

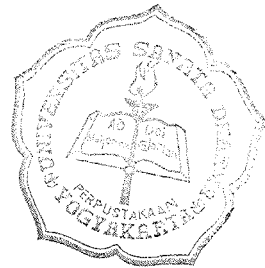
ENGLISH LANGUAGE TESTING  
AT  
THE SECONDARY SCHOOL IN INDONESIA

A Thesis  
Presented to  
The Department of English  
"SANATA DHARMA"  
Teachers Training Institute

In Partial Fulfilment  
of the Requirements of the  
Sarjana Degree

by  
Jenny Soegiarto

2002



Yogyakarta

March, 1973.

IV  
141  
58

Approved

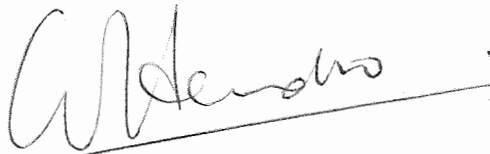
by



I.

Mr. Soepomo Ph.D.

and



II.

Drs. V.J. Hendroarsito .

The Head

The Dept. of English

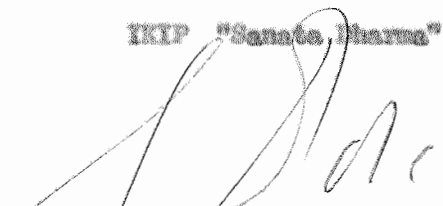
IKIP "Sanata Dharma"



( Mr. Soepomo Ph.D. )

The Rector

IKIP "Sanata Dharma"



( Drs. J.I.G.M. Brost S.J. )

ACKNOWLEDGEMENT .

I am deeply grateful to those who have read and constructively criticized my final draft and I wish to mention their names here so that we shall never forget them.

Dr. Soepomo , Head of the English Departement I.K.I.P. Sanata Dharma, who has given me invaluable support in the completion of my thesis.

Drs. V.J. Handrowarito , Lecturer at the Faculty of Letters, Gajah Mada University and Lecturer at the English Departement I.K.I.P. Sanata Dharma , who has been very helpful and cooperative in the reading of the paper.

I am particularly indebted to Dr. M.Murphy , British Council Officer at the English Departement I.K.I.P. Bandung , who has made suggestions on content and style throughout the entire manuscript.

I also express my deep appreciation to my husband , for his continual support and encouragement during the collection of data and the writing of this paper.

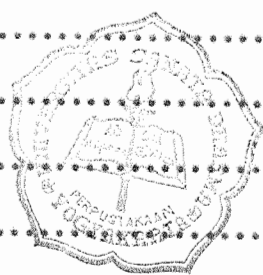
May I finally express my sincere gratitude to the entire staff of I.K.I.P. Sanata Dharma , especially the Rev. Drs. J.E. G. M . Brust , S.J. and the Rev. Drs. E.E. Bolsius , S.J. for the wonderful time I enjoyed during my study.

Jenny Soegiarto .

\* \* \* \* \*

CONTENTS.

<u>CHAPTER :</u>	<u>PAGE :</u>
ACKNOWLEDGEMENTS .....	iii
INTRODUCTION .....	1
I . RELATION OF LANGUAGE TESTING TO LANGUAGE LEARNING. ..	2
✓1. What is a Test ? .....	4
✓- What is the Purpose of tests ? .....	4
✓ a. Measuring device .....	4
✓ b. Teaching aid .....	5
✓2. Language Testing .....	6
✓a. The Process of Foreign Language Learning ..	6
✓b. What Problems are tested ? .....	7
1. Interference .....	7
2. Inaccurate habits .....	8
3. False Analogy .....	8
✓3. Characteristics of a Good Test .....	10
✓ a. Validity .....	10
✓ b. Reliability .....	11
✓ c. Appropriateness .....	11
✓ d. Objectivity .....	11
✓ e. Adequacy .....	11
✓ f. Administrability .....	12
II . TYPES OF TESTS .....	13
A <sub>1</sub> . Standardized Tests .....	13
- Definition and Nature .....	13
A <sub>2</sub> . Teacher - Made Test .....	16
B <sub>1</sub> . Proficiency + Aptitude tests .....	17
B <sub>2</sub> . Achievement + Attainment tests .....	17
B <sub>3</sub> . Diagnostic tests .....	18





	v
✓ C <sub>1</sub> . Subjective Tests .....	18
✓ - Definition .....	18
✓ - Advantages .....	19
1. Oral test .....	19
2. Essay - Type test .....	19
- Testing Higher Levels of Reasoning .....	20
- Influencing Study Procedures .....	21
✓ - Disadvantages .....	21
* Scoring .....	21
* Low Reliability .....	22
= Scorer Unreliability .....	22
= Content Unreliability .....	23
C <sub>2</sub> . Objective Type Tests .....	23
a. Definition and Nature .....	23
b. Advantages .....	25
1. Testing Recognition .....	25
2. Influencing Teacher's Strategies .....	26
3. High Scorability .....	27
4. High Reliability .....	27
a. Scorer Reliability .....	27
b. Content Reliability .....	27
c. Disadvantages .....	28
* Difficult to construct .....	28
✓ Subjective vs. Objective Tests .....	28
✓ - Differences between the two .....	
✓ 1. The Degree of Objectivity Involved in Marking .....	29
2. Length of the Student's Response .....	31
3. Summary .....	32
III . OBJECTIVE TEST ITEMS .....	34
- Outline of Objective Type Tests .....	34

A. Supply Type .....	35
1. The Direct Question .....	35
2. Completion .....	36
a. Open Ended .....	
b. Internal Blanks .....	
B. Selection Type .....	36
1. True - False .....	36
a. True-False .....	37
b. Double Choice .....	38
2. Multiple Choice .....	39
a. Correct Answer .....	40
b. 1. Positive Type .....	40
2. Negative Type .....	40
b. Incomplete Statements .....	40
c. Best Answer .....	41
d. Substitution .....	41
e. Multiple Response .....	42
3. Matching .....	42
a. Simple Matching .....	43
b. Compound Matching .....	43
4. Rearrangement or Rank Order .....	45
5. Interpretive .....	46
6. Other Miscellaneous Forms .....	47
IV . TEST CONSTRUCTION .....	49
✓ A. Planning the Test .....	49
✓ 1. Purpose of Test .....	50
✓ 2. Specific Aspects of the Course Content to be Tested .....	51
✓ 3. Age or Grade Level for Which the Test is Meant .....	53
✓ 4. Desired Range and Distribution of scores	

for the Test .....	53
5. Type or Types of Objective Items to be Used.	54
6. Form of Administration and the Time Limit Desired for the Final Test .....	54
B. Writing the Test .....	55
1. The Numbers of Items for the Try-Out .....	55
2. The Inclusion of Optional Items .....	55
3. Direction for the Test .....	56
4. The Need for Criticism from Others .....	57
5. The Form and Nature of Test Items .....	58
a. Types of Questions .....	58
b. Independence of Test Items .....	59
c. The Language Used .....	60
1. Clear, Brief and Economical Language..	60
2. Avoiding Improper Contest .....	62
3. Avoid Giving Away the Answer .....	63
a. Grammatical Form as Specific Deter- miners .....	63
b. Words or Groups of Words as Specif- if Determiners .....	63
d. The Distractors .....	64
- Definition .....	64
1. Number of Distractors .....	65
2. Plausibility of Distractors .....	65
3. Placement of Distractors .....	65
6. The Made Up of Tests .....	66
✓ C. Trying Out the Test .....	67
1. The Students Used .....	68
2. The Administration of the Try-Out .....	68
a. The Conditions Favoured .....	68
b. The Time Allowed .....	68

c. The Scoring of the Try-Out .....	69
✓ D. Analyzing the Test .....	70
V . THE INTERPRETATION OF TEST SCORES .....	72
* Statistical Concepts .....	72
1. The Frequency Distribution .....	72
a. Normal Curve .....	74
b. Skewed Curves .....	74
2. Variability .....	74
- Definition .....	74
a. Range .....	75
b. Standard Deviation .....	75
3. Central Tendency .....	76
a. Mean .....	77
b. Median .....	77
c. Mode .....	78
* Scores .....	78
1. Definition .....	78
2. Types .....	78
a. Raw Score .....	78
b. Derived Scores .....	79
3. Interpretation of Scores .....	80
a. Scores Related to an Absolute Standard ..	81
b. Scores Related to an Individual Standard .	82
c. Scores Related to a Relative Standard ..	82
d. Norms .....	83
- How to Obtain Norms .....	84
a. Content Norms .....	84
b. Age Norms .....	85
c. Grade Norms .....	85
e. Rank Norms .....	86
e. From Raw Scores to Marks .....	86

	ix
= Grading on the Curve .....	86
VI . CONCLUSION .....	88
BIBLIOGRAPHY .....	90

\* \* \* \* \*

## INTRODUCTION

Concerned with the professional status of the English teachers at our Senior High School in our area, the writer would like to present this paper as to make an attempt to provide her colleagues with some assistance in their daily work, particularly in the field of testing.

One of my duties as the Head of the Secondary School Teachers' Advisory Council (4Team Pembina Bahasa Inggris ) is to observe the classroom teaching of teachers of English in the Yogyakarta area. During these observations then, I picked out one or two of their problems revealed during the class periods.

Indeed, teachers of English, as a foreign or second language, face numerous problems. These problems are even greater, when the language is completely foreign to the teacher as well as to the student. We have to admit that not all of the teachers of English at our secondary schools are capable or competent to do the job, due to various reasons.

In my experience of observing the classroom procedures, I have found out various problems to be particularly common, urgent and serious. One of these difficulties is the problem of testing.

The truth is that most secondary school teachers teaching English do not have the necessary background for testing. They are not familiar with testing at all. What they do, if they are testing is just basing things on previous experience acquired or observed when they themselves were students. Therefore the tests they construct and the evaluation they make are not based on sound principles of testing. Such tests, therefore, are neither valid nor reliable as a measure of achievement in English.

Some S.M.A. teachers may also have had a separate course in tests and measurements during their study or at an upgrading or re-

fresh course, however, many such special courses are usually not geared to producing competence in test construction, but more often treat commercially available tests that may prove valuable or suitable in an educational setting. Thus the teacher often completes his educational program with little understanding of testing principles and practices.

Books on testing are indeed numerous, but again a great deal of academic ability is needed to get hold of their contents. Besides, these books are sometimes a bit too sophisticated, since they are designed for higher levels of study and various fields of employment.

In schools, standardized tests are far less common than those made by the teacher himself. These teacher-made tests, upon which students' marks are usually based, often reveal serious shortcomings. Their faults may arise from such causes as inadequate consideration by the teacher of the objectives of instruction, technical defects in construction or teacher's inability to adjust the test to the situational appropriateness.

This paper, therefore is intended to help these teachers in their work. Secondary school teachers in Indonesia will find this brief treatment of language testing helpful, particularly in providing their students the appropriate test on many different occasions. The writer would like so much, that this brief explanation on testing will answer questions like :

What is a test ?

What is the purpose of giving a test ?

What are the basic principles for constructing good test items ?

What kinds of test should we know ?

To what extent can a test be used not merely as a measuring device but perhaps more important as teaching

devices?

Is the true-false form likely to be as good as the multiple choice ?

To get the maximum out of it, this paper is deliberately made simple, meaning the language used is simple as well as the definitions and explanations.

It is the writer's sincere wish, that her colleagues at the secondary schools can find the right answer to various questions in the field of testing, after being acquainted with this paper.

We hope so !

\*\*\*\*\*



## CHAPTER I

### RELATION OF LANGUAGE TESTING TO LANGUAGE LEARNING

Any kind of learning cannot be separated from testing. How can we know that we have come to our goal? How can we know that we have achieved our objectives? These things can only be investigated through testing. In our case, since we are teaching English, and since our students have to learn English, it is that language that we must test.

English is a foreign language to us, and since foreign language learning is completely different from the learning of the native tongue, consequently the testing of the two languages will also be completely different.

#### 1. What is a test?

A test is a device to measure or find out what we want to measure or find out.

What is the purpose of tests? Tests have a dual purpose:

a. Measuring device.

b. Teaching aid.

#### a. Measuring device.

In relation to Language Testing, a test is to find out the language proficiency of the students. As a measuring device a test is used for predictive and diagnostic purposes. Achievement and Aptitude Tests are sample of predictors.

Although the distinction between aptitude <sup>kecakapan / kemampuan</sup> and achievement tests is by no means as definite as has often been supposed, the emphasis in the case of aptitude tests has been more on potential abilities, as for future learning or job performance, and less upon the current abilities or specific knowledge.

When either aptitude or achievement test are used in educa-

tional and vocational counseling, the counselor is assisting the student in making predictions.

In the case of a diagnostic test the emphasis is upon the patterning of relative strengths and weaknesses or of peaks and gaps in abilities. The use of tests for diagnostic purposes typically occurs in an educational setting, where the immediate aim is to spot areas in which teaching methods may need to be changed.

Further discussion on the usages of tests will be given in another chapter.

#### ✓ D. Tests as teaching aid .

Another significant use of tests in the classroom, is as a direct aid in motivation and thus in teaching. The most obvious aspect of this use of testing is that students, who know they are to be tested will often do more studying and hence learn more than would otherwise be the case ( 24, p. 6 ).

Studying very hard merely for the purpose of passing a test with no intent to remember beyond it, is poor practice. Nevertheless, even this way of learning undoubtedly result in more knowledge than would no studying at all. In addition, although the test may be poor, the student's regular preparation for periodic tests is often a positive motivating force.

Probably, a more fruitful way for instructional purposes is to ensure that learning takes place either at the very time the test is given or very soon thereafter. A teacher who fails to take advantage of this method of teaching is ignoring one of the most fruitful teaching devices. Too often, a test is given, perhaps in essay form and returned to the students several days or week later, with numerical marks assigned to

each question, but without a clear explanation of what would be the best answer to the question. Under such a system the student has lost his motivation by the time he gets some vague impression from the teacher of the answer that were expected from him. When a student learns immediately which items he has had trouble with and the correct responses, he is much more likely to correct his mistakes than if he must wait several days or perhaps never be informed of his specific errors.

Thus, in testing, teacher as well as students, each have their share, meaning that the students have to prepare themselves by way of studying hard and the teachers have to correct the testpapers as soon as possible, so that discussions can promptly be made on the students' mistakes.

Although there are other purposes of tests, e.g. in the field of testing interests, attitude and other personality characteristics, these are specialized fields, with which our teachers of English at the secondary schools are not directly concerned with.

## 2. Language Testing .

### a. The Process of foreign Language Learning .

Language learning is the forming of habits. In learning a foreign language, the basic patterns of sound, of word-formation, of sentence construction must be learned. Several problems are encountered during this process of learning, especially dealing with habits already established by the native tongue.

Foreign language learning is transferring the habit system of his native language and sentence patterns of his native language, in short, the entire structure of his native tongue is transferred.

The learning problems turn out to be matters of form, meaning, distribution or a combination of these. They can be predicted and described in most cases by a linguistic comparison of the two language structures.

The theory assumes that testing control of the problems is testing control of the language. Problems are those units and patterns that do not have a counterpart in the native language or that do have counterparts with structurally different distribution or meaning. The theory assumes also that the student does not know these units and patterns that are problems, unless he can use them at normal conversational or reading speed in linguistically valid situation, that is situations that parallel those language in use ( 15, p. 24 ).

Lists of words are not in themselves valid situations ; stating grammatical rules is not either, because native speakers are able to speak the language, but do not know to state the grammatical rules, while other people are able to state these rules, but cannot use them.

b. What problems are tested ?

In testing Second or Foreign Language Learning, we have to take the following problems into consideration :

1. Interference from the Patterns of the mother tongue ( Phonemic difficulties ).
2. Inaccurate habits learned at an earlier stage of education.
3. False Analogy.

1. Interference .

English is completely foreign to our Indonesian students. The entire structure of the two languages, English and Ba

hasa Indonesia, is completely different. On top of that, students at our secondary schools have the greatest difficulty in producing the correct pronunciation of the English phonemes, since the two languages differ also in their phonemic production (Compare the English stops and alveolars with the Indonesian ones).

*different  
phonemes  
(from 5/15)*

## 2. Inaccurate habits .

So many S.M.P. teachers are totally unqualified and can hardly speak good English. In teaching reading and conversation (pronunciation) their students will automatically take over the wrong or inexact pronunciation of their teacher.

*from Vietnam  
assistant  
Indon teacher*

## 3. False Analogy .

All too often students at the S.M.P. are drilled for too long in the meaningless patterns like 'I am sitting', 'I am standing', etc., so that it becomes natural for them to say 'I am having', 'I am hearing', 'I am seeing', etc.

Besides the problems mentioned above, we have also to decide which part of the language we are going to test. Are we going to test the elements or the skills ?

Elements of a language are sounds, words, arrangement or patterns of words as manifested in pronunciation, intonation, lexis (meaning) or syntax (grammar). Elements never occur separately in a language, they appear as elements in one of the four integrated skills of language, e.g. :

1. Speaking.
2. Understanding.
3. Reading.
4. Writing.

Taking this division into consideration, we make our choice.

Which do we test ? The elements, skills or both ? Needless to say, that the nature of the choice will depend on what the purpose of the test is.

At the Secondary schools we test both, elements as well as skills, for the objective is to measure the students' ability. It is indeed wisest to test both elements and skills, as the testing of one alone will not give a very complete picture. Needless to say, that it is necessary to test the elements and skills separately, as a satisfactory test of a skill can never include a sufficiently wide sample of the elements. A composition, for example, cannot possibly test fully the extent of a student's control of structures, partly because there will be no opportunity of using certain structures, and partly because the wise student will simply avoid the structures he is uncertain of.

Another example is, when a student is able to pronounce the words in the right way, but as soon as the words are arranged into sentence, this same student is not able to read the sentence properly.

Prof. Lado's test theory says that a second/foreign language learner has language learning problems and these problems have to be found out and to be tested afterwards. The problems to Lado are chiefly major points of contrast between a learner's native tongue and his target language.

Both learning and language are involved. If a test is constructed for a single group of students with identical language background and identical ability to learn the target language, then constructive analysis is essential. But most students have varying abilities which effect comparable exposures. Contrastive analysis is not practicable when a teacher or institution is faced with a group

of foreign learners from a heterogeneous native language background. Lado's theory is thus linguistically based. The result is that Lado's own proficiency test does not follow his theory : as he says himself :

"the task of preparing separate tests for all language background is so enormous that we may never hope to have such tests except for a limited few languages" ( 7, p. 13 ).

### 3. The Characteristics of a Good test.

We know the purpose of a test, we also know what we must test, and the next question is now :

What Sort of Test should we use ?

To obtain maximum benefit of a test, the test should be good. What are the characteristics of a good test ? There are several factors to determine the value of a test :

#### A. Good Test must be :

- a. Valid.
- b. Reliable.
- c. Appropriate.
- d. Objective.
- e. Adequate.
- f. Administrable.

#### a. Validity.

A test is valid, when it measures satisfactorily what is intended to measure.

Validity is not general but specific. If a test on pronunciation measures pronunciation and nothing else, it is a valid test on pronunciation, as it would not be a valid test of grammar or vocabulary ( 15, p. 30 ). The teacher cannot give a test to the first grade students on sophisticated patterns.

You cannot test an S.M.A. student by asking to write an essay on the outer space craft and how it works. ×

b. Reliability.

1. Mark-and Re-mark ability .

A test is reliable if atleast which has already been marked by one examiner is subsequently remarked with substantially the same result, either by the same examiner or by a second examiner.

2. Test-and Re-test ability.

If administered one and the same group of students to on more than one occasion, with no instruction intervening, the obtained results are substantially the same.

c. Appropriateness .

Appropriate is suitable. A test should be suitable. Appropriateness is closely connected with validity. A test should measure what we are going to measure. So before we give a test, we have to ask ourselves what we are going to test. If we want to find out the students' mastery of structures then we test their ability in structures. And if we want to find out their aural comprehension, then we test their ability of comprehending aurally.

d. Objectivity.

A test is objective if whenever, wherever the test is given and whoever gives the test, it will give the same result.

e. Adequacy.

A test is considered adequate when the test represents all the lessons taught in proportion to their importance. So a test should form a sample of the lessons taught.



f. Administrability .

A test must be easy to score and easy to administer. The ease of scoring and administration is an important point to be considered, i.e. whether the form of test used or proposed can in fact be administered in prevailing conditions and by the available personnel.

An oral test which required trained phoneticians for example is not a practical possibility in administrative terms.

The characteristics of a good test mentioned above are just a brief explanation of the way how to evaluate a good test. In the chapter on types of tests, these points will be further discussed .

\* \* \* \* \*

B<sub>3</sub>. Diagnosis, is related entirely to the use made of the information and not at all to the presence of skill in the learner. We can speak of a learner's Achievement, Proficiency or aptitude but not, of course, of his diagnosis. Diagnosis is the teacher's concern for what has gone wrong, it is a test for the teacher's own consumption. The information obtained from achievement, proficiency and aptitude tests are used by the teacher to make his diagnosis as regards the result of his teaching. Has he used the right method? Must he change his way of presentation of certain materials? By means of this diagnosis then the teacher can decide for himself what will be the most appropriate way or method in presenting the various materials he is teaching. A diagnostic test may be additionally used on achievement and proficiency tests ( 7, p. 7 + 8 ).

C. Another classification of tests, according to the scoring is

1. Subjective Type Tests.
2. Objective Type Tests.

As stated above, these terms are used to indicate two types or scoring. Objective tests are those that are scored rather mechanically without need to evaluate complex performance on a scale. Subjective tests are those that require an opinion, a judgement on the part of the examiner. In subjective tests there are differences in scoring by different examiners, hence the name subjective ( 15, p. 29 ).

\* C<sub>1</sub>. Subjective Tests.

Definition : Subjective Tests are those which require an opinion or judgement on the part of the examiner. There is a scale in correctness. If an item is correct, how correct is it? These tests consists of questions only, while the students

have to compose the answer. The subjective test follows a procedure, in which the testee answers, in his own words, all or some of a relatively small number of questions. Typical keywords in the questions set in examinations of this kind are : 'Discuss', 'Compare', 'Describe', etc.

Assessment of the examinee's work is subjective, in the sense that its "correctness" has to be evaluated or judged by the examiner.

The number of questions is usually small, and the responses may be written in whatever way the students think appropriate. The choice of words and organization of thoughts are left to the preference of the students, so that answers may vary from a very short statement to a long winding paragraph.

#### Advantages of the subjective tests.

Subjective type tests can be given in the form of an :

1. oral test (conversation between examiner and testee).
2. essay-type test.

#### 1. Oral test.

Again here it is entirely left to the discretion of the examiner what questions he is going to ask his testee and how he is going to score the answer. <sup>⑤</sup> Since this kind of test is not often taken at the S.M.A. level, no further discussion on the topic are included in this chapter.

#### 2. Essay - Type test.

<sup>⑥</sup> Advocates of the essay test claim that it is impossible or at least very difficult to elicit creative behaviour on the part of the testee in completely objective tests. Some people are of the opinion that the serious weakness of objective tests is that they elicit "mere recognition". The essay test

is said to be preferable because it forces the testee to recall information.

Several early experiments were concerned with this question. In one recall testing experiment the same material was tested by two various objective type tests; the multiple choice type-which requires recognition ability more than recall ability and the completion type, which stresses recall ability. The significant finding was that scores from the two forms were almost perfectly correlated. The scores were indeed somewhat higher for the recognition type ( 24, p. 94 ).

This leads to the conclusion, that subjective tests test recall rather than mere possession of knowledge. For knowledge to have value, one must be able to recall it when needed. (F (mully test))

#### Testing Higher Levels of Reasoning.

It is claimed that the essay-type tests can test higher level processes, such as required in the organization of ideas, comparison, contrast, judgement, imagination and inference. They test the ability of the student to interpret and organize what has been learned and to select that which is relevant. They require not only the ability to memorize details, but also the ability to see meanings in their relationships to other things and to apply principles to situation, calling forth some new application of a familiar principle. *situation*

"The person who is able to utilize information acquired in one situation and who applies it to the intelligent solution of problems in a new setting has arrived at true mastery" (19, p. 166). *exercise test*

Elements of knowledge form essential ingredients for thought and Reasoning. Reasoning as well as other forms of high level mental behaviour does not take place in a vacuum ( 24, p. 96 ). One, who does not possess the knowledge of, for instance, the principles of second language teaching, obviously does not *inspiration*

reason in this field.

The essay-type test allows the student to organize his ideas and present them in his own words, if he prefers, rather than parroting an answer that comes from another source. In this feature of the essay test lies the possibility of evoking creative behaviour and therein lies the value to support it.

However, the essay test is only valuable in courses where the student's ability to express himself effectively is the major objective of instruction.

#### Influencing Study Procedures .

The type of measurement used by the teacher will influence the type of the study techniques of the students. When students expect that the test will be of the essay type, they will very likely prepare themselves for a test according, <sup>ly</sup> e.g. making out lines, summaries and other relevant things.

They will try to see scattered points in a subject matter as a whole unit. They will try to understand that one idea is not separated from the other, that a synthesis results from two theses. In short they will see ~~the~~ subject matter in a different and better light; they will not forget the main thread of the subject matter easily.

#### Disadvantages .

##### Scoring .

The greatest weakness of the essay type test is the difficulty of scoring. The scoring cannot be left to amateur scorers, experts are needed. Furthermore, careful evaluation of an essay requires some fifteen minutes or even longer. When a teacher has to mark papers from 3 or 4 classes of 40 students each, it is obvious that the total time spent becomes indeed significant. A conscientious teacher simply cannot reduce this scoring time .

### Low Reliability.

#### Scorer Ureliability.

The marker of an essay type test must not only know the material being tested, thoroughly himself, but his judgement in many cases seem to be clearly influenced by subjective factors. If the first part of the essay impresses him favourably, the rest will get a lenient judgement. But in turn, the blunders found in the beginning of the essay will definitely affect the last part of the paper.

Teachers tend to mark essay papers in terms of their own particular values and expectations. It is common that sometimes teachers cannot agree with one another regarding these values. This might be due to different standards<sup>s</sup> of marking and different weighing of the questions. Oddly enough one teacher often would not give the same mark twice to the same paper. Also when the number of papers to be marked is so great as to require marking periods separated in time, the teacher's standards may shift in one direction or the other. This tendency, so to speak, to disagree with oneself, lowers reliability.

Therefore, to minimize inconsistency, teachers should recheck the early graded papers now and then to ensure that standards have not shifted appreciably or to avoid any tendency towards non-comparable standards.

The essay test also involves many things, that are not really relevant to the content tested. Factors such as legibility or nice handwriting, spelling and grammar or the effectiveness of the organization of the ideas presented, may influence the judgement of the teacher. The importance of having something to say is underrated.

Physical as well as mental conditions of the teacher al-

so tend to condition the marking of essay papers at a particular time, or in the words of Ashburn ( 2, p. 13 ).

" ..... the passing or failing ..... depends, not on what students know or do not know, but on who reads the paper", — also "on when the papers are read".

#### Content Unreliability.

Content unreliability can be caused by poor sampling i.e. the sample may be poor, meaning too small, so that the obtained scores will depend too much upon the accidental inclusion of certain test items and the exclusion of others. Or the sample may be large, but it may fail to be representative, so that certain subject matters may be neglected or overemphasized. Thus, the score of a student will depend too much upon that particular sampling of items. So the student is the one, who suffers. He might probably have received quite a different score on some other test.

Content unreliability is again some weakness of an essay type test, because due to the nature of such a test, it can only contain a limited number of questions. However, sampling unreliability can also affect an objective test, especially if the teacher constructs his test hastily and the items do not represent the material intended to be tested. The teacher should learn from experience, the length and content of a good test and consider carefully the necessary precautions, in order to avoid basic errors.

### x C<sub>2</sub> Objective Type Tests.

#### a. Definition and Nature.

An Objective Test is a test, which consists of questions

or items, to each of which only one correct answer is possible. It is from this, of course, that their objectivity stems : if only one answer is possible, the examiner's subjective judgement can play no part in the evaluation of the response. Many objec tive tests can be marked by machine without the intervention of the human examiner at all, even the possibility of arithmetical error on the part of the marker is excluded.

In many cases objective type tests only ask for a "check" as an indication of the right answer. The questions can be answered by just a word or by indicating the response in the form of <sup>a</sup> symbol, the one of several alternatives that is believed to be the best. This characteristic of objective type tests al low a great number of questions to be put in them.

Objective type tests have master keys of answers, or they are constructed that way as to allow the use of mechanical or semi-mechanical devices. The key for an objective test can be determined in advance so that any person with reasonable degree of clerical ability can perform the actual scoring. Even when a test cannot be tried out on a group of typical subjects before it must be used in an actual testing situation, techniques exist for detecting unforeseen blunders in the scoring key. Even a person, who know nothing about the subject matter being tested, can then compare the choices on the answer sheet with those predetermined answers

The less human bias that enters the scoring procedure, the more objective the scale.

A weak objective test may indeed provide less valid results than a poorly constructed essay test. Nevertheless, a good





objective test, made carefully, may provide considerably greater assurance of validity than a well-constructed essay test.

If the essay type test does not have a place in the early stages of language instruction, the objective type certainly does. Very early in the language course, the achievement of a student can be measured by means of an objective type tests. But it should be kept in mind, that the objectivity of these tests lies only in the scoring. Beyond that, for instance in the transfer from scores into grades, is something subjective. Brownell ( 4, p.p. 481 - 8 ) points out that there are more subjective than objective factors involved in an objective type test. He says :

"Well, first of all, in the practical circumstances of teaching, one decides to give a test. The decision is surely not based upon purely objective considerations. Second, one determines whether to make a test or buy one .....

Third, one makes up one's mind regarding the kind of test — whether it is to be of the traditional type, of the new er types or a combination → judgement once more.

Fifth, one selects the items to be included, little objectivity here, — Sixth, one chooses the form to be employed — true false, multiple choice or what not — again little objectivity.

Seventh, one frames the items as carefully as one can — and once more has only his judgement for guidance. Eighth, one prepares a key by listing the correct answers — a judgement which may not be acceptable to other teachers even of the same subject. Ninth, through opinion one defines the conditions of administering the test. Tenth, one scores the papers — at last objectivity. But eleventh, one assigns marks — another increment of judgement, and a big one."

#### b. Advantages .

##### 1. Testing Recognition.

The objective Type test is apt to test efficiently

a student's knowledge of facts rather than the ability to organize and interpret what one has learned. The response is merely identified from a list of suggested answer supplied by the teacher. It involves more recognition than recall, and such a recognition test usually yields higher scores than a recall exam. Surely good multiple choice comprehension-questions will test more than just recall. Therefore, objective tests encourage students to build up a broad background of knowledge.

Though recognition is the lowest level of knowledge it is an important thing in matters like one's reading vocabulary.

## 2. Influencing Teacher's Stratagem .

From time to time, or very often at the beginning of a schoolyear or a term, a teacher may administer an objective type test to measure the achievement of his students and to interpret the results of it. If his interpretation is correct, he can learn about the range of achievement in his class.

This knowledge will enable him to locate problems and to plan basic strategy for instruction accordingly. It will help him determine at what point he should begin in developing certain skills with his class. It also will help him find material of an appropriate level of difficulty for the class as a group as well as for the individual, who may be far behind or way ahead. When properly constructed and used, objective type tests can assist the teacher very much to make the necessary adjustments in his expectations to the abilities and present strength of the class. When these expectations are not in line with the abilities of the class, then, very likely difficulties will result.

Objective tests for planning may be administered at the beginning and half way through the term. The latter tests will also indicate progress achieved during that period.

### 3. High Scorability.

This type of test is quick and rather easy to score objectively. The scoring and interpretation can be done in the same way by different teachers or test administrators. Not only can the scoring achieve a high degree of accuracy and consistency, but it can also be accomplished in a relatively short time. Objective type testpapers or their corresponding answer sheets can be returned to the students, before much time has passed. This will enable students to know exactly what answer was intended for each question. They will also see the item with which they had difficulty. They are more likely to correct their mistakes in this way than when they have to wait several days for the tests results. Sometimes, and too often, the students are not informed at all of their specific errors.

The fact that students can be informed of the test results in a short time, will have important motivational implications.

### 4. High Reliability.

#### a. Scorer Reliability.

The teacher who marks objective type tests can do his work with high accuracy and great consistency, as the items are scored only right or wrong. Irrelevant factors, such as legibility of handwriting and conditions of the teacher are of no influence at all. And he need not be an expert in the field himself. Anyone with a little bit of clerical ability can correct objective tests with the help of a predetermined key, the masterkey.

Inconsistencies, that may occur, either within one scorer from one occasion to another, or among different scorers, can be controlled almost completely.

#### b. Content Reliability.

Objective type tests demand only very brief answers,

so that each of them may cover many items. Therefore, objective type tests are very often long tests and long tests are usually considered more reliable than short ones. The large sampling afforded by these tests tend to give good reliability.

In contrast to the subjective type test, the objective type can sample the content of the subject matter more adequately. An essay test of one clock hour may contain only three or four random and broad questions, not explicitly constructed, whereas the objective test may have as many as 80 or 100 questions, allowing it to have a more complete and representative coverage of situations permitted by the same testing period.

#### c. Disadvantages .

##### 1. Difficult to construct .

Objective testing indeed requires highly skilled personnel. Many things must be considered; the type of test that will measure a certain purpose best, the diction of each item, etc. Skill, time and experience are needed for the construction of an objective test as we shall see in the chapter on "Test Construction".

#### Subjective vs Objective tests .

There is no fixed rule or criterion as to when to use a certain type of test. However, having in mind the characteristics of each type will be of great help for the teacher who plans to make his own test.

However, it should be noted that a suitable test in one place may be not suitable in some other place, especially when learning problems of one place are quite different from problems in the other (Africa-China). Since we are mainly concerned

with tests at the secondary schools in Indonesia, learning problems of that kind will not occur. Yet learning problems of the rural areas may differ from urban areas, due to the different environment and conditions, incapable teachers are more frequently found in the outback areas rather than in the cities.

The remoter the place, the worse the conditions as regards building, teaching aids and other facilities.

For certain aspects of language the objective type questions are often more appropriate. A student's knowledge of a particular topic can be decided more explicitly by a series of objective questions than by an essay in which the relevant facts have to be located and counted.

#### 1. The Degree of Objectivity Involved in Marking.

In an objective type test, there is one correct answer to each question with a predetermined mark for each question. In an essay type test, there is no completely right or wrong answer and the quality of the essay must be judged subjectively by an expert and marked with a subjective evaluation.

When considered as measuring instruments, objective type tests compare favourable with essay tests in some respects and unfavourably in others. The score assigned to a subjective test of the completely objective kind is definite, apart from clerical errors, regardless who marks the test paper. In other words, reader reliability for this type of test is perfect.

In subjective type tests, there is no guarantee that a test paper would receive, the same mark if it were marked by two different teachers, or by the one and the same teacher, but on two different occasions. Differences in the marks and even in the ranking of students have been found to be great ( 17 , p. 498 ). This can be partially overcome by marking all the responses of the students

to one question first, before proceeding to the next question.

Essay type tests have an advantage over the objective type in being relatively easy to prepare. The careful preparation of objective test questions requires much time and effort. In the scoring of tests, the advantage is the reverse. Objective tests can be scored accurately by a clerical staff in a short time, but essay tests have to be marked by experts in the subject. This is a time consuming process, although this may not be too important a consideration for the classroom teacher with only 30 or 40 essay papers to mark.

The many aspects of a language course can be tested more thoroughly and explicitly by a large number of objective questions which would require the same amount of working time.

Nevertheless, which type of tests the teacher wants to employ depends entirely on the particular purpose he has in mind and which type of tests will suit best the particular set of objectives.

Essay tests would generally be more appropriate for the higher levels of learning, whereas objective tests would be adequate for testing the lower skills and for situations in which the teacher is more concerned with the product of the thinking than the process involved.

The difficulties associated with the use of each type of test can be avoided to some degree by following certain principles in setting and marking the test questions. The objectivity of objective tests can be combined with the depth of the essay tests. Care in the construction is the key factor. Special attention should be given to the wording of essay questions. Questions must be made specific, so that students will not be in doubt as to what is required.

Example : Item : Discuss the "Western Type Film" ?  
 Better : How can the "Western Type Film" be a  
 box - office success ?

or : Contrast the "Western Type Film" with the  
Musical Comedy Film ?

## 2. Length of the Student's Response .

Confronted with the first question, the slow student would most probably spend much time in deciding, where to begin, and he would not know where to end. The last two questions have more definite ideas to elicit. This will make it much easier for the teacher to make a scoring key.

If the question is too general, students are not running the same race ( 240, p. 103 ). Some students will respond to what the teacher intended to elicit, while others will provide their own different but legitimate interpretation of it. These different responses cannot be rated on the same scale and in turn the marks assigned cannot provide information as to the relative competence of the students. The use of the words such as "explain", 'contrast' are more likely to keep students on the point than the word "discuss".

Essays need not all be long. A number of essays of three or four paragraphs may serve the purpose just as well and likely to be marked reliably. Sparing use of optional questions should also improve reader reliability as teachers can compare the performance of students on common basis.

When essay papers are marked by more than one examiner, a preliminary conference of examiners discussion of sample answers to determine what qualities to look for and what weight, if any, must be attached to spelling, handwriting and punctuation, should lead to closer agreement in the marks given by those examiners. The marking of essays on the basis of general impression should be discouraged as this seems to be the least satisfactory.

It is just as important to know where to use certain types of tests as it is to know how to use them.

Summary .

The following summary will help decide the most suitable test type to serve a particular purpose in a certain testing situation.

EssayObjectiveAbilities measured :

Requires the student to express himself in his own words, using information from his own background and knowledge.

Can tap high levels of reasoning such as required in inference, organization of ideas, comparison and contrast.

Does not measure purely factual information efficiently.

Requires the student to select correct answer<sup>s</sup> from given options, or to supply an answer limited to one word or phrase.

Can tap high levels of reasoning such as required in inference organization of ideas, comparison and contrast.

Measures knowledge of facts efficiently.

Scope :

Covers only limited field of knowledge in any one test.

Essay questions take so long to answer that relatively few can be answered in a given period of time. Also the student who is especially fluent, can often avoid discussing points of which he is unsure.

Covers broad field of knowledge in one test. Since objective questions may be answered quickly, one test may contain many questions.

A broad coverage helps provide reliable measurement.

Incentive to Pupils :

Encourages pupils to learn how

Encourages pupils to build



to organize their own ideas and express the effectively. up a broad background of knowledge and abilities.

Ease of Preparation :

<p>Requires writing only a few questions for a test. Tests must be clearly defined, general enough to offer some leeway, specific enough to set limits.</p>	<p>Requires writing many questions for a test. Wording must avoid ambiguities and "give a-ways". Distractors should embody most likely misconceptions.</p>
---	--

Scoring :

<p>Usually very time consuming to score.</p> <p>Permits teachers to comment directly on the reasoning processes of individual pupils.</p> <p>However and answer may be scored differently by different teachers or by the same teacher at different times.</p>	<p>Can be scored quickly.</p> <p>Answers<sup>are</sup> generally scored only right or wrong, but scoring is very accurate and consistent.</p>
--	---

Illustrations of objective test items will be discussed in the next chapter. It must be remembered, that the examples presented, offer only some of the possibilities. For instance, the wording of the directions may be different depending on the type of answer-sheet used, or on the format of the test-paper itself if no separate answer-sheets used.

\* \* \* \* \*

## CHAPTER III

### OBJECTIVE TEST ITEMS

Certain type of Objective Tests are familiar to the average language teacher and they are widely used in language-testing. However teachers must be aware of the different characteristics of each type. Objective tests can be classified into two major groups :

1. Supply Type.
2. Selection Type ;

according to whether students have to supply the words or phrases for the answer or to select their answer from given alternatives. The first is also known as the

- a. Recall Type - whereas the latter the
- b. Recognition Type .

#### Outline of Objective Type Tests .

The main types of objective tests used by the classroom teacher or may be listed as follows :

#### A. Supply / Recall Type .

1. The Direct Question .
2. Completion .

- a. Open Ended
- b. Internal Blanks

#### B. Selection / Recognition Type :

1. True - False
  - a. True - False
  - b. Double Choice
2. Multiple Choice
  - a. Correct Answer
    - (1) The Positive Type
    - (2) The Negative Type
  - b. Incomplete Statements

- c. Best Answer
- d. Substitution
- e. Multiple Response

### 3. Matching :

- a. Simple Matching
- b. Compound Matching

### 4. Rearrangement of Rank Order.

### 5. Interpretive.

### 6. Other Miscellaneous Forms.

## A. Supply Type .

A Supply Type test demands answers which must be recalled from a student's past experience. Usually only short answers are required, sometimes one word or a short phrase. Hence it is often referred to as "Short-Answer" Test, grouped under objective tests rather than subjective tests.

A Supply type test virtually eliminates the possibility of guessing desirable study practices. A limitation of this type is the fact that it tests facts consisting of isolated bits of information and the scoring is not completely objective. Thus it cannot be turned over objective to the lay reader and it will surely take more time than the other objective forms.

The Supply Type may take two forms :

1. Direct Question
2. Completion .

### 1. Direct Question .

This test type has direct questions as its items. A direct question is usually preferable to the Statement Form or the Completion Type, because it is more natural and is likely to be easier to phrase. In addition to its ordinary use, such

a test can also be used in situations presented in the form of passages, pictures maps or diagrams, therefore testing visual comprehension.

Example :

Directions : Answer briefly

Items : a. Is the "melati" a native flower ?

b. Why is English taught, and not Hindi ?

## 2. Completion .

This is one of the oldest of the new tests, dating back to 1897 (C. J. p. 145). The Completion Type has a series of sentences, from which certain important words or phrases are omitted. The blanks may be at the end of the sentence ( the open --- ended type ) or within the sentence ( the internal blank type). One blank counts for one point. The statements may be organized into one short paragraph, or may consist of a number of unrelated sentences. The students must fill in the blanks with the appropriate words or phrases.

A Completion test must be prepared with great care, otherwise it will test rote memory rather than real understanding.

Example :

Directions : Fill in the blanks with one word !

Items : a. He has     his bike to get additional money.

b. Ice is cold and sugar is    

## B. Selection Type .

Among selection Type questions, the True-False and the Multiple Choice are most common. The Rearrangement, Identification and Incorrect forms are less common.

### 1. True - False Form .

This test has items that fall under two categories, two

possible responses only : true-false / right-wrong / yes-no / same-opposite / - or the double choice. Probably the most known and widely practiced are :

a. True - False.

b. Double Choice

They are easy to construct and can cover a wide sampling. They can also be applied to a wide range of subject matter.

a. True - False.

The negative suggestion effect and guessing are too often pointed out as weakpoints. To minimize the change gap which results from these two effects, a correction formula is often used. Specific clues must be avoided. Statements with "only", are often false, whereas the use of "should" may indicate something true. Also the tendency of the teacher to take the statement directly from a textbook or with just slight changes, will encourage rote memorization.

It is said that with young students, questions eliciting the yes-no response are preferable to declarative statements ( 19, p. 174 ).

Such questions however, are seldom found in standardized tests, which shows the low regard, in which they are held by test-experts.

To claim reliability a true-false question must be made longer than other objective types. But it is highly recommended for cases where it is difficult to construct more than two responses for a multiple choice test.

True-False items should only be used with statements which are unequivocally true or false. Even when they are generally less acceptable than multiple choice items for which it is less likely that the correct answer will be se-

lected by chance.

Example :

Directions : Mark a T for true if you think the statement is true or F for false if it is not.

- Items :
1. A zoo is a place, where various animals are kept for exhibition.
  2. Repeating a sentence many times will lead to automatic habit mastery.
  3. A pessimist is a person who sees the bright side of things.

b. Double choice .

The Double<sup>choice</sup> is by nature a variant of the Multiple Choice Type in which one item offers but two alternatives, one right and the other wrong. Thus it has only one distractor to offer.

In the construction of Multiple Choice items with more than two alternatives, it is often found that the greatest difficulty is in devising proper distractors. It is relatively easy to find a correct alternative and just one distractor that is really tempting. Usually the other distractors are not only more difficult to devise, but also easily discarded by students.

One advantage over the "conventional" "Multiple Choice" items is that it takes less time and less energy to construct.

From the students' point of view when graded tests are returned for review, there is no need to read over the entire test, indicating correct or incorrect answers. The grading itself indicated which of two choices in each item is correct and which incorrect. Class attention could, therefore, be focused on specific items.

One might argue, that though the double choice test lends itself to easy construction and is in many ways superior to the ordi

nary multiple-choice items, it is less reliable. But it seems likely that the Double Choice is at least adequate and there is no obvious reason why it should not be included in a testing program. This is

Example :

Directions : Cross out the wrong alternatives between brackets.

- Items :
1. Fire-wood is (fire, wood).
  2. It took Tony and (I, me) a long time to find out our seats.
  3. Rita received a ( compliment, complement ) on her excellent performance.

2. Multiple Choice.

The Multiple Choice Type is one in which the items contain several suggested alternatives, of which the best or right one is to be selected. The correct answer need not be the only disputably correct response to the questions, but it must definitely be defensible as the best among the suggested alternatives.

It is widely agreed that the Multiple Choice test is the most valuable and most applicable of all objective test forms. The answers can be decided ahead of time to secure maximum objectivity.

It is also flexible in the sense that the teacher can offer several items on a certain problem.

There are several variants of the Multiple Choice type test :

a. Correct Answer.

(1) Positive Type.

(2) Negative Type.

b. Incomplete Statements.

c. Best answer.

d. Substitution.

e. Multiple Response.

a. Correct Answer.

(1) Positive Type.

Example :

Directions : Choose the best answer and indicate your choice with a check.

Items : 1. I suggest -- this method in your teaching.

a. you to use.

b. you using.

c. you use.

d. to use.

2. Because she has to deal with people who are ill in hospital, a nurse should be --- person.

a. a soft.

b. an easy.

c. a kind.

d. a sad.

(2) Negative Type.

Directions : Put a v in front of the correct answer.

Item : Who is not a 1945 novelist ?

a. Nur Sutan Iskandar.

b. Muhammad Yamin.

c. Chairil Anwar.

d. Hamka.

b. Incomplete Statements.

Example :

Directions : For each item, write the appropriate letter of the correct answer !



Items : 1. If mother had one, she --- it to me.

- a. will give.
- b. would give.
- c. would given.
- d. gave.

2. At the Senior High School the teaching of a foreign language other than English is :

- a. required.
- b. Discriminating.
- c. elastic.
- d. optional.

3. The daily weather forecast can be found in :

- a. a daily newspaper.
- b. an encyclopedia.
- c. a dictionary.
- d. a magazine.

c. Best Answer .

Example :

Directions : For each item, write the letter of the best answer.!

Item : The best reason for the choice of English as the first foreign language as the fact, that :

- a. it has many speakers.
- b. it is used in the United Nations.
- c. it has been analyzed scientifically.
- d. it has a wide distribution.

d. Substitution.

Example :

Directions : Write one word for each of the following

phrases a groups of words.

- Items :
1. A hundred years.
  2. Founteen days.
  3. The house where a king lives.

e. Multiple Response .

The Multiple Response type is not identical with the other Multiple Choice tests discussed earlier. Therefore, this type must be distingiushed from the others as it requires, one, two, or more responses to a single item.

Example :

Directions : For each item, check the alternative/al-ternatives, which will suit the sentence.

Items : 1. This explanation has not convinced :

- a. nobody.
- b. any one.
- c. no one.
- d. any body.

2. The rose smells :

- a. nicely.
- b. sweet.
- c. good.
- d. allright.

3. Sawahs are usually :

- a. dependance. )
- b. dependant. ))
- c. dependable. ))) upon the rain.
- d. dependent. ))

3. Matching .

The Matching Type items are almost the same as Multiple Choice

items, only in Matching Type there are several problems of stems.

Matching items consists of two groups. Items in Group II are matched with those in Group I according to the instruction .

<u>Group I</u>	<u>Group II</u>
1. questions	1. answers.
2. incomplete sentences.	2. phrases completing the sentences.
3. statements.	3. comments on the statements.
4. pictures.	4. comments in the pictures,=

The choices should be put not in an orderly arrangement. The number of items in Group I and Group II may be the same or different. But ~~to~~<sup>to</sup> increase validity and to avoid guessing, there should be more choices than the problems, at least three more.

The Matching Type items enable a particular aspect of a subject to be surveyed quickly and economically. The association of two or more things can be checked. For instance, terms and definitions, laws and illustrations, books and authors or incomplete statements in the first column which require a word or phrase from the second column. This type is very effective for measuring such types of learning, as seen in the words of Lindquist :

"The matching exercise is particularly well adapted to testing in who, what, when and where types of situations or for naming and identifying abilities" ( 9 , p. 150 ).

But Matching is not so suitable for measuring understanding on a high level as distinguished from mere memory. Besides, it is time consuming for the student to do this type of test, especially when it is ill-constructed, or when the list consists of too many items. The student has to go through a long list first before he

can make his choice. There are two variants of Matching tests :

1. Simple Matching .
2. Compound Matching.

1. Simple Matching .

Example :

Directions : In the brackets before each item in Column I, write the letter of the corresponding item from Column II.

<u>Items</u> :	Column I	Column II
( ) 1.	Every one must die	a. who never made a mistake.
( ) 2.	He got an invitation.	b. me to go to the movies.
( ) 3.	On his way home	c. she had to buy some butter.
( ) 4.	There is no man	d. to the Sutherland's party.
( ) 5.	Opportunity rarely	e. remains in the building.
		f. he had to pass a haunted house.
		g. knocks more than once
		h. sooner or later.

2. Compound Matching .

Example :

Directions : In the brackets before each instrument in Column I, write the letter of its type and the letter how it is played.

Items :

Column I	Column II	Column III
<u>Instruments.</u>	<u>Type Instrument</u>	<u>How it is played</u>
( ) a. Drum.	1. Wind instrument.	A. With a plectrum.
( ) b. Accordion.	2. Percussion instrument	B. By striking keys.
( ) c. Xylophone.	3. 4-stringed "	C. With a bow.
( ) d. Violin.	4. 6-stringed "	D. By beating with
	5. 7-stringed "	two sticks.
	6. 7-stringed "	E. By pressing keys
	7. many stringed "	at each end.
		F. By blowing through a metal mouth piece.

#### 4. Rearrangement or Rank Order.

The Rearrangement or Rank Order type is a test in which students are asked to arrange the items according to a certain appropriate order, either chronological, logical or some other kind.

This type is not so popular among teachers, because of the special scoring problem that is involved. Suppose, the correct order of the four items, presented as a - b - c - d, is d - c - a - b, one student might miss all the four items or may have partly ranked them right, for instance d - c - b - a .

Teachers then can consider the item as a unit, or assign a total number of points that would be proportional to the correlation with the correct order.

Compared with a 4-alternatives Multiple-Choice test, where a student has a chance to have the item correct of one in four, the student doing a Rearrangement type scored as a unit, has only a chance of 1 in 24. And this will obviously put aside any mechanical aids in scoring, thus teachers are discouraged from using this type.

Example \*

Directions : In the space provided, write the letters of words or phrases in the correct order so as to make a good sentence .

Items : 1. She was born :

- a. in the year 1964.
- b. at half past one.
- (....) c. on November 22th.
- d. in a maternity clinic.

2. I don't know

- a. the.
- (....) b. is.
- c. time.
- d. what.

5. Interpretive.

Here the student is supplied with a printed passage or a table of figures, or a map or diagram or musical excerpt, and is required to answer a series of questions based on what is presented.

This type can go with almost any type of test.

Example :

Directions : Decide whether the statement is true or false based on what can be seen in the picture.

Items :



- 1. She has bracelets.
- 2. She is wearing a sleeveless dress.
- 3. There are two different women.

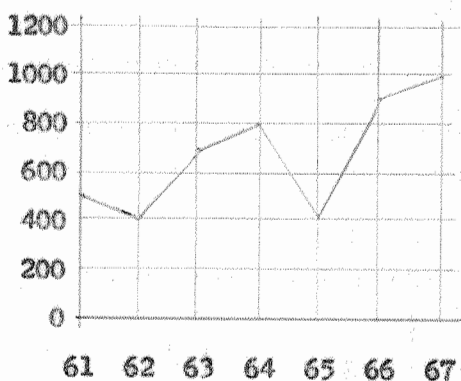
**Directions :** The proctor will play in succession excerpts of music. He will tell you when to write the answer on your answer sheet.

**Items :** For <sup>each</sup> excerpt encircle:

- a. if it is folk song.
- b. if it is a classical.
- c. if it is popmusic.

**Tourists visiting**

**Yogyakarta**



**Directions :** Refer to the diagram and then answer the questions!

- Items :**
1. Is the flow of tourists to Yogyakarta between 1964 - 1965 increasing or decreasing ?
  2. In what year did Yogyakarta have the most visitors?
  3. Was there a time when Yogyakarta did not have any tourists at all ?

**6. Other Miscellaneous Forms .**

There are still other types that cannot be grouped into one of the types earlier mentioned, such as the following .

**Example :**

**Directions :** Write the words between brackets in the correct form on your answer sheet !

- Items :**
1. He (write) two letters and is now writing a third one.
  2. Many (thesis) have been written on the teaching of

English .

3. Early today the school attendant (water) the garden .

Directions : Write a word, similiar in pronunciation but different in spelling and meaning, next to the following.

Items : 1. pair -  
2. made -  
3. flower -

Directions : Write the full form of the following words!

Items : 1. phone -  
2. pram -  
3. plane →

Directions : Combine these two sentences .

Items : 1. a. You will never reach your goal.  
b. You study very hard.

The examples given do not claim to be exhaustive, there are many other possible constructions.

\* \* \* \* \*



## CHAPTER IV

### TEST CONSTRUCTION

This chapter will consider the principles in the construction of tests, taking into account the fact that most tests are administered at our secondary schools without the necessary understanding and knowledge of what to test and how to test. Too often tests are made over-night, assembled from materials within the immediate reach of the teacher or from ideas that flash into his mind. Thus instead of materials selected to meet the purpose of a certain test, the usual practice is just the opposite, the available materials determine the test. Consequently, out of such a situation only unsatisfactory results are to be expected. Scores obtained from such tests would certainly reflect the student's ability falsely, which in itself is depriving the students of fair evaluation.

Therefore, the construction of objective tests, which is usually done by the ordinary classroom teacher, but which calls for some rather specialized professional skills in the preparation and analysis of items, must follow certain principles and procedures. These principles, however, can be easily understood and teachers should be able to apply them in their own classroom "testing-program".

Admittedly, constructing tests is a laborious task for the teacher.

A good test must have evolved through the following steps :

*general principles of test construction :*

A. Planning the Test .

B. Writing the Test.

C. Trying Out the Test.

D. Analysing the Test.

*Preparing*

*Evaluating*

964 - 171  
Stanley

#### A. Planning the Test .

In planning any test it is always advisable to begin with a



list of specifications for the required test : i.e.

1. The Purpose of the Test.
2. The Specific aspects of the Course Content to Be Tested.
3. The Age or Grade level for which the test is meant.
4. The desired Range and Distribution of scores for the test.
5. The Type or Types of Objective items to be used.
6. The Form of Administration and the time limit desired for the final test.

1. Purpose of the Test .

Since English examinations have profound influence on the way English is taught and since meaningful research on language learning is not possible without standardized testing, the whole subject of examination should be done in the right way.

The present methods of examinations <sup>do</sup> not relate well either to measurement of achievement or to stated goals of teaching English.

As stated in the Departemental decision of the Ministry of Education and Culture, regarding the teaching of English as the first foreign language to be taught at our secondary schools , the following objectives are put forward :

Heaton : 1979, 4

- a. Effective reading ability.
- b. Ability to understand spoken English.
- c. Writing ability.
- d. Speaking ability.

The decision also points out, that no full mastery of the first foreign language is expected, but just a "Working Knowledge" is required.

To be of maximum help for teachers in constructing tests,

the objectives should be stated as specifically as possible. It is not enough to say, that the objective is to require a "Working Knowledge of English".

Once the specific objectives are determined, the teacher may decide for himself, what he will test as to attain these objectives. By using techniques suitable to each objective, the teacher can evaluate his effectiveness in the classroom.

## 2. The Specific Aspects of the Course Content to be Tested.

The next thing to be decided upon the preliminary stage of testing is the specific aspect of the course content, that will be included in the test and their proportion one has, versus the other. The test should reflect the degree of emphasis of each section of the course.

As we have learned in an earlier chapter, Language Learning and Language Testing are inseparable in the sense that they influence one another. Therefore, the test should also put more weight on those parts that get greater stress in teaching. A large number of class periods devoted to a certain topic would be emphasized on testing. The more class periods devoted to a particular topic, the greater weight this topic would be given in the test.

For a test to achieve an adequate balance of the several aspects of the course, it is worthwhile to draw an outline, not just a topical outline, but also how much weight each portion is likely to receive. The following example clarify this :

Outline For a Test on  
General Proficiency

SMA - 2<sup>nd</sup> Year

Course Content	Number of Items	Notes
1: <u>Spelling</u>		
a. Punctuation	10	
b. Spelling		
2: <u>Pronunciation</u>		
a. Stress Recognition in Single Words	10	
b. Recognition of 'same-different' sounds		
3: <u>Vocabulary.</u>		
a. Word meaning.	15	
b. Antonyms.		
c. Analogy.		
4: <u>Structures</u>		
a. <del>Order</del> Order.	35	
b. Agreement.		
c. Conditionals.		
d. Voice.		
e. Tenses.		
5: <u>Comprehension.</u>		
a. Recognition of True-False statements.	30	
b. Short Passages or paragraphs as a basis.		
c. Idioms		
<b>Total</b>	<b>100</b>	

### 3. The Age or Grade Level for which the Test is Meant.

The item constructor must attempt to adjust the difficulty of the item to the subjects from whom it is intended. The age or grade level of the students taking the test must be considered as this will have great bearing in content and in form and the time allowed for completion. Younger students tend to be restless if they have to work on long tests. Also the type of structures and vocabulary must be kept well within their mastery-range. Factual questions would suit such a purpose better than questions involving high mental processes. If comprehension passages are required, they must be within the students' experience or be appealing to them.

### 4. The Desired Range and Distribution of Scores for the Test .

In planning a test, the teacher might have a certain level of difficulty in mind. He may do this for acceptable reasons. For motivational purposes he may plan to give a test in which the majority of the class will do well. In view of this purpose then, by the time he writes the item he will choose only those items which he is sure that the greater part of the class will have a fair chance of answering correctly. He might plan 75% or more passing marks for this.

A test difficulty of 50% is generally agreed to be effective and proper. This means that the average score is fifty. Unfortunately this is not always possible in practice.

Generally speaking the difficulty of a test must be geared to the students for whom the test is planned. Therefore, a teacher with some knowledge of the background of his students is likely to be more successful in writing tests.

It is worth noting that the difficulty of an item is not determined by the Statement of the items but more by the degree

of discrimination elicited by the alternatives. In the case of Matching tests, for instance, the difficulty may be increased by providing more responses than are need.

5. The Type or Types of Objective Items to be Used.

The type of test will be determined by the particular purpose for which it is intended and not because it is "in" at the moment the test is administered. It is not the newness of a test what matters, but the ability to measure what is supposed to be measured.

For school reports to parents, a test that is able to rank students is the one to be chosen. But if the test is trying to discriminate the weak from the strong, then a test with a diagnostic nature is the proper one to select and use. What is important in the first example is less important in the second.

It is good if more than one type of test is included in a test, because certain types, by their very nature, are able to test specific aspects better than others.

Furthermore, several types will bring variation especially in long tests.

6. The Form of Administration and the Time Limit Desired for the Final Test .

As soon as the teacher decides to give a test, he must make up his mind as to how he would administer the test, orally, in written form, or a combination of the two, considering the time available for testing, the facilities for mimeographing the test, cost of test and age and experience of students being tested. Also, enough time should be allowed to students to attempt all the questions, especially at the "try-out" stage, because at this time the teacher still does not know the degree of difficulty of each item or whether the discriminating power is great or not .

The exact time limit for the test in its final form can only be determined after the "try-out" and evaluation periods.

### B. Writing The Test .

When the plan is fixed, items or questions are then written with the specification<sup>s</sup> of the plan in mind. The following are recommended for consideration in writing test items :

1. The Number of Items for the Try-Out.
2. The Inclusion of Optional Items.
3. Directions for the Test.
4. The Need for Criticism from others.
5. The Form and Nature of test items.
6. The Make up of Tests.

#### 1. The Number of Items for the Try-Out.

For the trial test more items than are required in the final test, should be prepared. There must be at least 25% items more than are needed for each section of the test. Items that prove to be not discriminating at all, can then be dropped out.

If the teacher writes down possible questions ( as the course material is presented and studied by the students-in the course of classroom work ) - then he will have a fund of questions available, with which he can construct his test.

#### 2. The Inclusion of Optional Items .

Sometimes teachers may wish to include optional questions in the test, In the case of an essay type test this more as a compensation for the in-adequate sampling.

It is also a source of a problem in scoring, since students no longer take exactly the same test. The papers cannot be marked using the same judgement. Yet the exposure of several questions to choose from is often employed as a means of finding out what

parts of the content already taught are liked most or mastered best. The teacher may then plan to re-teach these parts.

### 3. Direction for the Test .

Before discussing the writing of the test itself, the directions for a test or each subsection thereof must first be considered. A test must have directions for students as well as for the test administrator.

Directions should be clear and precise, allowing only one possible interpretation. The words and sentence-structures should be such that even the weakest student can understand what he is expected to do. A student may not know the answer to a question, but at least he must know what is expected from him.

In the early stages of foreign language teaching this means that the teacher must use only those vocabulary words and sentence structures known to his students. For instance, "Draw a line under" is perhaps preferable to "Underline" .

Sometimes an example to illustrate what is expected, is worthwhile. Furthermore, tests meant for use in several classes other than one's own, which may not be familiar with a certain type of test, should certainly always have a complete set of directions.

Students must be told where and how to mark the items, how much time is allowed and whether the items are of equal value or not and also how the test is going to be scored. The last aspect especially is highly necessary in cases where the scoring will employ some reduction for errors made, the so called "correction for guessing". This is one way to prevent students from "wild guessing". It would not be fair to score papers this way without telling students about it. Such an additional instruction may read :

" Your mark will be the number of right items minus the number



ber, of wrong items. Therefore, if you are not sure of the correct response to a question, do not guess as this will reduce your score".

4. The Need for Criticism from Others.

Once the items are written it is desirable to obtain the opinions of other teachers of the same subject about these items. This is to ensure that the items are expressed clearly and unambiguously, that the proposed answer is really correct and that the distractors are attractive enough to be chosen by the students. The way a question is stated must be critically examined, because too often ambiguities are not realized. One answer, for instance, could perhaps be correct with one interpretation, seen in this example.

He sent a souvenir from Paris ( adapted from 18, p. 84 ).

- a. The souvenir was sent from Paris.
- b. The souvenir was a Parisian souvenir.
- c. The man sent a Parisian souvenir from some place.
- d. The man sent a Parisian souvenir from Paris.

Sometimes a change or addition of word is sufficient to erase the ambiguity. In some other cases the item must be rewritten or even discarded.

After items have been revised to take account of such criticism from colleagues, a native speaker, if one is available, is asked to read the test. When all this has been accomplished, a trial form of the test is ready to be administered to a group of students of an age and ability similar to those with whom the final test is to be used.

## 5. The Form and Nature of Test Items .

Item writing is a difficult task indeed. It needs skill and ingenuity. A way to make oneself acquainted with item writing is to get hold of a good standardized or teacher made test and scrutinize it. The items in those tests, that prove to be good must have gone through the following general considerations.

- a. Types of questions.
- b. Independence of test items.
- c. The Language.
- d. The Distractors.

(B24) - 108  
Stanley

### a. Types of Questions .

Though writing items for a test is a skill in itself, where one person may have greater ability than others, this does not mean that one should be afraid of attempting to write a test of one's own. In a way, writing test items is an art, an art which is mastered only through much practice.

The particular purpose of the test and the level of the students are the most important considerations in determining the type of questions, e.g. positive or negative statements or questions.

Making special reference to objective questions of the True-False and Multiple-Choice type tests, will not make much difference whether the one form or the other is used. The common practice of fellow teachers and state exams, however, shows that in Multiple-Choice, the incomplete statements are widely used. In writing items of this pattern, students taking the test, must be given enough context to read before they are confronted with the alternatives. This means that the lead, that is, that part of the statement leading to the alternative, should have enough context. If the choice is put at the beginning, a student has to read the

whole item first and then go back to the alternatives to decide the correct response. This is certainly not economical.

Example : Poor : He is (a) ambushed ) -- because there  
 (b) ambitious ) ) is no money to  
 (c) embraced ) ) ) pay his bills  
 (d) embarrassed )

Better : He has no money to pay his bills, that's why

he is (a) ambushed

(b) ambitious

(c) embarrassed

(d) embraced.

Coming back to what must be considered as enough context in the lead (15, p. 129) - the following illustration will make it clear .

Enough context : He won a prize (a) goods captured at sea.

Too much context : He won a prize (b) reward resulting in the 100 m race. from a successful

Not Enough context : A prize. competition.  
 (c) the amount of money asked for anything.

To conclude, whatever type of question the teacher may wish to use, it must be kept in mind, that the question must elicit one correct response only. The more limited the response, the more objective the scoring and less time will be needed to read and score the test.

#### b. Independence of Test Item .

Test items must stand by themselves, even in a subdivision of a test. No item may have any dependence upon other items, ei-

ther preceding or following. No item may overlap any other item, as this will probably give away the answer to both. If a student knows the answer to one item, he most likely will also know the answer to the other overlapping items, because the response to this latter is made clear by the first item. Any single item must be able to be interpreted in itself without reference to another.

If the teacher keeps this possibility in mind, then he is in the position to safeguard his items from overlapping one another by eliminating words or groups of words that otherwise may cause trouble. This effort will save double work, because then not many items need reconsideration after try-out.

*Some important points to consider in construct a test multiple choice item.*

c. The Language Used.

- (1) should be clear, brief and economical.
- (2) should avoid "improper" context.
- (3) should not give away the answer.

(1) Clear, Brief and Economical Language.

The language used, must be clear and precise without being too technical. It must be simply phrased, leaving out irrelevant things. Technical vocabulary of low frequency, figurative language or vague constructions must be avoided as these will, in many cases, blur the meaning of the item rather than clarify it. Care in phrasing the question will save trouble later.

Students, who are expected to know the point might miss the item, simply because of the way the question is phrased.

Being classroom teachers, we will find difficulty in the wording of test items, in limiting ourselves to the essential things that must be put in a question.

Ideally, next to being clear and brief, an item must also be economical, not only in terms of the time needed to read it,

*Since the test is on grammar, the lang. used should not be difficult. should use the words common to most students. vocab. in gram test should not cause any trouble in doing of the gram test*

(c)  
9/000

not

(b)

*but economical & not present the problem clearly*

but also in terms of the space it will take in the test booklet.

In testing vocabulary, for instance, in case you want to test just the single words, if the lead in every item is the same and consequently does not contribute anything to the context, than it is economical to put the test word alone as the lead, as is the case of the "Word Power" section of any "Reader's Digest" issue. This of course, applies only to words having one general meaning.

Example :

Directions : In the test below, check the word or phrase you believe is nearest in meaning to the key-word.

Item : Equivalent (a) related  
(b) relative  
(c) questionable  
(d) equal.

( Reader's Digest, Vol 10, no 55, Oct.1967,p.147 ).

But considering that the meaning of a word is conditioned by the context of which it is a part, then, instead of using isolated words as the lead, it is better that the testword be included in a context. To maintain the principle of economy, it is enough to underline the word or have it printed in bold type, as this would be a better practice for words having several meanings.

Example : I pay the bill monthly .

- (a) A statement of money owed for things bought.
- (b) The horny part of a bird's mouth.
- (c) A plan for a new law.
- (d) A prize of paper money.

( Alternatives, 11 , - p. 95 ).

(2) Avoiding "Improper" Context .

In the rush to finish a test on time, teachers are often careless in wording questions. In most cases they are too absorbed in distracting students from the correct answer and they fore forget the flow of sentences, allowing them to become some thing unnatural. A context like this will reveal its weakness easily when the test is administered.

Example : A ship gives wool . ( True-False Item ).

In this, Tend not to choose the correct answer (C), because it is wrong if put in the context.

Considering also that our learners may consists of students having different cultural backgrounds and beliefs, the teacher must also avoid construction that may cause offence to students in the group. Though this may seem superficial, sometimes it may have a far reaching effect as it may disturb the condition of learning. The teacher may not be liked by the particular group that feels offended and consequently also the subject he is teaching will be disliked. Teaching under such conditions certainly will not be easy.

Therefore, before administering his test the teacher must really check whether the context may cause offence. The sample following is an effort to check the understanding of the word "delicacy" using a lead that must be discarded, as it will offend students coming from certain parts of Indonesia.

Example :

Dogs are a delicacy in certain parts of Indonesia.

- (a) fineness of quality.
- (b) fineness of feeling.
- (c) a choice kind of food.
- (d) a state requiring careful treatment.

( Alternatives, 11, p. 305 ).

(3) Avoid Giving Away the Answer .

The teacher must strive to word his question in such a way as not to reveal the answer. It is the whole context of the item that should determine the correct answer. Therefore, he must be aware of these so-called "specific determiners" that may give clues to the right responses without necessarily being accompanied by an understanding of the item as a whole.

Some specific determiners that may give clues to the correct response, are illustrated in the following groupings.

a. Grammatical Form as Specific Determiners.

Example : Language learning is

- (a) learning things by heart.
- (b) learning new words.
- (c) a mastery of sounds.
- (d) a process of using and interpreting sounds meaningfully.

Remark :

The form of 'learning' indicates that it is a process.

b. Words or Groups of Words as Specific Determiners .

The use of the exact words of the textbook must be very much avoided as this will promote only rote memorization.

Especially in "True-False" items, words such as "always", "never" and the like tend to make the statements false rather than true, whereas statements containing words like "may" and "sometimes" are generally true.

Example :

True-False :

1. Much reading in English may help the mastery of that language.

2. Indonesians learning English will never be able to distinguish the vowel sounds in "man" and "men" well.

If a word is more known than others in the alternatives, then one is likely to use this as a basis to narrow or secure the correct response.

Example :

Double Choice : Poor.

Letters have (stamps, stumps) on them.

Remark : Stumps will automatically be discarded because of its low frequency.

Better : When You walk in this are in the dark, beware of (stamps, stumps).

#### d. The Distractors .

Definition :

By distractors is meant here, those alternatives, other than the intended correct one, that are suggested in an item to be so plausible that they may be chosen by the student without adequate knowledge in the particular field demanded by the item. So to put it more clearly, distractors are meant to attract those, who do not have sufficient knowledge to choose the right answer. Thus if the distractors do not <sup>do</sup> this, they are not good.

To be effective, distractors in one item must be viewed in terms of their : (1) number.

(2) plausibility.

(3) placement.

#### (1) Number of Distractors.

It is generally agreed that an adequate number of

Heaton, 15-66



choices should be used in a Multiple Choice test. A 4-choice item is sufficient and is considered as the minimum, as any item having less alternatives will allow a certain degree of guessing. Increasing the plausibility of the distractors will help reduce the chance of guessing.

(2) Plausibility of Distractors.

Distractors must appear to be reasonable answers to students. If not, they do not serve their purpose and must be replaced by others. Therefore, it is not just a matter of putting any word or phrase different than the correct response to serve as a distractor. It must meet certain qualifications.

Distractors must be plausible, because the correct response may not only be obtained by directly selecting the correct alternative, but also by discarding the incorrect ones. They must approximately be of the same grammatical pattern or length as the correct response. If one of the optionals is an adjective, for instance, then all the others must also be adjectives. Therefore, if the intended correct alternative happens to be the one that is different from the others, it is easily singled out for its obvious nature. Or, in rare cases, it may also be the opposite.

(3) Placement of Distractors.

When items have been written in draft, the teacher must consider the position of the correct or best answer among the other alternatives in one item and also in its total sequence of the test. Correct answers must be placed at random so that the order of the correct responses is more a chance order than a regular pattern (11, p.p. 364-8).

A sequence that follow a regular pattern, though it may

facilitate scoring, should be discouraged. Besides, students are too often very quick in detecting it. For instance, many teachers tend not to use or underuse the first position in a Multiple Choice test and students being familiar with their teachers' bias, will then easily narrow his choice, disregarding the first answer.

However, there are cases in which the optionals need not be randomized. When the alternatives consists of numbers or names, they can be arranged in chronological alphabetical order. In some other similar cases they can be put in either ascending or descending order.

Example :

Directions : Arrange in order from greatest to least.

- |                |                |               |
|----------------|----------------|---------------|
| <u>Items :</u> | 1. (a) cottage | 2. (a) canon. |
|                | (b) hut.       | (b) mortar.   |
|                | (c) mansion.   | (d) pistol.   |
|                | (d) palace.    | (d) rifle.    |

I enjoy watching --- the children playing in the park.  
 A. about  
 B. at  
 C. on  
 D. to.

Also assigning equal numbers to each position in a Multiple Choice test should be avoided. In the case of a True-False test, the tendency is towards putting more false items than correct ones. In itself it means no harm, but if this is practiced regularly the validity of the false items will be reduced, since the students will tend to mark doubtful statements as false.

6. The Make - Up of Tests.

The Make-up of tests is in many cases neglected or overlooked. Once the teacher manages to finish his draft, he gives it to a clerk to cut the sheet without any further instruction

as to the spacing, types of letters or lines used. He usually leaves it to the discretion of the typist.

The aim and practice of the typist is to put as many items as possible on one sheet, as this will mean less time for him to run off his mimeograph-machine.

Spacing as well as the types of letters used are important in tests. Testees must be able to locate at once one item from another and one group from any other. Younger students certainly need a bigger type of letter and proper spacing.

Alternatives in Multiple Choice and Matching tests must be placed on one page. When it is not put on the same page, it requires a turning back and forth to locate the correct response. This is not only time consuming but also confusing. They must also be arranged one below the other, rather than written one after the other on one or several lines.

The test items must be so arranged as to allow the responses fall on one side of the test paper, for instance, at the left or right side of it. Moreover, the place to indicate the correct response should be clear, making use of parentheses ( brackets ) or dotted lines.

### C. Trying - Out the Test .

After items have been written according to plan and revised to take account of the many principles of test construction previously mentioned, a trial form should be administered to determine how good the test is. It is not possible to know in advance whether the item will turn out to be really valid without a pre-test stage. The necessary try-out should take the following aspects into consideration .

1. The Students Used.
2. The Administration of the test.

Stanley 1931

### 3. The Scoring of the test.

#### 1. The Students Used .

The trial form should be administered to a group of students of an age and ability similar to those with whom the final test is to be used,. If the test is meant to measure the general proficiency in English of students in suburban areas , then the try-out must also use sample students studying in such areas and not urban students. And if the test is to measure progress made in a scholastic year in a particular grade, then it must also be administered towards the end of the schoolyear, the try out as well as the actual test.

#### 2. The Administration of the Try Out.

##### a. The Conditions Favoured.

Though it is but a Try-Out, the teacher must strive to create approximately the same classroom testing atmosphere as when it will be actually administered later. The pre-test must stimulate the real one in every way.

##### b. The Time Allowed.

The time allowed for the test in the trial process is an important factor, more important at this stage rather than in the administration of the final form of the test. The teacher in proctoring the test must be not too strict in this matter , allowing enough time for almost all students to cover all items. However, in the case of a speed test things are different. Here the proctor has to stick to the allotted time as strictly as possible, since the amount of time in this case is the most important factor. On a speed test the student works against time, since during such kind of tests the level of difficulty remains constant.

On a language speed test even the better students know that

they will probably not be able to finish.

### 3. The Scoring of the Try - Out .

Since preferably all items in objective tests must be of equal value, regardless the range of difficulty, it is more than fair that one correct score deserves one point. Fractional credits must be avoided.

With regard to essay tests, students must be told that spelling or grammatical errors will reduce the final mark or not, thus influencing the care in writing the essay.

When the scoring business is completed, another task is awaiting, i.e., the scores must be examined for interpretive purposes.

The score in a language speed test will depend not only on knowledge and comprehension but also on the speed at which they progress. In order to discourage indiscriminate guessing on a speed test, the numerical scores are often modified :

two options : ( for example : true-false )

adjusted score = number right - number wrong.

three options :

adjusted score = number right -  $\frac{1}{2}$  number wrong.

four options :

adjusted score = number right -  $\frac{1}{3}$  number wrong.

five options :

adjusted score = number right -  $\frac{1}{4}$  number wrong.

The rationale of this adjustment is the following. Let us assume that out of 100 items with four options, Ali had finished 60 items, when he realized that only 30 seconds remained. He quickly marked Response A for items no. 61 - 100 on his answer sheet; he hoped to get about one out of four correct and thus

raise his score by 10 points. If we assume that Ali actually did make 10 correct guesses and 30 wrong ones, then his adjusted score on Items 61 to 100 would be :

$$10 \text{ right} - 1/3 (30 \text{ wrong}) = 0 \quad (23 \text{ p. } 37).$$

Thus Ali gained nothing through his wild guessing.

On the power test, where all students have the time to finish all the items they are capable of doing, such an adjustment is unnecessary. No student would have to answer a long series of items by blind guessing in order to finish the test before the time limit.

#### D. Analysing the Test .

The results obtained by students in the try-out test should be evaluated.

"No matter how carefully the test is prepared, in the first place, its merits should be established and not merely assumed." (193, p. 160).

The results must be analysed and interpreted statically to determine how difficult each item is for the group and the extent to which it discriminates between the better and the poor students.

An analysis of the responses to the individual items of a test will prove helpful for two reasons. First, the teacher can discover if there are certain points that a sizable number of students have failed to master, so that remedial work may be done before proceeding to the next unit. Second, the teacher will know how well certain topics of the classroom material are mastered by his students.

This information will be useful in constructing new tests. Item difficulty is determined by observing what percentage of the students answer the item correctly. The more difficult the item is, the fewer will be the students who select the correct option.

Discriminating power tells how well the item performs in separating the better students from the poorer students. If the upper third of the students gets the item correct and the lower two thirds generally get the item wrong, then it is good discrimination between these two groups. Very difficult items should discriminate between the very good students and all of the others, relatively easy items should discriminate between the majority of the students in the class and the few poor ones ( 23, p. 38 ).

\* \* \* \* \*

## CHAPTER V

### THE INTERPRETATION OF TEST SCORES

The step following the completion of the marking of an objective test is the interpretation of the scores obtained. Test scores will only have value if they are interpreted in a justifiable way. For this, it is well to consider the glossary terms that are used in a discussion on tests. However, it is not the idea of this chapter to be involved in a highly technical discussion of a mathematical nature.

It is already earlier stated in the introduction that this paper wants to present a simple and 'down to earth' explanation on testing, which hopefully will be very helpful for our secondary school teachers in their work.

#### \* Statistical Concepts .

Measurement and evaluation at the present state are more and more making use of statistical concepts and teachers are assumed to know the language of statistics and to be familiar with the commonly used statistics, since the interpretation of scores requires understanding of these concepts and a fair ability in computing them.

Suppose that an objective test on structures has been administered, the answer sheets must then be marked by assigning a numerical score to each one. These scores are recorded in the teacher's rollbook and as they are there, they have no meaning. To have meaning and to be useful, the marked papers must be summarized and analyzed. And it is here that the statistical methods come into use.

#### 1. The Frequency Distribution .

The very first step in determining whether a test is working or not and therefore very essential, is the arrangement of the answer sheets according to scores, in order from high to low. This is



to facilitate the tabulation of the frequencies of different scores. The highest score is often written at the top and the tally marks are written next to the corresponding score. When all this has been done the result is a table, called a frequency count or frequency distribution or just distribution. The process itself is called tabulation.

Example : A Frequency Distribution without Further Grouping.

Scores ( $x$ )	Tallies			Frequency ( $f$ )
20				1
19				2
18	—			—
17				4
16				7
15				8
14				4
13				7
12				9
11				8
10				15
9				10
8				9
7				6
6				5
5	—			—
4				2
3	—			1
2				2
1	—			—

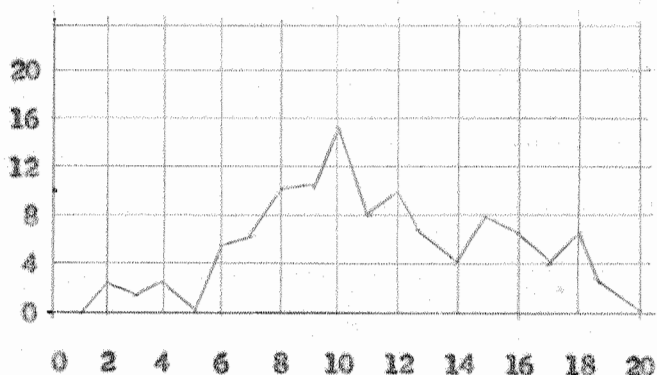
$n = 100$

Sometimes the symbol  $x$  stands for a substitute for the test score or variable and  $f$  for the frequency of a score and  $n$  for the numbers of answer sheets.

In this simplest form the scores take up some meaning and it can be made more meaningful, if it is converted into a visual display or graphical representation, being one of the following :

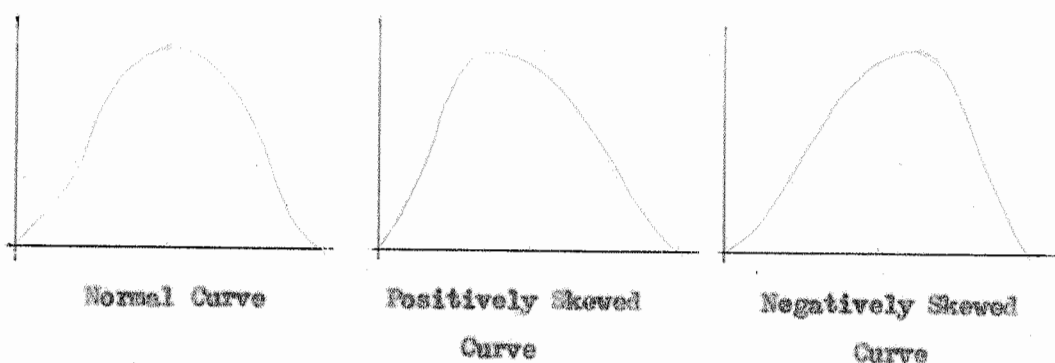
- a histogram.
- a frequency polygon.
- a bar graph.
- some form of chart.

Of the above mentioned, the polygon is the most popular and is often referred to as a frequency distribution

Example :

A Frequency Distribution may reveal :

- a. a perfect distribution known as the normal curve, where cases are equally distributed around the middle, forming a symmetrical shape.
- b. a positively skewed distribution ,  
where most of the cases are at the lower part of the range. In this case few students have a high score, the conclusion being that the test is too difficult.
- c. a negatively skewed distribution ,  
where cases are clustering around the upper part. Here it is just the opposite of the former. Most students do well in this test and therefore, the test has less discriminating power.



## 2. Variability.

### Definition :

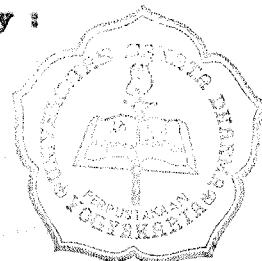
Variability is defined as the extent to which the scores tend to scatter or spread above or below the average (19, p. 81).

Students in a group may vary in their performance, though

they may have the same ability.

There are several measures to indicate the variability of a group. They all represent distances rather than points and the larger the distance, the greater the variability of the scores. Two of these measures will be discussed somewhat specifically :

- a. Range.
- b. Standard Deviation.



a. Range.

By range is understood the difference between the highest score and the lowest plus one, indicating the spread of the scores. In our sample the highest score is 20 and lowest is 2, which gives a range of  $( 20 - 2 ) + 1 = 19$ .

The range is employed in a grouped frequency distribution to determine the class interval.

b. Standard Deviation.

As we have learned, scores are not evenly scattered throughout the whole range. It will be more useful to know the range in which most of the classes fall. The measure used to obtain this partial range is the standard deviation. It is usually represented by a sigma (  $\sigma$  ) and may be roughly defined as the deviations of the scores from their mean, or the range above and below the mean.

A score may deviate from the mean. When each of such deviation is squared and then the result added up and finally its average computed, the result is a Standard Deviation, more often shortened into S.D, expressed by the formula :

$$S D = \sqrt{\frac{\sum ( X - M )^2}{n}} ; \text{ where :}$$

X is the score.

M the mean

$(X - M)^2$ , the sum total of the squares of the deviation of each score from the Mean.

$n$  : the number of scores.

When the mean and the S.D. of a group is known, we can determine both the general level and how wide the spread is. A small SD means that the group has small variability and the test is then relatively homogeneous whereas a large S.D shows that there is a certain heterogeneity.

The S.D is also used to represent the position of a student in a distribution. If the mean of a test is 70 and the S.D is 10, a student with a score of 80 is said to be one S.D above the Mean and the score is written as  $+ 1.00$  and likewise a score of 60 is one S.D below the mean and the score written as  $- 1.00$ .

The minus sign, however, is not an indication that the deviation is less than zero, for this is impossible. The symbol simply means that the deviation is below the Mean. These scores are called Standard Scores.

The S.D is the most effective to measure the variability of tests.

### 3. Central Tendency .

Many Statistical concepts are based on the idea of the normal Curve, which gives an ideal distribution. In a group of students selected at random the majority would be in most cases of average ability if measured by almost any criterion. The scores, therefore, tend to group around the center. A few students would be really gifted in English and a few others would have little talent for language. But this is true only for large groups selected at random.

One of the most effective concepts applied in tests is, therefore, the measures of central tendency. They are indices of the general or average level of performance of the group. These measures stating where the midpoints lies are :

a. Mean.

b. Median.

c. Mode.

a. Mean .

The Mean is just the arithmetic average of the scores and is determined by dividing the sum of the scores by the number of papers of students to which or to whom a score is assigned.

From our data, the Mean can be obtained as follows :

$$\frac{4 + 3 + 8 + 30 + 42 + \dots\dots\dots 38 + 20}{100} = \frac{1108}{100} = 11.08$$

The raw score Mean is formulated this way =  $M = \frac{\sum X}{n}$ , where the letter  $\sum$  stands for the sum total of the scores,  $X$ , and  $n$  for the number of papers.

It must be noted, however, that a mean can be misleading if the distribution is not even, as in the case of skewed distribution. Some extreme scores can effect the Mean value, so that the latter can give a misleadingly high score or low score.

A more desirable average than this computed one, is the Median.

b. Median.

The Median is the middle point which divides the distribution into two halves. Half of the scores lie above it and another half below it. In comparison to the Mean, it is simpler, especially when there is no need for a high degree of accuracy.

When the number of students happens to be an even number there won't be any midscore, unless the middle pair of scores

is averaged. If 50 students made a score of 10 or less and 50 others score 11 or more, then the midscore is 10.5.

The Median is obtained by just counting up or counting down half of the number of students taking the test, from there its other name, Counting Average .

The Median is also often used as a reference point for describing the location of individual students in a distribution.

c. Mode .

The mode of a test is the most frequent score. It is very obvious and it can be obtained by just going through the scores: an inspection. Therefore, a Mode is also referred to as an "Inspection Average" .

Of the three measures of CentralTendency the poorest is the Mode, because the change of certain scores may shift the Mode markedly and therefore, it is less trustworthy.

\* Scores .

1. Definition :

By definition, a score is a numerical description of an individual's performance on a test ( 19 , p. 276 ).

A Testpaper is usually marked by crossing out the wrong responses and then the right responses are added up. The sum total of the right responses is then the score.

2. Types .

There are several types of scores which can be distinguished as :

- a. Raw Score.
- b. Derived Score.

a. Raw Score.

When a testpaper is marked and the sum total of the

right responses added up, the score obtained is known as the raw score. If a student has got a score of 76, it means that he has 76 items right. It is known, what the total possible score is, whether his score is out of 100 and judged 76 in relation to this or whether that 76 must be taken to mean out of a possible score of 90. The number itself does not indicate whether the performance represented by it, is satisfactory or unsatisfactory.

Therefore, a Raw Score in itself, means very little. It may be the lowest score or the highest. If it is 76 out of 100, which is usually assumed, then the performance is relatively good, but what conclusion can be drawn if it is 76 out of 200, or when no other student scored lower than 76?

If there is need to compare two or more scores, the Raw Scores cannot be compared as they are. A common denomination must be found first for several test scores to be comparable. Then the Raw Scores are no longer maintaining their 'raw' status, but have become derived scores.

#### b. Derived scores.

A Derived Score is a numerical description of a student's performance in terms of norms. The norm itself is the performance of the group itself or any other defined group to be typical. Roughly speaking then, a Derived Score is any score that is obtained from computing the Raw Score.

If a student scores 60 on one test, and 80 on another, we may not directly conclude that he did better in the second test than in the first. If in the first test the total possible score is 100, he has got more than 50% of the items correct, but if nobody else in the class taking the test scores lower than 60, than there is only but one conclusion, that is, he must be the poorest.

Therefore, the Raw Scores are further worked out to give a stabler and more just description of the performance of the student. We can find the Mean of the Scores first. Assuming that it is 50, the student is then 10 points above average.

Another sample of a Derived Score, is the Percentile Rank. It is a description of a student's position in a group in terms of the percentage of students falling below that score. A student in the 60th Percentile Rank will have a score, in which 60% of his fellow students in the group score less than that.

### 3. Interpretation of Scores.

As indicated in a previous section, the process of measurement, whether carried out by test or other devices, merely yields a score. Measurement tells only the number of right answers and does not tell whether a student gave a good or bad performance. Correcting test papers is not done only to obtain scores but to evaluate them. Therefore, scores have to be interpreted before the performance, which they represent, can be evaluated. Teachers, students, parents, administrators and future employers are more interested in grades than in scores themselves.

The fact, that a student scores 75 on a test is of secondary interest. What teachers want to know is whether this represents a good or poor performance or a satisfactory or unsatisfactory performance. This requires that the score be evaluated against some standard.

For instance, if Tony scores 61 on a spelling test, is this good or poor? This question can only be considered satisfactorily if it is related to one of the three possible standards, referred to as :

- a. an absolute standard.
- b. an individual standard.
- c. a relative standard.



a. Scores Related to an Absolute Standard.

An Absolute Standard involves the idea that it is possible to achieve a perfect performance in the test or subject matter under consideration. It is a standard which is commonly used in schools and in the community and is one which appeals to the Public mind.

A student's score on a test is compared with the total possible score and is often expressed as a percentage of the total possible score. This carries with it the suggestion that the student knows, say, 70% or 80% of the material studied in the course, or included in the test. In other words, the quality of the student's performance is judged in relation to this absolute Standard. Yet, this is not as satisfactory an estimate of quality or mastery as it seems to be. If a slightly more difficult test had been employed, the same student might have been regarded as knowing only 60% of the material studied in the course. His knowledge of the subject matter would probably then be considered as less adequate, even though it was the same as before. Only the second test was just harder, the conclusion being that the score a student gets, depends on the difficulty of the test.

The employment of the absolute Standard in the interpretation of scores in teacher-made tests is reasonably satisfactory practice, provided that two precautions are observed. :

- (1) One is that due recognition should be given to the fact, that percentage scores on a test depend on the difficulty level of the test and that a student percentage score is not a precise and final indication of the extent of his knowledge about or competence in the relevant field or topic.
- (2) The other precaution applies to the common practice of

Adding scores or percentages in different aspects or subjects to obtain total marks. This practice is used widely by teachers and others who operate under the impression that it assigns either equal importance to each subject or the relative degree of importance indicated by the total number of marks allowed for each subject.

b. Scores Related to an Individual Standard.

When a student score is evaluated with respect to his own previous performance, or to what might be expected of him on the basis of an intelligence test result or some other criterion, the evaluation is made in relation to his own individual standard. His score of 6 out of 10 on a test on spelling might be considered as a satisfactory score for him if he had previously scored only 2 out of 10 on the same or on a similar test.

This type of comparison is involved in the assessment practices employed in the individualized instruction schemes. It is useful if a student wants to proceed at his own rate. Evaluating a student's work in terms of his own previous performance appears to be a good method of promoting learning, but unfortunately it is not very practicable in a competitive society with large numbers of students needing instruction.

c. Scores Related to a Relative Standard.

A Relative Standard is employed for interpreting scores when they are evaluated in relation to the scores obtained by some reference group of persons. The group must be one with which a student can be sensibly compared. It can be his own class, or a class of the same level of ability and learning or in his own age group throughout the town or country. A student's performance could then be judged to be among the best, about aver

age or below average.

d. Norms .

The terms normal and average are very much used, both in school and out and especially in the scope of tests we make judgements as to whether the students' achievement is normal, above average or below average. Diana's reading for example, is below average. This is just another way of saying that Diana cannot read as well as the average student in her group.

From a straight forward statistical statement meaning, the simple term "norm" has now become an implied expectation for each student. Now there is a feeling that somehow it is not a good thing for Diana not to achieve up to "norm". There is a stigma attached to the term below "average".

Norm has taken a new concept in the mind of the lay-man — a minimum to be achieved by all students — or otherwise you are defective. It is often assumed, that a student below average, is not trying hard enough or the like, thus forgetting that the average student does not exist, that it is undesirable and impossible to make everybody alike and that is just a statistical concept, not an expectancy figure. Forgetting also that there will be a range of scores with comparatively few for above or for below the average.

A student with a score below average may be lazy or he may be not. A poor score need not necessarily be evidence of poor ability in the subject under any and all conditions, provided the test is valid. He may have suffered from poor teaching which unfortunately is evident nowadays, or it may be caused by other factors such as emotional disturbance or dislike for the teacher.

The norm for a particular class on a particular test is the average score of the class, the amount of knowledge known

by the group of students that may help compare the measurement. However, norms refer not only to "average", but to levels of performance generally, at all points above and below average as well. They describe the actual levels of performance of defined groups and must therefore, be distinguished from standard. Schools vary greatly in terms of equipment, the training and experience of teachers, the social class of students and other aspects.

The idea that the desirable limit is reached when the class average is the same as the norm for the test, is unfortunately what many teachers believe. But it must be realized that the real function of the teacher is to lift the class above the average. If enough teachers managed to do this, the norm would have become outmoded and new norms would be prepared.

#### 1. How to obtain norms.

Norms are obtained by standardizing a test. Standardization is carried out by applying the test to a large and appropriate group to determine the distribution of the raw scores for the group and thus establish levels or standards of performance. A large number of scores are reduced to a form enabling effective comparison, by determining norms which can be stated in terms of certain units, such as :

- (a) content.
- (b) age.
- (c) grade.
- (d) rank.

##### a. Content Norms.

Content Norms are norms expressed in terms of units of content, of what it is to know a foreign language. These units have to be expressed in terms of the specific problems. Thus the con

tent might be the number of vowels mastered or the number of words.

It must be noted, however, that "mastery" has a relative meaning, it does not refer to all aspects. For instance, vowels may be mastered in certain environments and still in some other position they may form problems.

Content Norms are meaningful in measuring and interpreting performance on the elements of language.

#### b. Age Norms.

At each stage of development in the course, there is a maturity commensurate with it. In interpreting scores, Subject Age — a measure of subject maturity or level — is used in very much the same way as the use of MA — Mental Age — in interpreting scores in terms of chronological age, which is relatively stable, but the norms must first be standardized.

If a student of 10 in a comprehension test has achieved the same score as a second year S.M.P. student whose average age is 14, we may say that he has a Subject Age of 14 - 0. His Subject Quotient is  $\frac{14}{10} \times 100 = 140$ .

However, it must not be overlooked that age units on a certain test are not comparable to those on other tests, which may be testing the same aspects. But it does not mean that they have no value. They are just inadequate by themselves.

From the fact that two students have a S.A of 10 - 0 may not lead to the conclusion that they are identical in achievement. One might be very good in writing, whereas the other might be strong in comprehension and not in writing.

#### c. Grade Norms.

Tests can also be interpreted in terms of Grade Norms. Grade Norms are the average scores made on the test by students

in each grade in different schools.

In interpreting Grade Norms, one must be alert of the possibility of students with high scores who in fact belong to a high or class due to the strict promotion policy of the school.

As the case is with Age Norms, Grade Norms have low comparability.

d. Rank Norms .

Next to the norms previously discussed the Rank Norms is the most important. It interprets a student's score by ranking him among the group in terms of the percentage of students that score lower than he. From there its other name Percentile Norms.

A student in the 70th percentile rank must be interpreted that 70% of the students taking the test scored lower than he.

e. From Raw Scores to Marks.

Raw scores are usually converted into marks. It is desirable for marks to have the same value in every school system. However, the transfer from Raw Scores into marks is something subjective, even when the test is of the objective type and scored objectively. This is due to the fact that marks or grade are stated in general terms. This can also be achieved by the use of Standard Scores; a more practicable method consists in assigning a letter grade to each test score to indicate its relative quality.

Grading on the Curve .

A common grading system employs the letters A,B,C,D,E each letter grade being awarded to a fixed percentage of students, C being the lowest passing grade.

These percentages of students are often based on the Normal Distribution Curve, the percentage of students assigned to each cat

BIBLIOGRAPHY .

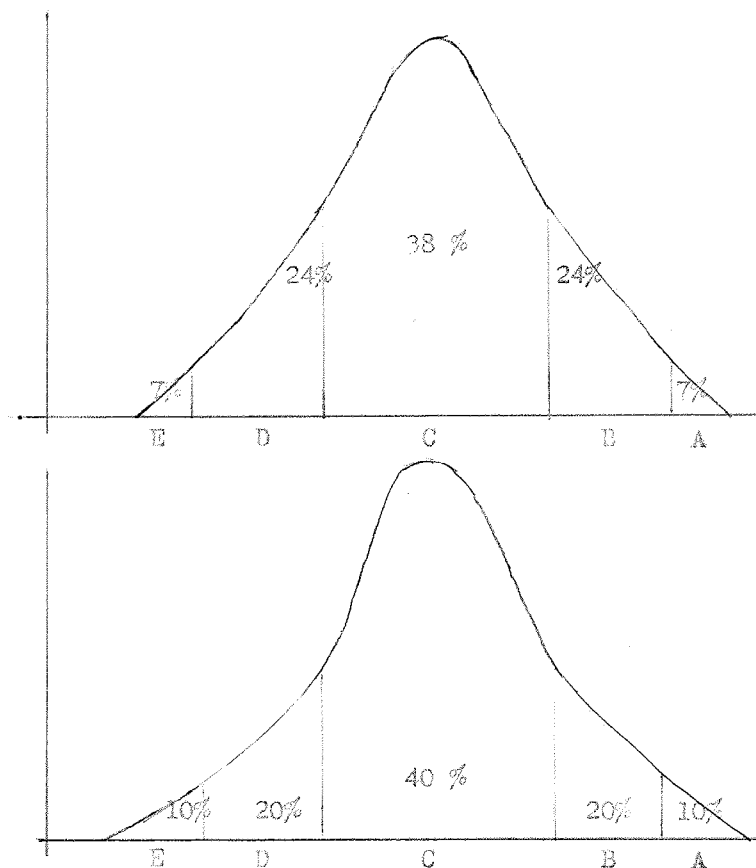
- Aerson, Scarvia B. Sequence in Multiple Choice Item Options  
Journal of Educational Psychology. October, 1952.
- Ashburn, Robert E. An experiment in the Essay Type Question  
Journal of Experimental Education, September, 1938.
3. Ballard, P.B., M.A.D. Lit. Teaching and Testing English. Uni-  
versity of London Press Ltd. Warwick Square, London,  
E.C.4. , 1961.
4. Brownell, William A. The Use of Objective Measures in Evalu-  
ating Instruction. Educational Method, 13 : 401 - 408 ,  
May - June, 1934.
5. Carroll, J.B. Notes on the measurement of achievement in for-  
ign languages. Harvard. Mimeographed, August, 1954.
6. Cronbach, Lee J. Studies and Acquiescence as a Factor in the  
True-False Test. Journal of Educational Psychology 33 ,  
September, 1942. Educational and Psychological Measure-  
ment 10, 1950.
7. Davies, Alan. Language Testing Symposium. A Psycholinguistic  
Approach. Oxford University Press, 1968.
8. Ebel, Robert L. Writing the Test Items. Ch. 7 in E. F.  
Liddgiust (Editor) Educational Measurement Washington  
D.C. , American Council on Education , 1951.
9. Hawkes, Herbert E, E.F.Lindgiust and C.H. Mann. The Construction  
and Use of Achievement Examinations. Boston, Houghton  
Mifflin Company, 1936.
10. Hawkins, E.W. Communication and Language The Importance of  
the early years. ( Mimeographed ).
11. Hornby, A.S., E.V.Gatenby and H.Wakefield. The Advanced  
Learner's Dictionary of Current English. Second Edition  
The English Language Book Society, Oxford University

Press, 1963.

- Kandel, Isaac L., and Edward L. Thorndike. Educational Measurement of Fifty Years Ago. Journal on Educational Psychology 4, November, 1913.
13. Kelly, Truman L. Interpretation of Educational Measurements Yonkers on Hudson, New York, World Book Company 1937.
14. Kuder, G.F. and M.V. Richardson. The Theory of the Estimation of Test Reliability. Psychometrika, 2, September 1937.
15. Lado, Robert. Language Testing. New York - Mc. Graw Hill, Book company Inc., 1964.
16. Lee, J. Murray and David Segel. Testing Practices of High School Teachers. United States Offices of Education Bulletin, No 9, 1936.
17. Lindquist, E.F. Educational Measurement. 1951, published by the American Council of Education Washington, D.C.
18. Myers, M. Claire - Charles I Mosier and Helen G. Price. Suggestions for the Construction of Multiple Choice Test Items. Educational and Psychological Measurement, 5 Autumn 1945.
19. Ross, C.C. and revised Julian C. Stanley. Measurement in Today's Schools. Third edition, Englewood cliffs, N.Y. - Prentice - Hall, Inc. - 1959.
20. Ruch, G.M. The Objective or New Type Examination Chicago, Scott, Foresman & Company 1929.
21. Smith, B. Othanel. Logical Aspects of Educational Measurement New York : Columbia University Press, 1938.
22. Starr, V.H. Proficiency Tests in Modern Foreign Languages . Publications of the Modern Language Association of America Vol. LXXVI No 2.  
Foreign Language Proficiency Tests for Teachers and Advanced Students. Publications of the Modern Language Association of America Vol. LXXVII No 4.



egory being, 7, 24, 38, 24, 7 or 10, 20, 40, 20, 10 respectively . Working on grading on a Normal Curve, therefore, can determine the percentage of students falling in each grade. This method is sometimes referred to as "grading on the curve".



To assume, that at least 60 is the minimum passing grade , leads to the fact that the test tend to be too easy and therefore, loosing its discriminating power. It is not that 60% of the items forming the passing score that must be considered foremost. The average score must be put about 50% of the items and then the grades must be set in terms of percentage. However, the teacher must not stick too rigidly to any fixed rules for establishing grades.

Therefore, it is not necessary that 7% of the class should get the lowest mark. It depends entirely on the capacity and ability of the students .

\* \* \* \* \*

## CHAPTER VI

### CONCLUSION

Most teachers would agree that tests are necessary in order to determine how much a student knows and whether the teaching has been effective or not. Without some assessment neither the teacher nor the student knows the amount of progress that has been made.

It is therefore of vital importance that teachers should know what testing means and the principles involved in the construction and evaluation of tests.

The test is indispensable for the :

1. teacher.
2. student.
3. test itself.
4. administration or statistics.

1. For the teacher :

- a. to see the result of his teaching.
- b. to see how far the students have learned.
- c. to see whether the students are ready for new lessons.
- d. to see which lessons need reviewing.
- e. to use as a guide to give grades.
- f. to find out the students' individual difficulty.
- g. to determine which students should pass, get scholarship or be promoted.

2. For the student :

- a. to know his capacity compared with the other students in the class.
- b. to compare with the former results.
- c. to encourage him to study regularly and to review the lessons.

3. For the test itself :

- a. to find out whether the test is too difficult or too easy.
- b. to find out whether it discriminates the bright students from the less bright or not.
- c. to find out whether the test represents the lessons covered or not.

4. For the administration or statistics :

to find out the progress of the students in one school or area with that of the other schools or areas or with that of the former or the term and annual reports.

To conclude, we do hope, that teachers of English at our secondary schools should show ingenuity in composing their test items and have critical attitude in judging them.

\*\*\*\*\*

