dilarang memperbanyak karya tulis ini dalam bentuk dan  
dengan cara apapun tanpa ijin tertulis dari penerbit*

## KATA PENGANTAR

Syukur Alhamdulillah kepada Allah S.W.T atas ridho dan petunjukNYA sehingga penulis bisa menyelesaikan Buku Ajar dengan judul *English Language Testing and Assessment* yang diperuntukan dalam pembelajaran mahasiswa Pendidikan bahasa Inggris pada tingkat perguruan tinggi.

Buku ajar initerdiri dari 5 bab dengan memfokuskan pada aspek-aspek teknis dan konseptual pengembangan dan penyusunan tes untuk pembelajaran bahasa Inggris. Cakupan materi dalam penyajian lima bab tersebut membekali kompetensi penyusunan soal tes bahasa Inggris mahasiswa. Materi seperti konsep pendekatan dalam Language Testing, tipe dan fungsi dari tes bahasa, cirri-ciri tes yang baik serta materi lain yang berhubungan dengan penyusunan dan pengembangan tes disajikan secara komunikatif sebagai sumber belajar.

Buku ajar ini dilengkapi dengan diskripsi materi setiap topik, relevansi dalam menunjang mata kuliah Assessment in English Language Testing serta sub capaian pembelajaran yang akan dicapai. Pada setiap bagian akhir dari setiap bab, rangkuman dan evaluasi juga dicantumkan dalam rangka umpan balik.

Melalui buku ajar ini diharapkan mahasiswa bias memahami secara menyeluruh tentang hal-hal yang berkenaan dengan penyusunan dan pengembangan tes bahasa Inggris yang baik dan berterima. Akhirnya, penulis menyampaikan terimakasih yang sebesar-besarnya kepada pihak-pihak yang membantu terselesainya buku ajar ini. Penulis menyadari jika buku ini jauh dari kata sempurna. Oleh karenanya, saran dan masukan yang konstruktif akan sangat berarti bagi penulis.



# INTRODUCTION

## A. Course Description

This course covers an explanation of various issues about the preparation of the major language tests on the types and functions of tests in the language of teaching, the characteristics of good language tests, the selection and constructing of good language tests, the administering tests, the scoring tests, the interpreting tests results. Basic statistical calculations are also presented in order to analyze test items and test interpretations.

## B. Courses' Pre-Requisite

This course has no pre-requisite course to be taken.

## C. Lesson Plan

**Name of the Course** : Assessment in ELT

### Learning outcome for the course

1. M1 Understanding the concept of testing in English language
2. M2 Knowing kinds of test types and their test format
3. M3 Understanding characteristics of a good test
4. M4 Analysing English test items
5. M5 Constructing English test items

Mg Ke -	Sub-CP-MK	Indikator	Kriteria & Bentuk Penilaian	Metode Pembelajaran	Materi Pembelajaran
1 - 2	Mahasiswa mampu memahami maksud dan konsep pendekatan dalam Language Testing	Ketepatan dalam memahami konsep dan pendekatan tes dalam bahasa	<b>Kriteria</b>  Ketepatan dalam memahami konsep dan pendekatan tes dalam bahasa  <b>Penilaian</b>	Tanya jawab dan ceramah	Intro to lang. test & Approach and Method.  Role and goal of Language testing  1. Madsen, S. Harold. 1983. Techniques in Testing. New York: Oxford University Press

			Lisan (tanya jawab)		<ol style="list-style-type: none"> <li>2. Brown, Dean.James.1996 . Testing in Language Programs. New Jersey:Prentice Hall Regents</li> <li>3. Heaton, J.B. 1988.Writing English Language Tests. Longman Group: UK and USA</li> <li>4. Djiwandono, Soenardi.2008. <b>Tes Bahasa</b>. PT. Indeks: Indonesia</li> </ol>
3	Mahasiswa mampu memaparkan tipe dan fungsi dari tes bahasa	Ketepatan memaparkan tipe dan fungsi tes bahasa	<p><b>Kriteria</b></p> <p>Ketepatan memaparkan tipe dan fungsi tes bahasa</p> <p><b>Penilaian</b></p> <p>Lisan (Tanya Jawab)</p>	Tanya jawab dan ceramah	<p>Types and Function of Test in English Language</p> <ol style="list-style-type: none"> <li>1. Heaton, J.B. 1988.Writing English Language Tests. Longman Group: UK and USA</li> <li>2. Madsen. S. Harold. 1983. Techniques in Testing. New York: Oxford University Press</li> <li>3. Djiwandono, Soenardi.2008. <b>Tes Bahasa</b>. PT. Indeks: Indonesia</li> </ol>
4	Mahasiswa mampu memahami format tes bahasa	Ketepatan dalam menentukan format tes dalam bahasa	<p><b>Kriteria</b></p> <p>Ketepatan dalam menentukan format tes dalam bahasa</p>	Ceramah, tanya jawab	<p>Forms of test : objective and subjective test. Jenis dan macamnya</p> <ol style="list-style-type: none"> <li>1. Heaton, J.B. 1988.Writing English Language Tests. Longman Group:</li> </ol>



			<p><b>Penilaian</b></p> <p>Lisan (menyebutkan format tes bahasa)</p>		<p>UK and USA</p> <p>2. Madsen. S. Harold. 1983. Techniques in Testing. New York: Oxford University Press</p> <p>3. Djiwandono, Soenardi.2008. <b>Tes Bahasa. PT. Indeks: Indonesia</b></p>
5	Mahasiswa mampu memaparkan ciri-ciri tes bahasa	Ketepatan dalam mendiskripsikan ciri-ciri tes bahasa yang baik	<p><b>Kriteria</b></p> <p>Ketepatan dalam mendiskripsikan ciri-ciri tes bahasa yang baik</p> <p><b>Penilaian</b></p> <p>Lisan tentang ciri-ciri tes bahasa</p>	Diskusi dan presentasi	<p>The characteristics of test in Language</p> <p>Validity, Reliability, Practicality, Simplicity</p> <p>1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA</p> <p>2. Madsen. S. Harold. 1983. Techniques in Testing. New York: Oxford University Press</p> <p>3. Djiwandono, Soenardi.2008. <b>Tes Bahasa. PT. Indeks: Indonesia</b></p>
6 - 7	Mahasiswa mampu melakukan pengukuran butir tes mengacu pada validitas dan reliabilitas tes	Ketepatan dan kesesuaian dalam melakukan pengukuran tes	<p><b>Kriteria</b></p> <p>Ketepatan dan kesesuaian dalam melakukan pengukuran tes</p> <p><b>Penilaian</b></p>	<p>Presentasi, diskusi</p> <p><b>Tugas 1:</b></p> <p>Menganalisa butir tes tingkat SD, SMP, SMA dan SMK</p>	<p>Validity and Reliability of a test macam dan jenisnya, Item analysis</p> <p>1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA</p> <p>2. Madsen. S. Harold. 1983. Techniques in</p>

			-		Testing. New York: Oxford University Press 3. Djiwandono, Soenardi.2008. <b>Tes Bahasa. PT.</b> Indeks: Indonesia
<b>8</b>	<b>Uji tulis (analisa tes item) Unjuk kerja</b>				
<b>9 - 10</b>	Mahasiswa mampu menyusun tes yang baik : soal dan opsinya	Ketepatan dan kesesuaian dalam membuat butir soal dan pilihannya	<b>Kriteria</b>  Ketepatan dan kesesuaian penyusunan tes  <b>Penilaian</b>  -	Unjuk kerja dan diskusi  <b>Tugas 2:</b>  Menyusun butir test	<ol style="list-style-type: none"> <li>1. Langkah penyusunan tes</li> <li>2. Ciri-ciri tes yang baik</li> <li>3. Kriteria stem dan opsi yang baik</li> </ol> <ol style="list-style-type: none"> <li>1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA</li> <li>2. Madsen. S. Harold. 1983. Techniques in Testing. New York: Oxford University Press</li> <li>3. Djiwandono, Soenardi.2008. <b>Tes Bahasa. PT.</b> Indeks: Indonesia</li> </ol>
<b>11</b>	Mahasiswa mampu mengembankan dan menyusun tes Reading	Kesesuaian dan ketepatan dalam menyusun tes Reading	<b>Kriteria</b>  Ketepatan dan kesesuaian penyusunan tes Reading  <b>Penilaian</b>	Unjukkerja, diskusi/presentasi  <b>Tugas 3:</b>  Menyusun butir tes Reading	<b>Penyusunan tes Reading</b>  <ol style="list-style-type: none"> <li>1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA</li> <li>2. Djiwandono,</li> </ol>

					Soenardi.2 008. <b>Tes Bahasa.</b> PT. Imdeks: Indonesia
1 2	Mahasiswa mampu mengemban gkan dan menyusun tes Listening	Kesesuaian dan ketepatan dalam menyusun tes Listening	<b>Kriteria</b>  Ketepatan dan kesesuaian penyusunan tes Listening  <b>Penilaian</b>  -	Unjukkerja dan diskusi/prese ntasi  <b>Tugas 4:</b>  Menyusun butir tes Listening	<b>Penyusunan tes Listening</b>  1. Heaton, J.B. 1988. Writng English Language Tests. Longman Group: UK and USA  2. Djiwandon o, Soenardi.2 008. <b>Tes Bahasa.</b> PT. Imdeks: Indonesia
1 3	Mahasiswa mampu mengemban gkan dan menyusun tes Vocabulary	Kesesuaian dan ketepatan dalam menyusun tes Vocabulary	<b>Kriteria</b>  Ketepatan dan kesesuaian penyusunan tes Vocabulary  <b>Penilaian</b>  -	Unjukkerja, diskusi/prese ntasi  <b>Tugas 5:</b>  Menyusun butir tes Vocabulary	<b>Penyusunan tes Vocabulary</b>  1. Heaton, J.B. 1988. Writng English Language Tests. Longman Group: UK and USA  2. Djiwandon o, Soenardi.2 008. <b>Tes Bahasa.</b> PT. Imdeks: Indonesia
1 4	Mahasiswa mampu mengemban gkan dan	Kesesuaian dan ketepatan dalam menyusun	<b>Kriteria</b>  Ketepatan dan	Unjukkerja dan diskusi/prese	<b>Penyusunan tes Grammar dan Writing</b>

	menyusun tes Grammar dan Writing	tes Grammar dan Writing	kesesuaian penyusunan tes Writing  <b>Penilaian</b> -	ntasi  <b>Tugas 6:</b>  Menyusun tes Grammar dan Writing	1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA 2. Djiwandono, Soenardi. 2008. <b>Tes Bahasa.</b> PT. Imdeks: Indonesia
15	Mahasiswa mampu mengembangkan dan menyusun tes Speaking	Kesesuaian dan ketepatan dalam menyusun tes Speaking	<b>Kriteria</b>  Ketepatan dan kesesuaian penyusunan tes Speaking  <b>Penilaian</b> -	Unjukkerja dan diskusi/presentasi  <b>Tugas 7:</b>  Menyusun tes Speaking dan Rubrik penilaian	<b>Penyusunan butir tes Speaking</b>  1. Heaton, J.B. 1988. Writing English Language Tests. Longman Group: UK and USA 2. Djiwandono, Soenardi. 2008. <b>Tes Bahasa.</b> PT. Imdeks: Indonesia
16	<b>Uji Tulis (Menyusun tes Bahasa Inggris) Project Based Learning</b>				

## **6. Learning Outcome / Competence**

S 9S	Showing responsibility attitude to their works
S 20	Behaving as stated on ethic code of teacher profession
KU 2	Showing independent, qualified, and measurable work
KK 9	Developing indicator and instrument for evaluation
KK 20	Developing instrument for evaluation, evaluating process and learning result
KK 23	Conducting process and learning evaluation
PP 14	Having knowledge on conducting and developing test.

## **7. Evaluation**

The evaluation is carried out through students' active participants in classroom teaching and students' individual point or score referring to task 1 which is analysing English test item they obtain from schools or English teachers. Task 2 is also provided to get consistent individual score. The task is requiring students to construct a good English test items following standardized test construction. Rubric is provided as reference to measure students' performances within the two Tasks.

## **Table of Contents**

CHAPTER I.....	9
Concept and Approach of English Language Testing.....	9
1.1    Concept of Language Testing.....	9
1.2    Approach in English Language Testing.....	14
1.3    Summary .....	24
CHAPTER II.....	26
Types and Function of Language Test.....	26
2.1    Types and Function of Language Test.....	26
2.2    Summary .....	41
CHAPTER III .....	43
Forms of Language Testing.....	43
3.1    Forms of English Language Testing.....	43
3.2    Summary .....	56
CHAPTER IV .....	57
Developing and Improving Test Items .....	57
4.1    Characteristics of Good Test .....	58
4.2    Test Items Guidelines .....	69
4.3    Summary .....	74
CHAPTER V.....	76
Analysing Test Items.....	76
5.1    Item Analysis.....	76
5.2    Summary .....	81
Bibliography.....	82

## CHAPTER I

### Concept and Approach of English Language Testing

#### A. Brief Description

This chapter discusses about concept and Different types of approaches used for constructing English test items. These concepts and approaches are used as an underlying theory to develop English language test. Terminologies such as evaluation, assessment, test are descriptively presented to determine their differences. Meanwhile, approaches are provided to base reference for constructing test items.

#### B. Relevance

This topic has significant importance and contribution to the make of English test items. It underlies basic reference to how a set of English test items should be composed and constructed.

#### C. Learning Outcome

Students are able to understand concept and approaches to English language testing or assessment.

### 1.1 Concept of Language Testing

Testing is a crucial part of every teaching and learning experienced by teacher. Every teacher performs test everytime he or she has completed his or her teaching of a particular topic to see how effective his or her teaching is. Teachers can determine the success of the teaching by performing test to see whether the teaching technique, strategy, method or approach they have set runs effectively or stucks within a routine obligation where they most importantly have committed their duty only by teaching without necessarily reflects or changes

their teaching modes. A test functions to fulfill teachers' need in obtaining information toward their teaching performances.

This teaching performance is not the only goal for a teacher to carry out a test. It also has main purpose to assess students skill or knowledge toward what they have achieved at particular period of time. A test result provides information a teacher can collect to determine students' competence concerning certain topic they have taught during classroom teaching or even at the end of teaching program. This information can be used by teacher to set another 'policy' for better treatment toward students' competence.

According to Phillips (1979:1-2) "a test is meant as a tool used to get data concerning with group or individual characteristics." This definition is very relevance to the essence of testing in the teaching and learning process since this process of teaching and learning requires a picture of both teachers and students involved in each of their domain. Teachers need information both their teaching performances while students want to know their competences. As what Lutan (2000:21) claimed that "a test is an instrument used to obtain information about someone or an object."

The above general concept of testing is also intended to language testing where the target of the essence of language testing is the ability of students to communicate both in written and oral forms. Assessing or testing in language teaching is an application of various ways to obtain information about the extent to which the learners learn language or the achievement of language competence of learners. Conventionally, language ability itself consists of four fundamental skills namely Listening skill, Reading skill, Speaking skill and writing skill.



These four skills principally are classified into two language skills such as passive-receptive language skill and active-receptive skill. Listening skill and reading skill befall into passive-receptive where students' role tend to passively and indirectly learn language. The students only attempt to understand what is exposed by language speakers or writer rhetorical message. Meanwhile, speaking and writing skill belong to active-receptive language skill where the students or language learners initiate very directly and actively communicate and use their ability to speak and write to express their thought. These two domains of language skills then become the main target to be assessed.

In addition to some points above, assessing students' skill in language learning is basically not the only domain the test or assessment can do. Language component such as grammar or language structure in the forms of grammatical patterns, vocabulary, phonology, pronunciation can be another target of assessing students' language competence although these language components are not necessarily measured when students learn general language in this case English language. The measurement as a matter of fact can go to those who learn language in to more advanced level such as those who are interested in studying English language at university level.

As has been described, assessment is activity implemented to obtain information about the extent to which the learners learn or the achievement of competence of learners. Assessment of learning outcomes is basically a question about how teachers can find out the learning outcomes that have been done. The teacher must know the extent to which the learner has understood the material already taught or the extent to which the objectives / competencies of the managed learning activity can be achieved. The level of achievement of the

competencies or instructional goals of the learning activities that have been implemented can be expressed with grade.

There are four kinds of terms related to the concept of assessment. These terms are often used to determine the success of learning from learners. They are (1) test, (2) assessment and (3) evaluation. Among the four terms the notion is however still often mixed since the four terms fundamentally have different understanding.

Actually the process of measurement, assessment, evaluation and testing is an activity or process that is hierarchical. This means that activities are done sequentially and tiered starting from the process of measurement and then assessment and evaluation last. While the testing process is part of the measurement followed by assessment activities. Definition of those terms can be defined as follows:

- A **test** is used to examine someone's knowledge of something to determine what he or she knows or has learned. Testing measures the level of skill or knowledge that has been reached.
- **Assessment** is the process of documenting knowledge, skills, attitudes and beliefs, usually in measurable terms. The goal of assessment is to make improvements, as opposed to simply being judged. In an educational context, assessment is the process of describing, collecting, recording, scoring, and interpreting information about learning.
- **Evaluation** is the process of making judgments based on criteria and evidence.

(taken from <http://tutorials.istudy.psu.edu/testing/testing2.html>)

English language testing in this case is an integral part of an evaluation process on language teaching and learning. It deals with any issues on English test construction, characteristics of good English test, administering test, scoring test kinds and function of tests in language teaching. Principally, it goes for a process beginning from preparing test and interpreting test score for language teaching domain.

The language test is intended as a tool or procedure used in conducting or evaluating language ability with measuring the level of the language level. Such measurement aims to determine the level of proficiency in language competence. Regarding on the purpose that wants to be achieved, language tests can be applied to measure the level of language in particular covering two aspects of language skills namely passive-receptive and active receptive language skill encompassing four expertise; listening, reading, speaking and writing completed with language components such as vocabulary, grammar or structure, pronunciation etc.

Thus, language testing is the practice and study of evaluating the proficiency of an individual in using a particular language in this case English language effectively. In a narrower sense it attempts to measure one's capability on his or her language proficiency through test instruments which are administered following rules of constructing a good test. Its role in teaching is very significant since theory of teaching involves constructing and planning of teaching starting from setting teaching objectives, implementing teaching process, and finally making evaluation. Here, in making evaluation teacher needs to understand very well of how to administer a test in order to get valid information regarding students' competences and teachers' teaching performances. The information is obtained through conducting a test.

## **1.2 Approach in English Language Testing**

Currently, language test is progressing from time to time. The emergence and changes emphasizing on language test is intended to update the tests in accordance with the views, approaches, and focus of language learning. The updates made are often based on existing weaknesses. The existence of previous views, approaches or focus may become basic element and reflection toward the incoming progress of updated approaches in language testing. The weaknesses, of course are very beneficial to become reference for better improvement toward any issues on language testing. We cannot easily neglect them since they are the pioneer of the existence of the non exist.

Approach in language testing is very significant to determine what kind of language test format and other aspects would be appropriate for Particular test setting and level. Since test belongs to evaluation then it is very relevance to process of teaching learning. Teaching and learning process deals with preparation, execution and evaluation. The last part plays important role to determine the success of teaching and learning process. Thus, a test cannot be separated from evaluation where evaluation itself is a part of teaching process. Djiwandono, (2008, 18) states that the approach to language determines the approach of language learning, and the language learning approach determines the approach in administering the test.

The tendency and practice of language tests followed the flow of teaching method. There are some changes in the approach of understanding what to measure and how to measure. There are pre-scientific and scientific era signifies or marks the dominance of a particular approach to language testing. The era of 'grammar-translation' dominates the practice of testing in the pre-scientific era whereas discrete-point, pragmatic-integrative, communicative, and performance-based

dominates testing practice in the scientific era. These are the approaches currently implemented among language teachers to construct test following the teaching approach they set with some strengths and weaknesses they carry.

The approach used may affect on the teaching process reflecting on the evaluation done teacher. There are commonly five approaches employed by test makers to develop language testing. They are (1) Traditional or Grammar Translation Approach, (2) Discrete Approach, (3) Integrated Approach, (4) Pragmatic Approach, and (5) Communicative Approach.

#### ***a. Grammar Translation or Traditional Approach***

The traditional approach is the term used to refer to planning and execution regarding teaching and learning process carried out traditionally. The terms traditional does not mean 'left behind' but it is rather old and conventional. This view comes up with a reason that at that time the teaching process was not yet developed and executed based on comprehensive study on language. This approach views that Language tests tend to adopt the principle that language tests focus on grammar and translation tests. The background is the mainstream influence of the teaching of language known as the grammar translation method. This method, as argued by Richards and Rogers (1988: 3-4), has the principles of teaching such as:

1. Learning a foreign language is to learn a language in order to read its literature;
2. Reading and writing are the main focus of teaching, accuracy in translation is strongly emphasized, and
3. Grammar should be taught deductively, ie moving from the rules and then to the illustration examples

In developing the test, this approach does not relate to particular theory of language learning. Any language teacher who teaches language can carry out a test. There is no clear reference to be referred to and no special skill required by teachers to administer a test. This means that those who can teach English may be possible to carry a test and construct it. Consequently, the evaluation may go very subjectively, the choice of the test material is also subjective and even the choice of language skill is not quite clear (no clear standard what component or skills being tested). The test subject is called English test although the test itself goes for translation, grammar test ect.

In the traditional approach of language learning is held just for certain limited needs such as; Communicating verbally and limitedly and focusing on grammar. Many of them only emphasize the ability to translate from one language into another. There are some test format commonly developed by teachers who follow this approach such as translating text and answering open-ended questions. Example of this approach can be seen below:

*Translate the following text into Indonesian language!*

1. Two little boys are playing football by the river.  
-----
2. There was a car accident here last night. The police is identifying the crime scene now.  
-----

*Compose two free sentences with correct Grammar!*

1. -----
2. -----

Most of the time the test emphasizes on grammatical aspect since the grammar translation method is very popular and applicable at that time. Thus, this tradition approach is also labelled as grammar translation approach. In carrying out this test format, some rules concerning with test administration are less concerned since they are not yet available at that time. These rules include the way how the test should be done and how the test is scored. However, currently this approach is still employed although rules, technique how to construct or administer the test is already available.

### ***b. Discrete Approach***

The discrete approach in language tests is inspired by structural linguistic opinions that regard language as something composed of parts arranged according to a particular structure. The structuralistic assumes that the whole can be broken down into sections. Or, the whole is the sum of the parts. Each of these parts (language and skills) can be taught and tested separately. The discrete approach views that language is formed by many parts supporting the entity. These small parts have their own role to accomplish with what so called language. Thus, achieving each part comprehensively will be helpful for language learners to complete the entity. One test item is intended to measure one aspect of the language component.

In the use of the discrete approach test, the test is aimed at measuring only one element of the language component. The discrete approach test is applied on the basis of conventional understanding of a language consisting of four language skills and four language components. These four skills are listening, reading, speaking and writing. Meanwhile, the four components are language sound, grammar, vocabulary and language fluency.

Following idea from this approach related to test construction will potentially take time and energy. The test maker is demanded to develop item within one component or even its smallest unit. The test does not include context to form a test. Each item stands individually separated from one to another resulting no verbal communication that should be occurred. The success of the test takers to answer correctly of each item does not reflect the true ability of the learner in building communication in a real situation.

Since this approach is attempting to test language on the basis of individual entity a test format may measure one particular language component or skill. The key lies in the instruction given above the items as seen from the following examples:

*Find the antonym of the following words!*

1. Buy >< .....
2. Dark >< .....
3. Clever >< .....

*Find the past form of the following verbs!*

1. Break = .....
2. Stand = .....
3. Run = .....

*Read the words below correctly!*

1. Beard
2. Chocolate
3. Luggage

### ***c. Integrated Approach***

Integrated approach is the opposite approach of the discrete. This approach brings back the fallen parts of the



language skills or components in to a context. It is concerned mainly with meaning and the whole communicative effect of a discourse. This approach stated that communicative competence is so global that it requires the integration of all linguistic abilities. According to Oller (1983), “if discrete items take language skills or components apart, integrative approach put them back into a context; wheareas discrete items try to test students’ language competence a bit at a time, integrative tests on the other way around attempt to measurestudents’ capacity to use many bits all at the same time.”

The integrative approach offers test takers to gather information toward the context of a discourse in order to assist them to guess and answer test items. The test item will not individually stand by itself but it is supported with other language elements. The students or language learners will not count on a particular language element to answer certain test item. They can involve other element of language which is possibly contributing to answer test item. Combining more than one language elements can help language learners guess and find the answer needed for the test. The measurement depends largely on the collision of the components and skills of the language since Language is a group of components of language. The integrative approach tends to provide ‘context clue’ to set a test item as seen from this example of a test item:

The Noble man is *rich*, while the beggar is .....

To answer such item language learner who possibly has problem with understanding the opposite meaning of the word ‘rich’ will still be able to answer it by looking at other words to gather information which then look them as awhole to see the context of the discourse. Other words such as ‘while’ and ‘beggar’ may help them to guess what the blank part of the item should be inserted in.

Another example can be seen below:

Herman usually (go) to his office by bus. (goes)

Riya (sleep) at my grand mother's house last night. (slept)

The two items give opportunity for students to see linguistically the context of the sentences. Instead of asking what the second forms of 'go' or 'sleep' are, the students try to look at the element that may help them to decide what the past form of those two words is. Thus, the instruction of the test will play important part as a good way to administer a test. The first item will lead students to answer 'goes' because there is an element that become a clue to determine the answer. It happens also for the second item. The word 'last night' will help learners to guess the answer of the item.

The integrative approach sees language as whole of an entity leading test makers to construct test items contextually. It of course requires teachers' creativity and sensitivity to create a situation where integrative approach wishes to expect. Thus, in this approach the testee in answering a test item is required to devote their language competence and ability simultaneously.

#### ***d. Pragmatic Approach***

The pragmatic approach concerns with the ability to understand a text or discourse. This understanding is not limited to the form and structure of sentences, phrases and words and elements used in text or discourse. Comprehensive understanding is obtained through extra-linguistic aspect. It is an aspect of language comprehension beyond what is expressed through language. It encompasses everything in the form of events, thoughts, feelings, perceptions, memories and others. Pragmatic approach is usually characterized by the task of

understanding discourse, through linguistic elements that are used fairly.

Both teachers and language learners' experiences toward language competence and ability or other linguistic aspects can contribute significantly to construct this test approach for a teacher and to answer items for students. Students are demanded to understand discourse or text by involving their prior knowledge of the discourse or text, discourse structure of a text, grammatical patterns, words or dictions used in a particular discourse or text. These linguistics and extralinguistics aspects are very important for students to comprehend a text or discourse. This is the principle notions of pragmatic approach or test.

The use of cloze testing, dictation, oral interview, translation and essay writing are included in this type of pragmatic tests. Cloze test for instance require comprehensive understanding for students to finish this test. It measures language learners' ability to decode interrupted or mutilated messages by making the most acceptable substitutions from all the contextual clues. There are two types of knowledge needed to be able to succeed in performing cloze test namely linguistic ability and extralinguistic ability which refers to knowledge of the world.

In the cloze test, a number of words in the text are deleted or omitted. It usually starts from the sentence to the two texts and ends in the second sentence of the end of the text. It is intended that the initial sentence and the final sentence remain intact. Students have the task to fill in the deleted words according to their understanding of the text as a whole (Djiwandono, 2008: 27). This understanding will involve

linguistic abilities as well as other areas of the field discussed in the text (extra linguistics).

An example of cloze test can be determined from the following text where students are required to fill the missing parts of the text.

Two students from Cologne, Germany, ages 17 and 18, are accused [ ] plotting an attack at their school on November 20, the first anniversary of the shooting at Geschwister Scholl school in Emsdetten, [ ] which 18-year old ex-student Sebastian Bosse carried [ ] an attack [ ] left six people wounded by gunshot and [ ] wounded by detonated smoke bombs [ ] taking his own life. One of the [ ], age 17, killed himself on Friday after a confrontation with a high school teacher and police [ ]. The plot [ ] uncovered when a school principal [ ] discovered a web page the 17-year-old suspect had [ ], showing material that glorified the two perpetrators of the Columbine High School massacre.

The second [input type="text"], age 18, is in police [input type="text"], and has confessed [input type="text"] the plot. Materials [input type="text"] two crossbows, two air guns, instructions [input type="text"] how to build a pipe bomb and a list of potential victims had [input type="text"] seized by Cologne police when they [input type="text"] alerted that the plot [input type="text"] have taken place at the Georg-Büchner Gymnasium, located in the western district of Weiden. The items had been confiscated at their homes.

(Taken from *Elek Mathe* - [EnglishLearner.com](http://EnglishLearner.com))

### ***e. Communicative Approach***

The communicative language test approach is used to measure students' ability to use the target language in the authentic situation. This approach believes that a student is considered successful in learning the target language if he can communicate using knowledge and skills such as listening, speaking, reading and writing authentically. This approach leads to a communicative and authentic language test. The implementation of this test approach can be found in role play. Role play is able to bring the student to a real situation depending on the role and setting set by the teacher. The

teacher assesses the student's performance based on the aspect set as a criterion in the assessment. In this case, the teacher usually provides a rubric as a reference in assessing student performance in various aspects of the assessment. Tasks in the test should resemble as far as possible to the ones as would be found in real life in terms of communicative use of language. Test instructions and scoring plans should touch on effective, communication of meaning rather than on grammatical accuracy (Anon, 1990 in Komala and Sari, 2012).

This approach stresses how important non linguistics aspect plays its role. The communicative approach attempts to extend the idea of pragmatic approach and integrative approach where non linguistics aspect and non contextual aspect become key role to use language based on real situation (authenticity) both receptively and productively. This kind of approach would be ideal for assessing students' performance through authentic assessment or alternative assessment which require students to perform their competences in multiaspects that sometimes do not emerge during common test.

### **1.3 Summary**

The concept and approaches to English language testing are necessarily to be recognized as underlying basis for constructing Set of English test. They contribute very significantly to the role of teaching and the evaluation a teacher or a test maker needs so badly. These two factors assist and lead test maker to construct and develop English test items in order to achieve a good and acceptable set of English test.

The concept itself can lead students to comprehend thoroughly the essence of evaluation, test, and assessment and how to deal with them. Meanwhile, the approaches can guide students or test makers to choose which approach or approaches

appropriate to be implemented for the test items they are going to construct.

### **Exercises**

1. Which approach is appropriate for you to be applied? Why?
2. Could you please identify some weaknesses of each approach?
3. Find example of each approach in terms of the appropriate test construction!

## CHAPTER II

### Types and Function of Language Test

#### A. **Brief Description**

This chapter presents types and function of English Language Testing referring to different bases. They attempt to introduce students with many kinds of test function and test examples for certain basis.

#### B. **Relevance**

This topic is very relevant to designing test. It helps test makers or students understand very well why the test is necessarily carried out and how to deal with it.

#### C. **Learning Outcome**

Students are able to describe different types and function of test in English language.

### 2.1 Types and Function of Language Test

Language test is a tool or procedure used in conducting assessment and evaluation to know the language skills of language learners. The measurement covers two elements namely language skills and language components. Based on the criteria of how the language is studied and examined, the test is developed based on different views in understanding the nature of the language. The linguistic test may involve only one aspect of the language itself, or it may be directly linked to the factual use of language according to the communicative function of language (Djiwandono, 1996:19).

There are various type of tests conducted by teachers to measure students competences toward those two elements. These test types vary depending on the goal set and the approach employed by the teachers. Stern (1983: 340) notes that



“if the ultimate objective of language teaching is effective language learning, then our main concern must be the learning outcome”. This point of view is also supported by Wigglesworth (2008: 111) who states that “In the assessment of languages, tasks are designed to measure learners’ productive language skills through performances which allow candidates to demonstrate the kinds of language skills that may be required in a real world context”. This is because a “specific purpose language test is one in which test content and methods are derived from an analysis of a specific purposes target language use situation, so that test tasks and content are authentically representative of tasks in the target situation” (Douglas, 2000: 19).

Another thing determined the types of test in language is the approach employed by the teachers in administering a test. From the background of the language approach, the types of language tests can be grouped according to the following basis:

#### **A. Test Based on the Goal**

This type of test focuses its function on the purpose of conducting a test. There are 5 kinds of this type of tests with different function that follows.

##### *1. Placement Test*

This type of test has purpose to place someone or language learner on the group related to his/her language level. It is usually carried out in the beginning of English language program to classify participant for particular group of learner. In academic setting such as students class room placement this type of test is conducted to identify students score achievement within certain score interval decided by school authority. The range of the score interval is determined based on the school policy.

Those who get certain score will be placed at certain class category. Although this way of placing students into certain group of class can help teachers manage their teaching well but this strategy has its own flaws.

Categorizing classes into certain level in academic setting such as determining classes students within particular score level based range interval will create a discrimination that consequently will affect on the treatment the school or teachers give. However, this strategy of placement can be modified by distributing students' score heterogeneously to get equal treatment. This placement test is also commonly conducted by institution dealing with language courses. They do this to place candidates to certain level such as beginning level, intermediate level or advanced level.

## 2. *Achievement Test*

Achievement test is test type that has function to discover the result achieved by students in certain period of time regarding to the language teaching. It always relates to the material the teacher may have taught as stated in teacher's lesson plan embedded in the teaching objectives section. Students' performance on their language competences will be the target of this test type. Summative test, formative test, Midterm test, final test are examples of this types of test.

Its purpose is very clear which is to obtain information about students ability toward theme comprehended at certain time period by students. This information can be used as reference and reflection for a teacher to step further on their teaching concept. For students, it can be used as parameter concerning with his success in learning. The test maker i.e the teachers must

be able to choose proper material for the test to assure its validity.

### 3. *Proficiency Test*

It is a test that has no relation to teaching or learning program. This type of test is carried out to evaluate someone's proficiency in his overall language competence obtained through his language learning experience either formally or informally. It functions to present someone language competence with a comprehensive and standard language testing.

In language testing, this kind of test is closed to TOEFL, IELTS, TOEIC which is standardly constructed following standard rules to take it. TOEFL test for example is created to accommodate someone who wants to know their global language competence used to fulfill certain requirement or see their language competence merely. The result of the test will be a valid information for third party to take further step. This TOEFL test format is usually delivered in the forms of multiple choice item where test takers can select best options among 4 choices available.

TOEFL itself offers a rather comprehensive language test since it does not consist full language skill and component. It covers Listening skill presented in Listening comprehension section which measures the ability to understand English as it is spoken in the United States, Writing skill presented in Structure and Written Expression section that measures mastery of structural and grammatical patterns in standard written English. The language used is more formal than conversational one, and Reading skill covered in Vocabulary and Reading Comprehension section which attempts to test one's ability to comprehend the meanings and words

usage in English written forms as well as the ability to comprehend various reading materials.

Those sections are using 4-choice objective test items. Listening comprehension section covers paraphrase, short dialogue between two people, and spoken text comprehension. Structure and written expression is presented in the forms of completion and error identification. The Reading comprehension and vocabulary section is formed by answering items based on the text either explicitly or implicitly. The Speaking excludes the TOEFL presentation on language testing. Here the example:

### **Listening Section**

Time: 47 Minutes

37 Questions

Part A

*Directions: In this part, you will hear short conversations between two people. After the conversation, a question will be asked. Choose the answer that most accurately answers the question based on what is stated or implied by the speakers. Mark the answer in your book or on a separate piece of paper.*

1. What do the speakers assume about Adam?
  - A. He already left the meeting.
  - B. He is lost.
  - C. He does not intend to come to the meeting.
  - D. He is already in the room.

### **Structure and Written Expression Section**

Time: 20 Minutes

25 Questions

*Directions: This section measures your ability to recognize language appropriate for standard written*

*English. One type of question consists of incomplete sentences, with a blank showing where information is to be filled in. Choose the word or phrase that most correctly completes the sentence. A second type of question consists of sentences with four underlined words or phrases. For each sentence, choose the one underlined word or phrase that is incorrect in standard written English. Mark the answer in your book or on a separate piece of paper.*

1. Overeating, in addition to lack of attention to  
A  
nutrition, are said to be the major cause of  
B C D  
obesity in the United States.
2. The committee has met and \_\_\_\_\_.  
A. have approve the budget  
B. budget was approved  
C. its approval of the budget  
D. approved the budget

*Taken from: TOEFLCBT  
(Cliffs Test Preparation)*

#### *4. Aptitude Test*

It is predictive test which attempts to predict someone language mastery or success prior to learn the second or foreign language. This type of test is designed to test or measure someone's potential ability to learn foreign or second language and to be successful in taking it. Someone's sensivity to comprehend both written and spoken forms of language learning play very importantly to process language learning and to the success of the exposure. The use of appropriate diction, understanding to use language structure propoerly, comprehending text contextually are among several language competences

that someone may possess that can contribute to the success to learn target language. the test commonlyask students to perform tasks such as answering and responding numbers, listening to spoken text, identifying spelling clues and structural patterns, and memorizing.

##### 5. *Diagnostic Test*

This test type has purpose to analytically determine cases of language barrier faced by students or language learners. It emphasizes on the specific aspect of language competence to be diagnose and accomplished later. The teacher or the test makers try to identify and discover language difficulty. From this process, teachers are supposed to be able to provide language teaching focusing on that particular problem.

Test of Writing for example, teachers ask students to develop a paragraph then teachers identify components or factors contributing frequently to the problem of organizing paragraph writing such as the grammatical pattern used, the spelling, diction, organization and mechanics. The teachers base on these focused elements to determine students' needs in order to have special emphasis on.

Often students encounter difficulties in language learning. The difficulties encountered can be seen from the use of language both in communication and during doing a test. Errors or difficulties that occur continuously and fundamentally is a sign of students having difficulty in learning the language. Mistakes that indicate such language learning difficulties can be obtained intentionally and well planned by conducting tests administered specifically for that purpose. Such a language test is known as a diagnostic test. With the diagnostic test, information will be obtained as the basis

of the implementation of language learning in accordance with the needs and abilities of students.

#### 6. *Try-Out Test*

This type of test aims to determine the extent to which the test items that are in the process of construction show good test features and acceptable. Results of implementation of this type of test will provide comprehensive information regarding the aspects of validity, reliability, practicality, simplicity and other information such as the level of difficulty of test items, the discrimination power of the test, clarity of instructions, clarification of writing and so forth. As a result, improvements to the test device will be done in order to fulfill a good and acceptable set of test. Improvements made may involve choices when it is multiple choice items, stem of test items, layouts, and others that are lightweight.

### **B. Based on Time to Carry out**

#### 1. *Entrance Test or Selection Test*

This kind of test is carried out to select candidates or participants who are qualified for particular activity requiring certain qualification on language. The skills needed depend largely on the skills proposed by the one who carries the test/selection. This test functions to give information to test maker regarding the selection they make. This test is popularly conducted in the beginning of language program or courses. The test makers set criteria to determine participants that will go through for a selection.

The criteria is of course based on academic consideration which is set in the forms of quantitative score (passing grade). Participants who achieve to that particular passing grade will be going through and successfully completed the test. The example of this type of test can be found in SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri), Seleksi Test Masuk Peserta Didik Baru tingkat SMA atau SMP). Thus, this kind of test is also called an selection test.

## 2. *Summative Test*

This type of test concerns with an evaluation on English language teaching program conducted by teachers. Its main function is to know the result of a whole process of language teaching as an evidence to the achievement of teaching objective. Currently, this test is popularly called Final test. It is done after one full semester of teaching program has been accomplished. Its purpose is very clear that is to see students' achievement on language competences they have understood within one semester.

## 3. *Formative Test*

This test type in terms of time of the implementation is different from Summative test. This formative test is conducted after a half of teaching program has been running. The purpose and the function is the same as Summative test.

# **C. Based on the Function of the Test**

## 1. *Pre Test*

This test functions to recognize one's initial language ability at the beginning of a teaching program.



It is done before language teaching program is carried out in order that teachers have general picture of their students. By that time, teachers can anticipate any technique, method, strategy, material difficulty or level, learning objective appropriate for those learners capacity. Teachers can prepare proper and appropriate lesson plan.

## 2. *Post Test*

Post test is done to know the progress of language competence achieved at the end of the course and compare it with the earlier language teaching program determined from the pre test.

# **D. Based on How to Construct the Test**

## 1. *Teacher - Made Test*

Teacher made test is clearly a test constructed by teachers. As language teachers, they have to conduct an evaluation to have information regarding their teaching and students performances. In this case they need to assess students language competence through a test. The test is prepared and made by the teachers. In constructing test items they do not need to make deep analysis as shown by standardized test construction. No standard procedure they make to make each item of the test. As a teacher, constructing test is a routine work he or she must deal with. In its construction, this test relies more heavily on the teacher's own considerations in which the method of composing is done without following the standard test-making pattern.

## 2. *Standardized Test*

Different from teacher made test, this test requires deep analysis and critical process to develop it. It does

this process since it will be used for wider scope. National Examination is one of the examples of the scope for this test type. The items before being printed and developed must be analysed through high standard process. Several points will viewed and reviewed to get representative, proper and appropriate items, valid and reliable. Aspects such as The process takes much longer since some procedures must be carried out. The procedures are glossed as follows:

### **Planning the test**

Before constructing test item, some points must be considered seriously such as time allotment for the test should be accomplished by test takers, numbers of items that can cover the content presentation in the syllabus or curriculum and who the test takers are, table of specifications and the format of the test. These points are the initial process to begin constructing test items.

### **Constructing test**

After completing planning stage, constructing test items becomes the main activity in the process of developing language testing. Here, a knowledge about constructing a good test item plays very important part. Test makers must critically be able to accommodate any background of language competence and and their ability in developing test.

### **Critical Review**

This stage can consume some time because it needs reviewers to check each item made. The reviewers are persons who are good and qualified to review the test items. They know much about developing test and are good at language competences.

Some aspects concerning with items construction will be identified carefully and critically. Aspects such as the stem, the option, the distractors, the grammar, spelling, diction, punctuation, lay out and other physical looks are the target to be critically reviewed. When the review has been settled down, the test maker can fix or revise necessary aspects and arrange the whole presentation of test items with a complete look as what a test should be (added with test identity and heading). It is not finished yet since the revised test items need to be tried out.

### **Try out**

At this stage, test makers need to look back each items they make to assure that the items are in the level of the target learners. The items must not be too difficult nor too easy. They must be proper and appropriate. The result of try-out step will determine whether the test items are appropriate and ready to be used. To achieve this goal, each item of the test will be identified their level of difficulty and their discrimination power by implementing item analysis deeply.

The result of item analysis will come up with a statement whether particular item is rejected, accepted, or revised. The accepted items will be put as finished items and will be placed in the scheme of the test. Meanwhile, the revised items should be revised to meet its appropriateness and the rejected ones will be dropped and no longer used as accepted item. This section will be discussed further at the other chapter of this book.

### **Revision**

From the result of the try out, information of each items' weaknesses and strengths can be collected such as its index of difficulty level and index on its

discrimination power, the items score – how many test takers can answer the item correctly and how many cannot. Through these information, this stage will discover which items are then able to be used as finished items. Necessary act must be done to revise the weaknesses determined from the results of the try out.

### **Final Draft**

When necessary revision has been completed. The items of the test are organized in such away to be presented as a final draft before being delivered to the target learners.

## **E. Based on Scoring the Test**

### *1. Objective Test*

It is a test which is scored very objectively although it will be scored by other scorers. “The nature refers to the way in which assessment can be made steadily or consistently, with the same result, unchanged result even if they were repeated or performed by different assessors.” (Soenardi, 1996). One characteristic of the objective test is that it should be developed and arranged in such a way that the correct answer to the items is ascertained beforehand and made one in the form of a key answer. The key answer is used as a benchmark and a binding handle. Participants' answers are considered correct when the test participants answer according to the key answers.

Objective test provides efficient way both for test makers and test takers. Test makers will find it easier to score only by checking with available key answer. This type of test also gives wider material coverage to be made considering the number of items commonly comprise more than 20 items that can give opportunity

for test makers to put various themes or topics they have taught.

The test takers can also benefit from this test type. They have option which give them opportunity to answer. It is very efficient with just cross, circle or complete the blank part of a text with alphabetical writing. The test makers can do the test with very short time. However, it has weakness that is very complicated to construct items. Examples of this type of test can be found in Multiple choice test items, Matching, True False and completion with options provided. These test items can be seen from the example below.

### Matching

<b>Premises</b> <b>Column A</b>	<b>Responses</b> <b>Column B</b>
___ 1. Person who performs mysterious tasks no one understands	A. Facilitator
___ 2. Person who provides schooling for children	B. Trainer
___ 3. Person who enables a group to find solutions	C. Instructional Designer
___ 4. Person who instructs adults in a classroom.	D. Meeting Organizer
	E. Teacher

Taken from [http://thelearningcoach.com/elearning\\_design/writing-matching-test-items/](http://thelearningcoach.com/elearning_design/writing-matching-test-items/) by **Connie Malamed**

The above example of matching test is the most common format. Malamed (2010) states that “the matching test item format provides a way for learners to connect a word, sentence or phrase in one column to a corresponding word, sentence or phrase in a second column. The items in the first column are called *premises* and the answers in the second column are the *responses*”.

*Multiple Choice*

Andi quickly \_\_\_\_ his dinner.

- a. Drank   b. Ate   c. Drove   d. Slept

*True False*

(T / F) Based on paragraph 2 it can be said that last night football match was horrryfyng.

2. *Subjective Test*

It is a test that cannot be possibly carried out objectively. The questions given invite various answers. In this test the most important thing to note is to avoid the slightest form of assessment so as not to be too subjective. Although this test is called a subjective test but the way of assessment is expected to be more objective by providing clear techniques of scoring.

Another possible way is to reduce the level of subjectivity by assigning more than one assessors, or equipping the assessor with detailed assessment (Djiwandono, 1996). These tests are usually found on writing tests by developing essays or paragraphs. It can also be used in performance tests and others that lead to inabsolute judgment. Reading comprehension especially advanced reading requires this type of test as well.

This test presents advantage and disadvantage for both test makers and test takers. Test takers will find it very fortunate since they have opportunity to get points although their answer are not fully correct. Nonetheless, they have to think harder and find the answer very critically because most of the times the answers are various and very difficult to answer. Meanwhile, the test makers can make the test item very quickly but it takes very long time to do the correction.

## **F. Based on Criteria to Assess**

### *1. Norm Reference Test*

It is score interpretation to change raw score become final score referring to the average score of the group. This means that one's score might be interpreted by comparing his score with others who do the same test within the group. The norm here is the level of achievement gained from that group that mirrors the level of achievement of most test participants. The results of this group test only apply to the group and can not be compared with other groups. A high-value test participant in the group does not mean that he or she has a high score in any other group.

### *2. Criteria reference Test*

Final score is based on the minimum target that can be accepted as a proper achievement or as indicator of mastering the achievement. This reveals that the determination of the final score based on the acquisition of small or the lowest language skills is not associated with the level of attainment of other test participants who are working on the same test. The achievement of the lowest level of ability is defined as a sufficient level of ability or sufficient.

## **2.2 Summary**

Test types and functions of the English language test present very important aspects of comprehending any issues on English tests' construction. By knowing them, students can prepare and anticipate any process of conducting and constructing test. With their six basis and their divisions for each will lead students to understand why and how test should be carried out.

## **Exercises**

1. Mention several differences of the following test types;
  - a. Pre-test and Post-Test
  - b. Objective Test and Subjective test
  - c. Teacher-made test and Standardized test
2. What can we do to avoid or reduce high subjectivity when implementing subjective test?
3. What are the functions of conducting 'try-out test'? What should we prepare to deal with?



## CHAPTER III

### Forms of Language Testing

#### A. **Brief Description**

This chapter III presents various kinds of English language testing format. It discusses about test format for both skills and components of English language. the topic will provide explanation and exemplification of what and how to make English test.

#### B. **Relevance**

This topic will equip students competence with some various options to make a test on different skills and components. They can select and even creatively develop English test format referring to the examples given.

#### C. **Learning Outcome**

Students are expected to understand various forms of English skills and components of test.

#### 3.1 **Forms of English Language Testing**

The test in the language gives an idea of how the language test is generally carried out. The elements of good test preparation with the criteria of proof aspect is also applied in order to get a good result of on test item as well. The test in the language itself has several parts that are elements of the language itself as a component of language and language skills. There are several language test forms that are used to measure language skills partially but there is also an integral and comprehensive one.

The main target is the level of mastery of language skills, and the level of mastery of pronunciation or mastery of grammar. This chapter will show the language test used to determine the language skills of students.

## 1. Test on Students' Language Skills

This test focuses on assessing students' language skills such as test of Listening, test of Speaking, test of Reading, and test of Writing. Each of these language skills test is discussed below.

### a. Test of Listening

The main objective of listening test is the ability of test participants to understand the contents of the discourse spoken by native either directly or recorded speakers. Here are some examples of listening tests:

#### 1. *Answering question (phrase)*

In this test form, students will listen to a short spoken text (phrase mode) from a recording cassette, cd or other sources. After the text is delivered a question is given and students need to answer it. To answer the question, student are commonly given a paper or answer sheet with blank part that must be filled for each number following the spoken text delivered or there will be choices that students may cross for the correct answer.

Spoken Text → "Temperature below zero"  
Question → How is the weather?  
Students possible response → (very cold)

#### 2. *Answering question (sentence)*

This test item is similar to the above one but its spoken text is constructed in the forms of sentence.

Spoken Text → (It may soon rain)  
Question → What should I  
take? Response → (an umbrella)

#### 3. *Finding out main idea*

This type of language test form is also principally the same. The text will be spoken through a cassette

or cd by a native speaker. The different only lays in the length of the text. The spoken text may consist more than 2 sentences where the students are required to determine the main idea of it.

<p>1. A. Rail services B. Christmas holiday C. Airports D. Snowfall E. Snowstorm</p>	<p>(Standar SKL 2012)</p>	<p>storm, says official service. One person was killed late Friday in a single car crash. Two other deaths were reported Saturday as there was more heavy snow than expected. Narrator : What is the monologue about? Monolog tersebut memberitakan tentang badai salju di pantai timur "A major snowstorm slashed the east coast".</p>
<p>Jawaban: E Pada audio diperdengarkan: CNN. A major snowstorm slashed the east coast and foiled the busy holiday as airport shut down the runways, rail services were slowed and bus roads were suspended on the last weekend before Christmas. Three deaths in Virginia were blamed on the</p>	<p>2. A. Two B. Three C. Four D. Five E. Six</p>	

(Taken from Strategi & Kupas Tuntas SKL UN 2016 SMA-MA)

#### 4. Answering question (spoken text)

This form of test require test participants to listen to text which then they need to answer questions in the forms of open ended question. It is like answering reading comprehension text but it is in the forms of listening comprehension. The questions are sometimes already available at the questions sheet but it is often spoken where the test participants can respond them through their answer sheet. It can be found in some parts of TOEFL test item.

#### 5. Retelling

Retelling is one of the ways to test students' listening skill. The students will be asked to listen to the spoken text where later they need to retell using

their own words to what they have listened. Test makers need to make criteria in scoring this kind of test.

**b. Test of Speaking**

The test of speech is intended to know the use of language be it aspects of fluency, pronunciation, content, diction, intonation and grammar. In performing this test, criteria or assessment rubrics should be made to obtain the desired level of student skill or ability. Here are some examples of speaking tests.

**1. Telling a picture**

Students are given pictures to be told. The pictures can be a single picture displaying various objects and or activities done by people or animal. Pictorial series indicating sequential picture can also be used as test item for students to tell. The instruction given is merely asking students to describe the pictures. It is suggested that teacher has rubric with him or her used as a guideline to score the students' performance.

**2. Retelling using students' own words**

In this type of the test, students are required to tell again the spoken texts which can be in the forms of story, news, ect where students need to retell them using their own words (text paraphrasing). Test makers only prepare spoken text or written text as reference for students to narrate or describe orally using the own sentences completed with rubric and scoring technique to be used as a guideline for giving score.

3. Speaking freely (deciding the topic to be discussed)  
Speaking freely here does not mean that students are totally free to speak anything without any control but they are free to describe, to narrate, to expose things the teachers have already decided particularly on the topic the students should speak about. They do not refer anymore to pictures, spoken texts or written texts to be delivered orally. They can develop or speak about topic given with time allotment that has been set and other possible criteria set by the teacher.

4. Conversations

It is a test of speaking where students are asked to perform a dialogue or role play with particular setting or situation made. Since it is a performance test it requires rubric and scoring criteria prepared by the test makers. Its purpose is to minimize subjectivity.

**c. Test of Reading**

Tes kemampuan membaca lebih menekankan pada aspek pemahaman siswa terhadap isi teks baik itu yang bersifat eksplisit maupun implisit. Butir test yang disusun bisa dalam bentuk open ended ataupun pilihan ganda yang diletakkan langsung dibawah teks bacaan. Berikut beberapa contoh tes membaca yang bisa dilakukan.

1. Completing text

Read the following unfinished text by inserting a proper word.

LATIHAN SOAL	CLOZE PASSAGES
<p>I recently took a trip to New Hampshire and had what I would call my first gonzo ... (21) ... taking up a mountain. I'm going up hiking before but this was without a doubt the best experience of my life. My friends, Sean and Madeline ... (22) ... some pictures during the trip. We went two hours straight before we reach the summit. The storm clouds were heading right toward us and you could actually see the rain falling from the cloud. It was like a wall of water moving at an alarmingly ... (23) ... pace ready to engulf us.</p> <p>1. ...</p> <p>A. Experience B. Influence C. Experiment D. Expedition E. Independence</p> <p>2. ...</p> <p>A. broke B. bought C. brought D. took E. put</p>	<p>3. ...</p> <p>A. Wide B. Strong C. Long D. Bright E. Fast</p> <p>(Standard SKL 2012)</p> <p><b>The Man and The Monster</b></p> <p>Once upon a time there lived a young man, named Nazar. He lived a year poor life. One day, Nazar went to town to look for a job. He had to climb up the hill and ... (43) ... a thick jungle to get to a narrow town. Suddenly, he heard ... (44) ... from a bush. When he approached, he heard some people gathering around a big tree. Nazar realized that the people were doing against God's order. Then he took out a big ... (45) ... axe and planned to cut down the tree.</p> <p>4. ...</p> <p>A. Checked B. Examined C. Inspected D. Went through E. Hunted through</p>

(Taken from Strategi & Kupas Tuntas SKL UN 2016 SMA-MA)

## 2. Answering question

Most of the times this test is composed to comprehend reading text placed before questions that follow. The questions are commonly WH-questions confirming either implicit or explicit ideas found on the text. It is often followed by questions related to vocabulary mastery regarding to the text. This test format of vocabulary may form various formats such as matching, completion, or multiple choice items depending on teacher preference and time allocated.

## 3. Summarizing

This test of reading type requires students to read passage/s in order to understand comprehensively the text is mainly discussing about. The students are asked to make a summary (important points) of

the passage/s using their own sentences. The summary of the passage/s mostly will be assessed on the basis of the students comprehension regardless their grammatical patterns used. However, the teacher will not neglect the language used since they will put scoring criteria to assess. It of course tends to be subjective by providing scoring criteria it will reduce teachers' subjectivity.

**d. Test of Writing**

Writing skills test aims to measure the level of mastery of the ability to express the mind to others in writing using the language of good and true and plot of writing presented plainly and coherently. Writing skills test leads to a performance assessment whereby students are comprehensively required to demonstrate their mastery of language, thought ideas, and writing organization.

1. Telling a picture

The test makers or teachers provide series of pictures to be developed and written by students in the forms of composition. The composition made must follow what the pictures are presenting. Students are required to describe and tell what is drawn in the pictures logically and chronologically. The pictures are set to come up with a composition of paragraph or essay.

Instruction:

*Compose the following jumbled pictures into a chronological composition within 100 words.*



*Taken from EAP Book of University of Muhammadiyah Gresik*

## 2. Free writing

This writing test type requires students to make a free composition where the test makers only provide instruction with topic the test makers have to develop in. However, the topic can be free either. Several requirements are also given to do the task such as the number of words should be used (the length of the composition), the language used ect. The topic is given freely and students are free to develop their composition. Rubric for guideline scoring is badly needed to reduce subjectivity for this performance test.



### 3. Writing academically

This test type asks students to develop paragraph or essay with well organized composition, appropriate language used, and logic content. Academic writing here refers to developing composition following rules of writing paragraph or essay academically. There will be an attention on many aspects to be assessed such as grammatical patterns used, words or diction, spelling, ideas, organization, and mechanics including punctuation and capitalization. Limiting number of words or composition length is commonly suggested in the instruction. Since writing test is claimed as performance test then preparing a comprehensive rubric completed with scoring technique is highly suggested.

## 2. Test on Students' Language Components

The linguistic component capability test is intended to measure the level of capability in more detail than the mastery of one's language skills. This test emphasizes the use of language technically and more operationally. This is because the desired language skills are more directed to the technical aspects of language that are generally reserved only for those who want to learn the language profoundly and thoroughly. There are several levels of mastery of the components that can be measured in language skills component test that is:

### a. Vocabulary Test

The vocabulary test is intended to know one's mastery of the meaning of the word either in the form of a

passive-receptive or active-productive understanding.  
Here are some examples or tests of vocabulary mastery:

1. Showing object

In showing an object test format, the test item shows the words followed by pictures that must be chosen by students as the answer.

(Show the object of the following words)

Patio

Balcony

Terrace

2. Performing

This vocabulary test requires test makers or others to act or perform any activity tended to the intended words.

Whispering (the test takers/participants act as suggested)

Nodding (the test takers/participants act as suggested)

3. Translation

This vocabulary test ask students to translate words into their first language or on the other way around.

Reluctant – enggan

Passage – Lorong/ Lompongan

4. Synonym

Calm = quiet

Pretty = beautiful

5. Antonym

Receive x send

Empty x full

6. Mentioning Words

A place where students go to study = school

Having a lot of wealth = rich

7. Completing a sentence

We can read news and advertisements in a ...  
He went to ... to see film.

Of the above various test types on language components, the test format can be various as well. The test format may come into objective or subjective ones depending on the test makers' goal. However, passive-receptive commonly requires objective one while active-receptive needs a subjective test format. The instructions made is very crucial despite its various test formats.

**b. Test of Grammar**

The grammatical component test demands a test participant to demonstrate their ability to grasp the right and correct grammar that includes the ordering of sentences (syntax) as well as the wording (morphology). The followings are examples of this kind of test.

1. Words formation

*Identifying the root*

(Write down the root of the following word)

Comfortable ... (derive from comfort)

*Derivational affixes*

(Tambahkan kata turunan berikut dengan imbuhan)

king ... (kingdom)

*Adjusting word forms*

The sun (to rise) in the east. (rises)

You have to finish your work (quick) – (quickly)

2. Phrases formation

*Arranging words*

Story, interest (interesting story)

*Completing phrase*

Dog ... (barking dog)

*Forming phrase*

A season when it rains a lot... (rainy season)

*Defining phrase*

Village people = people who live in a village

3. Building sentence

*Knowing sentence*

The man behind the gun. (Not a sentence)

The man behind the gun is my uncle. (Sentence)

*Constructing sentence*

driver, wash, car – (the driver washes the car)

*Arranging sentence*

Have to, the song, again, sing, they

(They have to sing the song again.)

*Changing sentence*

Change the following affirmative sentence into (- and ?)

Truck driver often washes his car every week.

Truck driver does not often wash his car every week.

Does truck driver wash his car every week?

**c. Test of Pronunciation**

1. Knowing language sound

(Circle the words that use sound –sh)

ship            sing

nation        cash

2. Differentiating Language sound

(write 'S' if the two words are the same and 'D' if it is different)

(...) Pen – Pant

(...) ship – sheep

- (...)Tent – tend                      (... ) sick – six
3. Pronouncing Language sound  
(Pronounce the following language sounds clearly)  
Phonetic transcription
  4. Pronouncing words  
(Pronounce the following words)  
Myth                      rhythm                      leisure
  5. Pronouncing sentence  
She sells sea shells on the sea shore.  
Peter Piper picked a pack of pickled pepper.
  6. Pronouncing words pairs  
Penny – benny                      write- ride  
Few – view                      treat – trade

#### **d. Test of Dictation**

1. Standard dictation  
This type of test is very common in the early 80s to 90s. Dictation becomes very popular test used by teacher to determine students' competence in listening and in spelling competences. This type of a test needs teachers' carefulness in reciting every single words. The teachers will read the whole text clearly and carefully with certain tone and intonation. Then, it is read partly as the student writes them. At last the whole text rereads.
2. Partial dictation  
Different from standard dictation, this type of dictation test is rather simple. The test participants are provided with a text where some parts of the text are deleted to be fulfilled by the students. The text then is read with appropriate tone and intonation referring to students condition.

### **3.2 Summary**

Test format of English language skills and components are necessarily to be recognized and best implemented by students. Constructing English test requires students' or test makers' creativity and innovation. Test formats from various forms can assist students to develop many alternative test item from skill courses to components of language. Different kinds of test format provide test makers new horizon in constructing English test items.

## CHAPTER IV

### Developing and Improving Test Items

#### A. Brief Description

This topic is very crucial concerning to English test construction. It includes things to avoid aspects such as ambiguity, acceptability, ect in developing English test items for different kinds of test formats. It aims to improve the quality of the test items by looking at the characteristics of constructing good items.

#### B. Relevance

Developing and improving test items play very important and crucial part in constructing test. It helps test makers to understand how to develop and improve their test covering aspects of validity, simplicity, practicality, and reliability that may guide them to create acceptable a set of English test.

#### C. Learning Outcome

It is expected that students are able to develop and improve English test based on characteristics and criteria how to make a good test.

A good test is constructed very carefully by considering the goal of administering the test and other factors relating to factors contributing to making a good test such as the language used, material coverage, physical look of each item. In developing test item, principally some requirements and procedure to make it must be applied to have an ideal test items.

There are guidelines to develop and to improve the quality of the test items. These guidelines are applied differently for different test formats. However, general guidelines are available for both objective or subjective test.

These guidelines are a condition that test makers must refer and meet to in order that the test presents ideal and good test items properly and appropriately testable. But before discussing technically related to the development and improvement of test items, it is better to know characteristics of good test used as fundamental base in constructing test items. Understanding the characteristics of a good test will lead to the quality of good test measurement (evaluation). Consequently, the information obtained related to students' performance is accountable and can be used as reference to decide further policy or strategy to determine language teaching performances.

#### **4.1 Characteristics of Good Test**

Characteristics of good test can be identified from the following qualities of the test:

##### **A. Validity**

A test is valid when it measures what it is supposed to measure. When we design our tests, we need to be sure that we are not asking students to do things which are completely different from the activities they have taken part in during our lessons. If a test is for checking students's reading ability, then it should check reading skills but not e.g. writing or listening ability. To make sure that the test is valid, it needs evidences. These evidences can be checked through kinds of validity the test items possibly have.

##### *1. Face Validity*

Face validity is defined as a test activity showing relevance of conducting test. It shows what is being tested is reflected in the activity of the test. When testing on speaking skill or class the activity shows an activity of students having conversation or any other activities indicating speaking test not other test. When testing



writing for example, the activity will show students develop a composition or activity to write not to speak nor to read.

## 2. *Content Validity*

“Content validity is defined as any attempt to show that the content of the test is a representative sample from the domain that is to be tested (Fulcher, 2007: 6)”. Content validity requires test makers to adopt the test items and the table of specification to the presentation of material presented in syllabus. The relevance of topic presentation within test items to syllabus indicate the validity of the content being tested. The items of the test must adequately cover test specification or the language domain specified and this specification must also be relevant to the curriculum and the syllabus developed which then translated into a lesson plan completed with assessment with it.

It is worth to say that to establish or to check content validity It needs an expert judgment to guarantee that the test has been valid. This expert judgment is suggested not the one who gets involved directly in developing the test items and he or she must be the one who is an expert in language testing.

Another thing that should be noted that content validity is not the only validity the test makers must establish. They need to establish another type of validity such as face validity, criteria validity, or construct validity.

## 3. *Criteria validity*

This type of validity expects a comparison of previously held tests by correlating with tests that have

currently been made. Of course the tests used as criteria or benchmark have already qualified as a good test as the validity and reliability aspects have been met. If the correlation results show a positive or significant result then the measuring instrument has good criteria validity aspect.

There are two types of validity of this type namely Concurrent validity and Predictive validity. Djiwandono (2008: 166) says that “the difference between these two validities is the spacing (time spacing) of the two test scores compared to the one held at almost the same time, the other at two times with the longer difference”.

#### 4. *Construct validity*

Construct validity is a way to prove that the test made has a validity aspect based on the concept of the field of knowledge (construct) measured. This means that the construct validity adjusts between the results of the measuring instrument and the ability to be measured. The English test score result, for example, should measure and reflect the same construct as the ability to be measured. Thus, a measuring instrument containing construct validity is a measuring instrument whose results are relevant and appropriate to the characteristics of the behavior or capability being measured. Each test item shows conformity with details of the ability to be measured.

Therefore, in preparing the test items it is necessary to make the test specifications (kisi-kisi) so that the measuring instrument will base the development and preparation of the test item on the specified grid or specification. In short, items written in syllabus or RPP such as competency standards, basic competencies and indicators can be elaborated and described and measured in each item on a test. Arikunto (1997: 64) asserts that “a test is said to have construct validity if the items

constructing the test measure every aspect of thinking as mentioned in a specific instructional objective.”

## 1. Reliability

Reliability is a measuring tool used to measure the consistency of a test. Tests that satisfy the reliability aspect provide an unchanging result even if the test is done more than once. This means that the test is reliable because the result does not change significantly. A good test is a test that has passed the reliability aspect in which results are consistent even though the test is performed more than once on the same test participants at different times. The test with the same level of reliability as the previous test makes the test a good test. The same level of test reliability was scored with a score of 1.0. Nevertheless, this achievement almost never happens because of the element of change in answers made by the test participants. Thus, generally the level of reliability falls within the score interval of 0.1 - 1.0.

There are kinds of reliability offered to assure that the test is reasonably consistent. The followings are the kinds of reliability taken from Djiwandono (2008:171)

### *a. Test-Retest Reliability*

Test item is tested twice to the same participant at different time (Repetition). As a result, there will be two set of English scores which then is put in particular formula to check the test reliability. One of this formula is using Pearson Product Moment or Kuder-Richardson formula.

### *b. Equivalent Test Reliability*

This type of reliability is made by using two tests constructed equally then it is given to the respondent or test object at the same time.

c. *Split-Half Reliability*

Participants are given a test at once then they are assessed. Test item is divided into 2 sets of test deriving from odd and even numbers of the test items. From here, the test makers will receive 2 lists of scores from test takers. Similar to the previous types of reliability, these two set of English scores are measured or correlated employing Pearson Product Moment or Kuder-Richardson formula as two among other reliability formulas.

d. *Cronbach Alpha 1*

This kind of Reliability can be used as an alternative calculation for Split-Half Reliability. Here, there is only one test unlike the two previous kinds of Reliabilities which employs 2 list scores of test takers (using Pearson Product Moment - PPM)

e. *Cronbach Alpha 2*

The reliability is measured in a whole because this type of Reliability is different from CronbachAlpha1. Test of writing or essay test is applicable for this type.

f. *Kuder Richardson K-R20*

This type of Reliability serves only once test taken. The score of the test is scored by dichotomy-point 1 given when the answer is correct and no point when it is incorrect

g. *Kuder Richardson KR-21*

This type of Reliability is only different in the form of its simplicity with the K-R20. This reliability measurement is often used although it gives lower result than of K-R20.

h. *Rater Reliability*

The scoring of the test is done 2 times by the same scorer (1 scorer). This reliability type goes very well only on test of writing or speaking(Intra rater). There will be two scores for each test taker. The scores, then, are

analyzed using Pearson Product Moment. The scoring of the test is done by two raters to the same test takers (2 scorers). The difference here is only on the time. The test can be scored at the same time. PPM is employed here.

*i. Intra-Rater Reliability*

This type of reliability suggests the existence of two sets of values given by an independent rater (rater) in which the rater scores the same test and participants but the test is performed 2 times at different times. Thus, there will be 2 sets of values to be correlated by using correlation calculations to produce the estimated level of test reliability. Thus, this type of reliability emphasizes the test item or 1 set of measured and correlated tests.

*j. Inter-Rater Reliability*

In contrast to Intra rater reliability, inter rater reliability emphasizes the level of reliability in the judgment (rater) rather than the test question. In this case, the inter-rater agreement becomes an important and principal element for determining the inter-rater reliability of the test item (inter rater agreement). Consistency of assessment by more than one assessor is important. Consistency of this rater is the reference in determining the reliability of this type of reliability and not a matter of the test item.

Therefore, a pointer, criterion or rubric of assessment is necessary in applying this type of reliability in order to obtain consistent results. This type of reliability is usually applied in the assessment or implementation of subjective tests such as writing or performance tests. The agreement on the results of the assessment of the same test participants shows that the reliability is high.

### k. Estimated Reliability

In calculating this level of reliability it is estimated based on the number of test items ( $k$ ), average score ( $x$ ), and standard deviation ( $S$ ). The first step is to determine the level of difficulty of the test which is realized in the form of average score. If the average score reaches an interval between 70% to 90% then the test is relatively easy and if the average reaches the 50% to 70% interval it is quite difficult. From this estimate then 2 tables are made representing the two poles (easy and difficult table). The following 2 tables are taken from Diederich (1971:10)

Approximate Reliability of Easy Tests (average 70% to 90% correct)									
Number of items (n)	20	30	40	50	60	70	80	90	100
If S.D. is .10n	.21	.48	.62	.69	.75	.78	.81	.83	.85
If S.D. is .15n	.68	.80	.84	.88	.90	.91	.92	.93	.94
If S.D. is .20n	.84	.90	.92	.94	.95	.96	.96	.97	.97

  

Approximate Reliability of Hard Tests (average 50% to 70% correct)									
Number of items (n)	20	30	40	50	60	70	80	90	100
If S.D. is .10n	—	.21	.41	.53	.61	.66	.71	.74	.77
If S.D. is .15n	.19	.67	.75	.80	.84	.86	.88	.89	.90
If S.D. is .20n	.74	.83	.87	.90	.92	.93	.94	.94	.95

As an illustration, there is an English test containing 50 multiple choice questions with a total of 100 test takers. From the test results then obtained the average score is 25 with the standard deviation is 7.50. The next step is to determine the use of table A or B by dividing the average score (25) by the number of questions (50). The result is 50%. This means that the problem goes into the difficult category of table B. Of the number of 50 test items, in table B (column number of test 50 items) there are 3 results of reliability that is 0.53; 0.80; and

0.90. These three alternative scores are an indicator of the reliability level of this type.

The other determination can be done by dividing the amount of standard deviation by the number of test items and the result of this division will bring up a certain number to be seen in the table. So  $7.50 : 50 = 0.15$ . in table B in the number of items column, point 0.15 falls to the number 0.80. So the level of reliability is 0.80.

Among those types of Reliability, the following will discuss one among others to meet criteria of being consistent and reliable of the test item. Someothers may be found through program application which have been available in the market place or computer program. Below is the formula of reliability test using Kuder Richardson formula.

$$r_{xx} = \frac{K \cdot S_x^2 - \bar{x}(k - \bar{x})}{S_x^2 \cdot (k - 1)}$$

Where

X : Scores obtained from sample group of 16 items with 1 item has 1 point

F : Number of test participant who get the particular score

FX : F x X

X : X - Mean (14 - 10.25= 3.75)

Mean : FX : Number of students

$\chi^2$  : X x X

$$F \chi^2 : F \times \chi^2$$

Going through the formula above, it requires some units to be determined such as:

$$\bar{X} = \frac{\sum FX}{n}$$

and

$$\chi = X - \bar{X}$$

And

$$S = \sqrt{\frac{\sum F\chi^2}{N-1}}$$

To get clear understanding to this technical calculation, the following illustration is made: Suppose the process of the try out has been carried out which attempt to check the appropriateness and properness of the test items through item analysis. It discovers that among 22 items there are 16 item are accepted. These 16 items are then identified whether they represent the table of presentation made earlier or not. When it is done, this set of test items then is tested to 40 test participants to see its consistence. Referring to the formula above the process can be seen from the following: First, the data is put in the table shown below and second the numerical data is processed by inserting them into the formula above.



X	F	FX	$\chi$	$\chi^2$	$F\chi^2$
14	3	42	3,75	14,06	42,19
13	8	104	2,75	7,56	60,5
12	8	96	1,75	3,06	24,5
11	5	55	0,75	0,56	2,81
10	2	20	-0,25	0,06	0,125
9	1	9	-1,25	1,56	1,56
8	4	32	-2,25	5,06	20,23
7	2	14	-3,25	10,56	21,13
6	3	18	-4,25	18,06	54,19
5	4	20	-5,25	27,56	110,25
		410			$\Sigma F\chi^2 = 337,505$

The table above help us put numerical data into the formula of Kuder Richardson as described sequently and completely as follows:

$$\bar{X} = \frac{FX}{n} = \frac{410}{40} = 10,25$$

$$\chi = X - \bar{x}$$

$$S = \sqrt{\frac{\sum F\chi^2}{N-1}} = \sqrt{\frac{337,505}{40-1}} = \sqrt{\frac{337,505}{39}} = \sqrt{8,65} = 2,942$$

$$r_{kk} = \frac{K \cdot Sx^2 - \bar{x}(k - \bar{x})}{Sx^2 \cdot (k - 1)}$$

$$= \frac{16 \cdot (2,942)^2 - 10,25(16 - 10,25)}{(2,942)^2 \cdot (16 - 1)}$$

$$= \frac{(16 \cdot 8,66) - (10,25 \cdot 5,75)}{8,66 \cdot 15}$$

$$= \frac{138,56 - 58,94}{129,9}$$

$$= \frac{79,62}{129,9} = 0,613$$

The reliability test of the 16 test items through the formula above is 0,613 which means that it has reliable enough considering reliability interval starts from 0,2 to 1,00.

### 1.3 Simplicity

Simplicity has the meaning that test questions should be compiled using clear, correct and simple language. Preparation of instructions, for example, should be avoided any confusing or ambiguous statements. Likewise, in making statements, ambiguous statements should be avoided to produce the ideal item.

### 1.4 Practicality

Practical test is a test that is easy to be carried out, easy to score with clear instructions within and more importantly it spends less financial support to conduct the test and it does not spend much time to held on. Practicallity refers to the inter resources availability to design, to develop and to use the test. The resources here include material, human, time, and financial resources. These resources have relationship and complete one to another to form a test, to process a test and to conduct a test.

Practicality can be achieved and measured by the success of the availability of those sources. Spending high cost for a test particularly on language testing and teaching would be impractical. This means that test should not be excessively expensive, stay within appropriate time constraints, be relatively easy to administer, and have a scoring or evaluation procedure that is specific and time-efficient.

## 4.2 Test Items Guidelines

As mentioned earlier in the beginning of this chapter that test items would be ideal and physically look good if it follows guidelines to assure its acceptance. The guidelines for every test item format will be different one to another. Nonetheless, there are general guidelines applicable for type of tests either objective or subjective test type. The following discussion is emphasized on both general and specific guidelines for developing and improving test items for various types of test formats.

### a. Multiple Choice

This test is used to measure more complex learning outcomes. It is concerning with aspects of memory, understanding, application, analysis, synthesis and evaluation. This form of test consists of the main problem and answer choices. The bearer of the subject matter can be put forward in the form of questions and statements called stem. While the answer choices can be words, numbers or sentences called options. Within the options, there is one correct answer among other three choices. The three choices are called diatractor which functions to distract students' intention to choose the answer of the four option. Only students who are ready and understand well toward the topic area will not be distracted with the distractors.

Multiple choice test item is one of the most common test formats used by test makers to do an assessment. It is comon because it can cover a comprehensive material coverage. Although it takes long time to construct this test format but it is very efficient to score it. In this case, the test makers can count on someone else to do correction as long as they provide key answer completed with how to score the test. Constructing this test format, however,

require critical thinking in order not to present ‘easiness’ for test takers to answer each items. There are some points to ponder to make this test format look perfect and ideal especially on the items made.

1. Each item should be written at the level of students’ proficiency.

Multiple-choice item for any one question should be about the same level of difficulty, and ideally the sentence context should not be difficult for students to read. The subject matter or the stem should be written clearly and decisively not to cause different understanding or interpretation. Each number contains only one problem with language that can be understood by the students.

2. Avoid giving clues in one item that will help students to answer another one.

A good test item will not give clues for test takers to answer another item. This circumstance happens a lot whether test makers do it purposively or unpurposively. The step sometimes may provide clue to answer particular item somewhere in different number of the test. This must be avoided.

3. Avoid making unintentional clue for test takers to easily answer test item unless the test item is proposed for testing grammar for example. Look at the item below.

*Fahri needs to wake up earlier so he wants to buy an ...clock.*

*a. time   b. alarm                      c. watch                      d. bell*

4. All the parts of each item should be on one page.  
The lay out of the multiple choice item must be well constructed to avoid confusion for test takers to do the test. It must be guaranteed that each item of it must be presented within one page. Putting parts of test item in different page will cause confusion to happen. The stem and the option must not be separated in different sheet.
5. Each question must provide one correct answer.  
There must be one correct answer for this form of test to occur. There are commonly 4 choices of answers with 3 of them are the distractors and 1 is the correct answer. Test makers must carefully pay attention on this part to avoid ambiguity.

*She sent the ... yesterday.*

*a. letter      b. gift   c. food   d. books*

6. Avoid making choices stating ‘all the above answer is correct’ or vice versa, making choices stating ‘none of the above answer is correct’. Or, ‘choice a, b, and c is incorrect’.
7. The options have similar entity (length, quality, level).  
This means that the options must be homogenous, plausible, and well functioning. The length of the choices for example must be the same to avoid easy guessing that may trap students to answer either correctly or incorrectly. This will ruin the quality of the item. When the options are not logic with the stem it will cause ambiguity and confusion and when the options are not homogeneous it will lead test takers to

suspect that the unique option shown with various forms can be the answer.

*“Many teachers and principals argue...”  
What word would best substitute the principals?*

- a. students*
- b. parents*
- c. headmasters*
- d. employers*

8. Avoid formulating double negative stem to prevent misinterpretation toward that statement. For language skills, double negative usage is permissible when the aspect to be measured is precisely the notion of the double negative itself.

Example:

*The following statement is best representing the main idea of the second paragraph of the text except...*

9. The item must be relevance to the indicator proposed in lesson plan to meet the test validity especially on content validity.

**b. Matching**

This test format requires test takers to understand and know the basic concepts of matchmaking tests. The format of this test consists of 2 columns in which column 1 serves as a matter or premise while column 2 is a choice of answers or responses arranged randomly. This test format is intended to measure students' memory. Students should be able to determine the choice of answers as requested in column 1 (premise).

Some consideration must be made to administer this form of test as follows:

1. The premise is placed at the left side provided with number for each item of the premise and the response is placed at the right side provided with alphabet to show the difference of the two sides.
2. The whole parts of item (premises) and responses should be on one page to avoid practicality and easiness for students to read.
3. The options should usually be shorter than the premises/stems.
4. The premises and options should be logically related to one central theme that is obvious to the students.
5. The number of items is suggested below 15 to avoid time consuming and guessing mode or opportunity since the longer the items made the more chance provided for the students to guess the answer.
6. The responses part should consist more options to avoid easy guessing and right answer.
7. The premise and response are constructed very briefly.

**c. True and False**

This form of 'true' and 'false' questions is an objective matter by showing two opposite choices of 'true' or 'false' statement. This form of questions is structured to ask very complex concepts, ideas or facts. The basic concept of making 'true' and 'false' tests needs to get the most attention in preparing the tests. In this case, the problem created must either show true or false claims. The point is that developed statements should not present dubious statements.

There are several things that must be considered in preparing this right-wrong problem is as follows:

1. The instructions for working on the questions should be clear.
2. Statements made should be a capture that contains no confusion or uncertainty such as: perhaps, sometimes, in general, mostly.
3. Statements with multiple negative expressions should be avoided.
4. The statement should be short and simple.
5. Statements are clearly arranged and show absolute 'true' or absolute 'false'
6. The number of true answer questions is recommended equal with the number of False answers to avoid the tendency of students in guessing answers.
7. Setting or placement of correct answers and wrong answers arranged randomly.
8. Each of the 'true and false questions contains one idea and stands alone by not hanging on to other questions or statements.
9. Statements are made and developed appropriately and not quote directly from textbook sources. This is done to avoid aspects of memorizing for students.
10. Statements with multiple or more meanings shall be avoided.

### **4.3 Summary**

There are criteria that must be fulfilled for test makers to develop and improve the quality of the test they make. These criteria are made to present a good and acceptable English test items. To fulfill these criteria, characteristics and things related to what to do and what not to do should be paid into a great attention. Aspects of Validity, simplicity, practicality, reliability and even physical look of the test item are among them.



### **Exercise**

1. Give examples of English test items for each point in the things that should be avoided in constructing a test for the following test format:
  - a. True-False
  - b. Matching
  - c. Multiple choice
2. Practice to make two sets of scores obtained from English Test that you can find, then put them into either Pearson Product Moment or Kuder-Richardson formula to find its reliability aspect.

## CHAPTER V

### Analysing Test Items

#### A. **Brief Description**

This chapter discusses test item analysis. After students have been familiar with developing and improving English test items, they must be able to determine whether each item of a test they make is acceptable through a process of simple measurement.

#### B. **Relevance**

This chapter significantly contributes to the competence of the test makers in justifying whether the item is worth to be used or not. Indeed, characteristics and criteria of making good item have been met but test makers must ensure that each item statistically is accepted. It can be seen later through a process of measurement to find out each item index of difficulty level and its discrimination power which then determine the acceptability if the item.

#### C. **Learning Outcome**

Students are expected to be able to justify whether the item they are constructing qualifying or not.

### 5.1 **Item Analysis**

As described early in previous chapter, there are two types of test based on the way how to score it namely objective test and subjective test. The function of these two test type has been exposed clearly including the meaning of each. This chapter will focus its discussion on analysing objective test items on 2 or more than 2 choices such as Multiple Choice Items, True Fals or the likes.

Item analysis is clearly very crucial process to compose standardized test. This step is done after the raw test items have



There are three columns and three rows drawn in the table. The first column represents the number of items (22). The second column is test participants that consists of 40 participants. The last column displays items that are correctly answered by the test participants. This column can also represent students' score with 1 item 1 point. So, if there are 22 items can be answered correctly by students they will get highest score 22. Since the main purpose is not to know students' ability toward mastering language the score then will not be the main target. The main target is the items not the students. Students are used as target of the try- out although actually teachers can use this try out result to test students' competence.

We now come to the three rows drawn from the table. The first row describes the number of the test participants and the second row presents each item score presentation obtained from each student ended with its correct score for each. The third row displays raw score obtained by each participant in answering the whole items. When we finish with the presentation of this table we now continue to put those data into the following table of Frequency distribution of Students' score on the Try-out.

Table 2. Score Distribution and Group Classification

RAW SCORES	f	GROUP
9	4	LOW
10	3	
11	2	
12	3	
13	3	MEDIUM
14	5	
15	8	
16	6	HIGH
17	3	
18	3	
N = 40		

The table displays students' score distribution obtained from the try out. The table presents three columns and three rows. The first column displays raw scores collected from students' try-out result. It begins from the lowest to the highest scores gathered. The second column is the frequency of distribution where it tells us about the number of certain score obtained by students. Score 9 for instance is collected or obtained by 4 students. The last column displays group classification to determine whether score obtained by some students is categorized into low, medium, or high group. Meanwhile, the rows of the table present three rows with the second row of the table tells us about groups' distribution into three categories; low, medium, and high.

Classifying into groups of three derives from the number of students multiplied by 30%. This 30% is taken from expert judgment of 25% - 30% interval. It is possible to take another percentage between the intervals. So,  $40 \times 30\% = 12$  (take the first total 12 of the upper sequence from the 'f' and another total 12 from the bottom of the table above). From the table, it can be drawn that high group score starts from 16 to 18 and lower group starts from 9 – 12. The rest of the interval score belongs to medium group. This process principally gives information related to group classification obtained from students' score collection.

From this process the next step to do is to give judgment to the whole items (22 items) whether they are appropriate and accepted as standardized test item (fixed items). Going through this last process we can have a look at the table below:

Table 3. Items' Justification

Items	Correct Answer	Wrong Answer	HGC	LGC	IDL	IDP	CONCLUSION
1	23	17	8	5	0.57	0.24	Accepted
2	29	14	8	6	0.65	0.25	Accepted
3	8	34	3	0	-0.15	-0.25	Revised/Disarded
4	24	15	8	5	0.63	0.31	Accepted
5	26	14	8	5	0.65	0.31	Accepted
6	33	7	11	12	0.83	-0.08	Disarded
7	26	14	10	5	0.65	0.42	Accepted
8	23	17	8	4	0.58	0.3	Accepted
9	25	15	10	5	0.63	0.42	Accepted
10	26	14	8	5	0.65	0.31	Accepted
11	4	36	3	1	0.1	0.16	Disarded
12	24	15	8	5	0.63	0.31	Accepted
13	29	13	8	6	0.68	0.25	Accepted
14	25	11	12	6	0.75	0.5	Accepted
15	27	13	9	8	0.68	0.33	Accepted
16	26	14	11	5	0.65	0.67	Accepted
17	34	5	10	11	0.9	-0.08	Disarded
18	23	17	11	2	0.58	0.75	Accepted
19	26	14	11	4	0.65	0.58	Accepted
20	33	7	8	12	0.83	-0.21	Disarded
21	33	7	11	12	0.83	-0.08	Disarded
22	27	13	10	5	0.68	0.42	Accepted

There are eighth columns displayed from the table above. This table presentation displays a process of determining each item within the whole set of test to justify the acceptance of each item as standardized and fixed item as seen in conclusion column. The process starts with column 1 informing the number of raw items. The second column indicates number of test participants in answering the item of the test correctly. The third column informs the number of test participants who answer the item incorrectly. To obtain a clear picture of the status of each item, HGC column which stands for high group who answer item correctly is presented. HGC derives from previous table regarding to group classification. Based on the analysis, the HGC score is 8 for item number 1 (see table 2) whereas LGC which is low group who answer the item correctly reaches 5 participants for item 1.

This process is then continued to determine its index of difficulty level (IDL) and index of discrimination power (IDP). The IDL score is obtained from the number of students' correct answer for each item divided by the number of test

participants multiplied by 100%. Take an example of item 1 where correct answer is 23 and the test participants is 40. The result is 0.57 of its level of difficulty. This level can be accepted as appropriate and proper test since the accepted tolerance for this IDL is in the interval of 0.2 – 0.8. The IDP has its own way to justify its score level. IDP score is obtained from HGC subtracted by LGC then divided by 12 (expert judgement point).

Again, an example of this technique can be taken from item 1. The score of HGC is 8 and the LGC is 5. After going through mathematical calculation the IDP reaches 0.25. This level is accepted since Index of Discrimination power score starts from  $\geq 0,2$ . When both IDP and IDL is accepted the item is accepted as well. This means that this particular item is fixed and appropriate to use but when either one is accepted it suggested to be revised. The conclusion or the status of the item will be rejected if the two indexes show score beyond the score interval suggested by the two indexes.

## **5.2 Summary**

Item analysis is a complicated process since it deals with simple statistics calculation. This item analysis provides a good justification to determine whether test items constructed are acceptable or discarded.

### **Exercise**

1. Find examples of English test to be analysed critically through their simplicity, practicality, validity, reliability, and its physical look by considering criteria to do and not to do in constructing test items for particular format.
2. Construct and develop an English test items following standardized made test procedures.

## **Bibliography**

- about-us.html*. (2015, June). Retrieved June 26, 2017, from <http://www.biologimu.com/2015/06/validitas-reliabilitas-tes-asesmen.html>
- Arikunto, S. (1997). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Asari, S. (2004). *The Differences of Students' Reading Comprehension across Different Types of Linguistic Simplification*. . Malang: Unpublished Thesis.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Longman.
- Diederich, P. B. (1973). Short-Cut Statistics For Teacher Made-Test. *ERIC*, 10.
- Djiwandono, M. S. (1996). *Tes Bahasa Dalam Pengajaran* . Bandung: ITB.
- Djiwandono, M. S. (2008). *Tes Bahasa: Pegangan Bagi Pengajar Bahasa*. Jakarta: PT Index.
- Eduka, T. M. (2016). *Strategi & Kupas Tuntas SKL UN 2016 SMA-MA IPA*. Solo: GentaSmart.
- Fulcher, G. &. (2007). *Language Testing and Assessment. An advanced resource book*. New York: Routledge Taylor & Francis Group.
- Heaton, J. B. (1988). *Writing English Language Tests*. London: Longman.



- Jaya, H. (2013, June 22). Retrieved September 11, 2017, from <https://hanriansyahjaya.wordpress.com/reabilitas-inter-refer>
- Komala, R,D and Sari N,N (2012), From <https://thejoyoflanguageassessment.wordpress.com/2012/12/19/approaches-in-language-testing-2/>
- Lutan, R. (2000). *Pengukuran dan Evaluasi Penjaskes*. Jakarta: Departemen Pendidikan Nasional.
- Madsen, H. S. (1983). *Techniques in Testing*. Hongkong: Oxford University Press.
- Malamed, C. (2010, July 20). Retrieved August 23, 2017, from [http://thelearningcoach.com/elearning\\_design/writing-g-matching-test-items/Tips For Writing Matching Format Test Items](http://thelearningcoach.com/elearning_design/writing-g-matching-test-items/Tips For Writing Matching Format Test Items)
- Mathe, E (2014), From (<https://www.englishlearner.com/intermediate/cloze-1.shtml>)
- Nurham, H. (2013, May 2). Retrieved October 26, 2017, from <https://hamimnurham.wordpress.com/2013/05/02/pengertian-validitas-dan-jenis-jenis-validitas/>
- Oller Jr, J. W. (1983). *Issues in Language Testing*. Mass.: Newbury House Publication.
- Philips, A. D. (1979). *Measurement and Evaluation in physical education*. Canada: John Wiley & Sons, Inc.
- Shohamy, E. (2004) . *Assessment in multicultural societies: Applying democratic principles and practices to*

language testing. in B. Norton & K. Toohey (Eds.),  
Critical pedagogies and language learning (pp. 72-92).  
cambridge, UK: Cambridge University Press

Suwarto. (2007). Tes Menurut Teori Tes Klasik. JILID 16  
NOMOR 2. *JURNAL PENDIDIKAN*.

Wingglesworth, G. (2008). Task and performance based  
assessment. In E. Shohamy & N. H. Hornberger (Vol.  
Ed.), *Encyclopedia of language testing and assessment*  
(pp. 111-122). New York, NY: Springer

## Biography



**Slamet Asari** was born on August 20, 1970 in Pasuruan, East Java Indonesia. He is presently studying English Language Teaching for his Doctoral Degree at State University of Surabaya funded by Directorate Higher Education. He accomplished his master degree (S-2) in English Education particularly on the area of English Language Teaching in 2003 at State University of Malang. Long before achieving his master degree, he has achieved his strata 1 (S-1) in 1993 at English Department University of Muhammadiyah Malang.

His long experience in teaching English at University of Muhammadiyah Gresik brought him into several national or international academic events. Participating actively as speaker for International Conference on Lesson Study in several universities and cities in Indonesia is among things he has concerned with. He also had been given an opportunity to visit some countries to improve his teaching expertise.

Currently, he teaches some courses such as Language Testing, Authentic or Alternative Assessment, Academic Writing, Microteaching, and English for Specific Purposes (ESP) for undergraduate students at English Department of University of Muhammadiyah Gresik.

This book is his fourth book he has written for students of undergraduate level. Books such as 'Reading for Meaningful Passages', 'English for Law 1 and 2', and 'English for Psychology' are his previous publications.