Chapter IV

T E S T C O N S T R U C T I O N

(A) Planning The Test

Economy that planning ensures can hardly be exaggerated. The present investigation also directed big efforts towards a cereful planning, before actual work of designing the test started. This enabled the investigator to anticipate difficulties and problems and, so, to avoid or to minimize them.

Checklist-Questionnaire to Locate Components of Language Ability

(a) The present test is designed to measure an individual's language ability. The first requisite of a good test is that it should measure what it is meant to measure. This is not possible, unless the test designer is quite clear about what language is and what language ability means. For the first, he read a considerable amount of literature on the topic,

specially books on linguistics, and gained clarity about the nature and functions of a language. Regarding the second, he followed a more objective and scientific line of action. On the basis of his reading, he prepared a checklist-questionnaire which listed about 10 component abilities that might comprise language ability. He sent the checklist-questionnaire to 25 persons of eminence in the field of Gujarati linguistics and literature. (The list is given in appendix 1.) They were asked to check those abilities that they believed to be entering the general language ability. They were also asked to put weightage that, they believed, should be given to each component ability if the total test items were 100. The checklist -questionnaire is given in Table: 4.1.

Table : 4.1

A CHECKLIST-QUESTIONNAIRE

given to Experts in Linguistics and Language Teaching to Elicit Their Opinion regarding the Components of Language Ability and Their Relative Weightage.

Area:	Component Trait	How many items (out of 100) should be devoted to it?
(A) Comprehension	(1) Comprehension of a prose passage	
	(2) Comprehension of poetry	

Table : 4.1

A CHECKLIST-QUESTIONNAIRE

(Contd.)

Area	Component Trait	How many items (out of 100) should be devoted to it?
(B) Expression	(3) Vocabulary (a) Synonyms (b) Antonyms (c) Proper use of words	
	(4) Meaning and proper use of idiomatic phrases and proberbs	
	(5) Language items involving formal grammatical points	
	(6) Sentence patterns and structures	
	(7) Presentation and style	
	(8) Paragraphing	
(C) Orthography	(9) Spelling	
	(10) Punctuation	
(D)	Please suggest any other component ability that is left out and indicate its weightage. (1) (2) (3) (4)	
	Total ...	100

The investigator has designed the test on the basis of the theoretical discussion about language ability (as it appears in the previous chapter), as well as on the basis of the data collected through this checklist-questionnaire.

But this much would not suffice. One who attempts the practical job of test construction must also be clear about the principles regarding the following aspects of the work :

(a) Description of the population, that is, range of applicability.

(b) The aspects to be measured, that is, the sample of behaviour to be tested.

(c) The way it is to be measured, that is, the principles for writing test items.

The following sections describe how these problems are taken care of in the present study.

(b) Defining The Range of Applicability for The Present Test

The population to be tested needs to be precisely defined before the investigator commences to design the actual test items. It is essential because the related phases of test designing have direct bearing on the population to be tested.

Human abilities are measured through their manifestations. Therefore, modern psychologists differentiate between 'innate ability' and 'tested ability'. Human abilities as measured by psychological tests have to take into account the influence of training or experience. The language ability test measures language ability as it has developed upto this date - due to the interaction of innate ability and post-natal experience. The environmental influences that shape the innate ability for language acquisition are so many; but they can be grouped as follows:

1. schooling,
2. age,
3. cultural background,
4. sex.

It is hardly possible to isolate the influence of these forces. It is almost impossible to find performances and responses in which all the individuals who are to be measured have had equal opportunities for training, or equal acquisition, or in which training is a negligible factor, and that is not necessary also. The investigator has tried to control the abovementioned variables in the following manner.

1. Schooling: The present test was intended to be made applicable to students - both boys and girls-

who enter college after passing their S.S.C. examination. All who pass the S.S.C. examination do not enter college; but those who enter college are schooled in the same language content as those who do not do so. From the view of content it would be perfectly admissible to design the test on the basis of the content course prescribed for the S.S.C. students. It would be the quantum of content that a student, after 11 years' schooling, would be expected to master. The test is specifically meant for the college-entrants, and is meant to be administered when the student has just entered college, or just before his admission to the college. Here a college means a college of any of the three major faculties, situated in either rural or urban area. As the college-entrants are admitted to the pre-university class, the population here means students of Pre-University Arts, Pre-University Science and Pre-University Commerce classes.

The investigator's chief aim in constructing this test has been to develop an instrument which sorts out students worthy to be admitted in college from those who are not - on the basis of language ability; it has also been to sort out students worthy to be admitted in Arts, Science and Commerce faculties on the basis of the minimum language ability these

different faculties require of their students while they enter college.

Hence content has been selected in such a manner as would give proper weightage to the differential patterns of future experiences that Arts, Science and Commerce students might be expected to undergo after their entrance into college. This is specially true with regard to vocabulary and content material for reading comprehension and expression. Proper care has been taken by the investigator in selecting items from all spheres of life and related to all the above-mentioned faculties. Special care has been taken so that the test does not get loaded with or biased towards any particular sphere of life or branch of study.

Even after doing this, if different levels of language ability are required for successful and effective functioning in different faculties of university, that would be reflected in different norms for these different faculties. This has been one of the hypotheses intended to be tested in this study.

2. Age: The test is designed for college-entrants. Now, as the position prevails in Gujarat, a student entering college could be of any age between 15^{+} yrs. to 19^{+} yrs. But most of the students enter

college at the age of 16^+ yrs. or 17^+ yrs. So special age norms are not developed.

3. Cultural Background: Students entering college differ widely in their cultural background and a sizeable degree of difference among their levels of language ability might be due to this reason. The differences found in different cultures - urban, town and rural; higher caste and lower caste; upper-income, middle-income and lower-income groups; Savarnas and Adivasis etc. - can not be overlooked. They are more strongly revealed when language is considered. The investigator has been fully aware of these cultural differences and has provided for that by properly selecting the representative sample through the stratified random sampling. So far as the content selection was concerned, those stimuli-responses were selected which were common to all these cultural groups. Thus cultural differences were taken care of, while selecting the content of the items, and while selecting the sample.

4. Sex: As in the case of differences found in cultural background, the problem of sex differences also demands attention. The investigator has a feeling that sex differences affect language ability. So he has tried to maintain in the sample the proportion

obtaining between the two sexes in the total population - that is, total number of students entering college in Gujarat. Moreover, sex-wise norms are also developed.

One of the crucial aspects of mental measurement, namely, restricting the range of individuals to be tested has been fully considered above. This takes us to the next one, namely the selection of activities which are largely common to their experience.

(c) Sampling the Behaviour to be tested

(i) Need for sampling the Behaviour

Anastasi (1954) has defined a psychological test as "essentially an objective and standardized measure of a sample of behaviour". The mental ability to be measured is bound to manifest through various behaviours or various responses to stimuli or various performances. The tester can not and need not measure them all. While discussing the sampling of traits and functions, Freeman (1962) suggests,

> In developing a psychological test, it is impossible, and in fact unnecessary, to use an unlimited number of items. It is not necessary to attempt to present the 'subject'

> or the 'testee' with problems that will ascertain his responses for every conceivable situation involving a given trait or function. It is sufficient to get an adequate sampling of responses in a particular area or range of behaviour, the assumption being that the sampling is representative of the whole.

But even after accepting this, the fact remains that the test should be comprehensive enough to sample all representative modes of behaviour. Insufficient coverage of behaviour to be tested will decrease the validity of the test. So even while restricting the number of items, the investigator has taken pains to see that all representative modes of behaviour are sampled.

A testmaker has to be modest in his claims and should caution the test-users to interpret the norms after adjusting them for a margin of sampling and measuring errors. At the same time, the tester should do his utmost to make his measurement accurate and precise, taking due precautions at every stage of test development.

(ii) Review of the Literature on Language Ability (Aptitude) Testing

Considerable work has been done in this sphere in western countries, while similar work in India is rather limited. Even the few tests on language aptitude that already exist cover only one or two aspects of the total, general trait. The investigator had mostly to rely on his own concept of language ability, in this position of lack of good external criteria. Yet he has tried to correlate his test with the Language Ability Test for High School Students constructed by Mrs. Urvashi T. Desai and has found the correlation to be sufficiently high. The investigator has also referred to the Reading Comprehension Test of Dr. Jamnadas A. Bhagatwala.

The investigator has also referred to the Verbal Ability Test of the DAT battery and incorporated its salient points in his test. The investigator thinks nothing wrong in doing so, because though every language has its own uniqueness, there is much in common in all languages. He has also referred to the abstracts of six other American verbal ability tests, which are summarized in Chapter I.

With this background the investigator attempted to know how previous testers had viewed the problem.

The following types of relevant literature on test construction were studied:

1. Some original language aptitude and language ability tests.
2. Review of tests in the Mental Measurements Yearbook.
3. Description of tests in books on psychological testing.

This literature was studied, keeping in mind the following aspects of testmaking:

1. The language skills measured.
2. Content of the tests and relative occurance of different types of items.
3. The variety of ways in which the same content is evaluated.

(iii) The study of the test content revealed that the tests measuring language ability covered a vast area. Some of the frequently used topics are as follows:-

1 Vocabulary
2 Word building
3 Choosing the right word
4 Idiomatic usage
5 Verb usage
6 Grammatical usage
7 Sentence structure
8 General comprehension
9 Paragraphization
10 Effective expression
11 Spelling
12 Punctuation and capitalization.

On the basis of the study of the above-mentioned literature, the investigator prepared a checklist-questionnaire which is already mentioned before. The construction of items and deciding their relative weightage were done on the basis of :

(a) the study of other language ability tests.

(b) the data collected through the checklist-questionnaire.

The list of language areas covered in this test is as follows:

Part I : Vocabulary

1. Synonyms
2. Antonyms
3. Correct use of words
4. Sandhi (Euphonic combination)
5. Idioms and usages
6. Proverbs with similar meanings
7. Proverbs with opposite meanings
8. Meaning of proverbs

Part II : Sentence Structure

1. Sentence structure
2. Grammatical elements in sentence structure
3. Organization of sentences in a coherent narration

Part III : Spelling and Punctuation

1. Spelling
2. Punctuation
3. Paragraphization

Part IV : Comprehension

1. Reading comprehension (with the content sheet available at the time of answering)
2. Eliciting central ideas of a passage
3. Reading comprehension (with the content sheet not available at the time of answering)

The blueprint for the test, which was prepared on the basis of this area-analysis, is given below:

BLUEPRINT

Ability to be tested	Approximate Number of Items in the Try-out Version	
	Multiple Choice	Matching
Part I		
1. Ability to recognize synonyms	90	--
2. Ability to recognize antonyms	10	75
3. Ability to recognize incorrect use of words	5	--

BLUEPRINT
(Contd.)

Ability to be tested	Approximate Number of Items in the Try-out Version	
	Multiple Choice	Matching
Part I (Contd.)		
4. Ability to recognize the correct 'Sandhi' (euphonic combination)	15	--
5. Ability to recognize the correct meaning of an idiom	10	20
6. Ability to recognize the correct meaning of a proverb	5	--
7. Ability to recognize proverbs with similar meaning	--	20
8. Ability to recognize proverbs with opposite meaning	3	--
Part II		
9. Ability to recognize the correct structure word	32	--
10. Ability to recognize structurally-correct sentence	20	--
11. Ability to arrange jumbled sentences in lexically logical order	54	--

BLUEPRINT

(Contd.)

Ability to be tested	Approximate Number of Items in the Try-out Version	
	Multiple Choice	Matching
PART III		
12. Ability to recognize the correct spelling of a word	150	--
13. Ability to recognize the correct letter which, when supplied, would result in correct spelling	20	--
14. Ability to recognize the correct punctuation mark	10	--
15. Ability of paragraphization	2	--
PART IV		
16. Ability to recognize the correct response to a stimulus question intended to test comprehension of a prose passage	14	6
17. Ability to summarize a prose passage into a number of fixed words (recognition)	10	--
18. Ability to comprehend a prose passage without the content material in hand for reference at the time of answering	13	--
	463	121
	584	

(iv) Some clarifications

The following points have been clarified in this section:

1. Why has the investigator decided to have a battery of four tests?
2. Why is the same component ability tested in a variety of forms?
3. Why is the number of test items so large?
4. Justification for the inclusion of some of the above-listed component abilities.
5. Justification for the non-inclusion of some component abilities.

(1) A test battery is a series of interrelated tests standardized on the same sample. When a tester feels that some mental abilities are partially interrelated, but not fully interrelated so as to be considered as one ability and to be included in one general test, he constructs a separate test for each of these abilities. Then he standardizes these tests on the same sample so that their separate norms can be compared and, if need be, merged to make an integrated distribution of scores. The patent fact here is the partial uniqueness and partial interrelation of sub-abilities. So the component abilities are not completely distinct and at the

same time are not completely identical. Another important fact is that the abilities are measured on the same sample. Consequently comparability and integration of norms is secured.

In this test, the investigator has felt that the four main sub-abilities comprising the group factor called language ability - namely command over vocabulary, language structure, spelling and punctuation, and comprehension - are distinct and yet highly interrelated. If language ability is to be properly measured, four distinct scores have to be considered, rather than one general score; and yet, to get a comprehensive picture of general language ability, the four scores have to be integrated.

It is due to this consideration that the investigator has chosen to have a test-battery, consisting of four distinct tests.

(2) It will be seen from the perusal of the test that in some cases the same ability is tested in two different ways. This variety of modes of testing is adopted for three reasons:

(a) To bring variety and change in the method of testing, that is to break monotony of the testing tool.

(b) To bring in the benefit of rotational technique in testing; suppose a particular item of knowledge is favourably or unfavourably related to a particular type of testing technique, then the result would be biassed. To avoid this, the same ability was tested in two different ways.

(c) To provide check for guess-work and chance response.

(3) The final version of the test consists of 400 items. The number might seem rather large. But two considerations have entered in keeping it so large.

(a) Actually this is a test battery. The final test-form is divided into four sub-tests or parts. Part I consists of 180 items, Part II of 73 items, Part III of 111 items and Part IV of 36 items. The four parts can be administered at different times, or on different days. The total mean time for all the four tests is 170 minutes; but the investigator administered the four parts separately. The mean time for each of the four parts separately is as follows:

Part I	:	45 mts.
Part II	:	40 mts.
Part III	:	40 mts.
Part IV	:	45 mts.

With this division, the test length seems to be well justified.

(b) The general ability that goes by the name of language ability is very comprehensive. It is a complex of so many sub-abilities. This can be seen by the list of component abilities given in the blueprint. Even a component trait like vocabulary has many facets. Again, this is a test for measuring the language ability (aptitude) in the area of the mother-tongue, that is L_1. Naturally the students have gained a comprehensive control over a wide linguistic area by the time they pass the S.S.C. examination and are about to enter college. The investigator wanted this test to be comprehensive enough to cover as much ground as possible and not to leave any important component ability uncovered. Consequently, such a large number of items are included in the test. But the investigator believes that, because the test is divided into 4 parts which can be administered as separate times, the number of items will not detract from the effectiveness of the test. On the other hand, it is likely that the largeness of number of items will add to the content validity of the test.

(4) The investigator wanted the test to be fairly inclusive of as many essential components of language

ability as possible. From that view point, he has included in the test those component traits which, he thought, were basic to the structure of language ability in Gujarati.

As for example, a student, who has a good command over spoken and written Gujarati, must be well-versed with synonyms and antonyms in the language. In the same way, correct use of a word - proper word in proper context - is surely a good indication of command over a language. As with other languages, Gujarati also abounds in idioms, usages and proverbs. If the student can understand the correct meaning of different idioms, usages and proverbs, can recognize their correct use, and can also recognize idioms and proverbs with similar and opposite meanings, that is a fairly good indicator of his language ability. All these component capabilities might be said to comprise the sub-area called vocabulary.

Sentence-patterns or sentence-structure is the second major area in language-ability. Gujarati is the mother-tongue (L_1) of the students. Most of its basic structures are already assimilated by the student orally, before he enters school. But during the eleven years that he spends in school, his command over these basic structures gets consolidated through frequent

use in speech and writing and reading. He also learns some more advanced and complex structures of Gujarati. Any test that claims to measure language ability must test the student's mastery of the use of structures - especially its subtle points. The correctness of patterns, grammatical points involved in these patterns, word order in a sentence, syntax - all these should be tested. Along with these, the student's capacity to put into order a set of jumbled, disorganized sentences should also be tested.

The investigator has always felt that style and presentation are an integral part of language ability. So he included some tests on style and effective, appropriate presentation in the pre-tryout version. But the experience of pre-tryout showed that the items were either too difficult or became too subjective to be reliable. So in the try-out version he eliminated those tests. But still his opinion stands that some tool should be constructed to measure objectively style and presentation. Absence of such a test in the final version is one of the limitations of this test-battery.

Spelling is a very important component ability in language-learning. So is punctuation. To divide the language material into proper paragraphs is also an

important ability. Hence tests on all these traits have been included.

Lastly, comprehension - accuracy and speed of intensive reading - is undoubtedly an important component of language ability.

(5) In the preceding section, the reasons for including certain component abilities in this test have been discussed. In this section are discussed considerations that led to the omission of certain other component traits.

This is a paper-and-pencil test. Hence oral expression, however important it may be from the linguistic point of view, has not been covered by this test. Similarly listening comprehension is also beyond the scope of this test, though the investigator is well aware of its importance in language learning. Expression, as it gets manifested in the written form, and reading comprehension are the two areas covered by this test. This is one of the limitations of this test and it is imposed by the inherent nature of a paper-and-pencil test as such.

Secondly, about 80 % of the test-items are of multiple-choice type; but some of them are of matching type also. As far as possible, at least three options are given as plausible responses to a stimulus question;

but at some places, it has not been possible to provide three distractors and hence, two-distractors items will also be found in a few sub-tests.

Thirdly, as stated in the preceding section, tests on style and presentation have not been included in the try-out and the final version, though they found a place at the pre-tryout stage. Style and presentation are rather difficult to measure objectively and precisely. There is a danger that the judgments of several judges as to the correct response out of several ones might differ considerably.

Lastly, formal grammar has found very little place in this test. It is because grammar itself is not language and language ability might not have much relation to the knowledge of formal grammatical rules. Instead, grammar points are covered as structures.

To summarize:

1. The choice was restricted to the areas which could be measured accurately with a testing tool.

2. The selection of skills was in accord with the fact that this was meant to be a group test and a paper-and-pencil test. Consequently, areas such as literary appreciation, paragraph building, literary stylistics, etc., though important as components of language ability, were not included. Again this was

to be a language ability test and not a test of literary aptitude.

3. Keeping in view the nature of mental measurement, areas which required specific training were not much stressed. Among them are formal grammar, literary history and literary form.

4. Selection of areas were made so as to make the measurement extensive rather than intensive.

It can be concluded that two parallel considerations of validity and practicability have guided the test construction. A fair and honest attempt has been made to select suitable and purposeful content with a view to making the instrument valid and useful.

Following the suggestion of Vaugham (Lindquist, ed. 1951), the content has been used rather broadly to cover both the subject matter of the test and the types of abilities that, it is thought, the students should possess. Sometimes these two aspects of the test content can conveniently be treated together; but sometimes, the specific content might have to be separated from the type of behaviour to be tested. So, the behaviour specifications are given below:

(d) Choice of behaviours to be tested

Vocabulary (Figures in the brackets indicate the test-number in the try-out version.)

1. The student recognizes the correct synonym of the given word. (1.1)
2. The student recognizes the correct antynym of the given word. (1.2-a) (1.2-b)
3. The student locates the word improperly used in a sentence. (1.3)
4. The student recognizes the correct dissolution of sandhis (euphonic combination). (1.4)
5. The student recognizes the correct meaning of a given idiom or a given usage. (1.5-a) (1.5-b)
6. The student recognizes the correct meaning of a given proverb. (1.6-a)
7. The student recognizes the proverb having meaning similar to a given proverb. (1.6-b)
8. The student recognizes the proverb having meaning opposite to a given proverb. (1.6-c)

Sentence Structure, Expression, Organization

1. The student recognizes the correct word that would, when supplied to fill up the blank, make the sentence structurally correct and complete. (2.1-a)

2. The student recognizes the sentence which is correct structurally and grammatically. (2.1-b) (2.1-c)
3. The student arranges a number of sentences in order to make the sequence proper from the view point of coherent, systematic presentation and organization. (2.2-a) (2.2-b) (2.2-c) (2.2-d) (2.2-e)

Spelling, Punctuation and Paragraphization (Orthographical skills)

1. The student recognizes the correct spelling of a word. (3.1-a)
2. The student recognizes the correct letter, which, when supplied to fill the blank, would form a correct meaningful word (spelling). (3.1-b) (3.1-c)
3. The student recognizes the correct punctuation mark for a given position in a sentence. (3.2)
4. The student divides an unparagraphed group of sentences into suitable paragraphs. (3.3)

Comprehension

1. The student recognizes the correct detail, or fact, or meaning as a response to a stimulus (question asked), on the basis of his reading of a paragraph. (4.1-a) (4.1-b) (4.1-c)

2. The student recognizes the word that would be a correct heading for a given paragraph. (4.2-a)
3. The student recognizes some key words in a given paragraph that would be useful in telling the central idea of the passage. (4.2-a) (4.2-b)
4. The student, without having the content sheet with him at the time of answering, recognizes correct detail, fact or meaning on the basis of having read a paragraph a few minutes ago. (4.3-a) (4.3-b)

CONSTRUCTION OF TEST ITEMS

So far, the types of content as well as abilities thought essential were considered. The next phase was to make decisions regarding the test length and the types of items. These two are important factors which directly govern the writing of test items.

(a) Test Length

The test length depends upon the purpose of the test and the statistical characteristics of the items. The test must contain a large number of items, if

(1) important decisions regarding individuals are to be based on the test;

(2) part scores are to be obtained; and

(3) the items of the test are heterogeneous.

All the three situations listed above were applicable to this test, so the investigator decided to make the test fairly long and comprehensive.

Another factor which influences test length is the amount of time likely to be available, and also the optimum time length to avoid any setting in of fatigue. Convenience of administration is also a factor to be considered in this respect.

As this was a test to be administered on college-entrants, the average duration of 40 to 45 minutes for each one of the four tests was quite practical from all the above-mentioned considerations. College-periods are generally of 45 minutes' duration. As mentioned above, the test has been divided into 4 parts. Here, the possible rejection of some items as a result of item analysis was also accounted for, and for the try-out version four tests that might take about 50 to 60 minutes each were constructed. 584 items were included in the try-out version; after rejecting 184 items on the basis of item analysis, the investigator retained 400 items in the final version.

(b) Selection of Types of Items

Item writing is an art. It requires an uncommon combination of special abilities on the part of the test-maker. The process is essentially creative. Each item, as it is being written, presents new problems and new opportunities. Ebel (Lindquist, ed. 1951) comments,

> Just as there can be no set formulas for producing a good story or a good painting, so there can be no set of rules that will guarantee the production of good test items.

Principles can be established. Suggestions can be offered. But above them all, it is the item-writer's judgment that determines the quality of items.

Ebel (Lindquist ed. 1951) has distinguished two major types of objective test items: the 'supply' items and the 'selection' items. He has further divided the selection type into the following kinds:

1. Constant alternatives, that is items requiring choice between constant alternatives.
2. Changing alternatives (multiple choice), that is items requiring choice among changing alternatives.

3. Matching which might consist of two types: simple matching and compound matching.

Ordinarily at a fairly early stage in planning a test, before test construction is started, decisions must be made concerning the types of items the test-maker should include in a test. In the present case the factors worth considering were:

i. relation to scoring
ii. relation to administration facilities and
iii. use of few versus many types of items.

(i) Relation to Scoring

One of the characteristics of standardized measurement is that of objectivity of scoring. The phase of test construction was taken care of even at the stage of selecting the specific skills to be measured. The skills and abilities that could not be measured precisely through objective test-items were discarded, just to avoid subjectivity. Even in the areas that are usually difficult to be measured through objective test items, an attempt has been made to develop objective techniques to measure them, rather than discard them, and hence it has become possible to include some such areas. Objective scoring is very essential for reliability. As Vaughan (Lindquist ed. 1951) says,

> .... Among the various forms of so-called tests, the more objective, such as the 'multiple-choice' are to be preferred to the less objective, such as the 'completion'.

(ii) Relation to Administration Facilities

In the above section the problem of scoring facilities was taken care of. In addition to that, the probable administration facilities also have a bearing on the selection of the types of items. To minimise the difficulty of administering the test requires precautionary measures, specially when the test is designed for wide use, and is likely to be administered by different persons. In such a case, objective-type items are very suitable, because instructions for every exercise are given in the beginning, and the testees are not required to write anything in the test booklet except to make a tick-mark or draw a line or write a number.

(iii) Use of Homogeneous versus Heterogeneous Types of Items

The advantages enjoyed by the heterogeneous type, namely the sustenance of interest through variety of forms, and facilitation in terms of testing the content in a form which readily suits it, are specially

noteworthy. So the investigator has included two different item-types. But, at the same time, he has taken care not to alter the type too often.

The type of items is changed only after allowing proper time. The advantages of the system of homogeneous types are that, it induces a mental set up in the testees for a particular type of testing-pattern, and that it is specially suited to subjects of subcultural levels. Interest depends both on the content as well as on the form of the item.

So, after duly considering all the above-mentioned aspects, the investigator decided to construct a heterogeneous type of test having items of both multiple choice as well as matching variety.

(c) Guidelines observed while writing Test Items

For writing good test items, the following measures are taken:

1 As far as possible, uniformity in the nature of eliciting item response is maintained; and when diversity is introduced, it is never at the cost of advantages accrued through uniformity; and only at a point when uniformity might show diminishing returns.

2 As far as possible, four alternatives are given for each stimulus; in some items, however, the number of options is three, as to find the fourth one was rather impossible. Even in the matching type, the response column has three alternatives more than the stimulus column.

3 Correct responses are arranged in random order.

4 The distractors are made as attractive and plausible as possible.

5 In any one subtest, all the responses are made grammatically consistent and uniform in sentence-construction.

6 The investigator has got all the test items checked by experts to the effect that only one response (out of 3 or 4 alternatives given) is correct.

7 The correctness of the responses has also been checked through individual try-out.

8 Psychological points such as motivation and interest are considered at the time of item-construction.

Thus, the construction of items for the present test was based on a thorough understanding of the above-mentioned principles.

Additional Care Taken
for Ensuring Good Items

But occasions arose when the present investigator found it difficult to get good alternatives for an item, or to find out a good stimulus that might test a particular content-bit or ability. At such times, the investigator resorted to the following ways:

1 taking the expert help.

2 administering the items to some individuals of the normative population in a free response form. (This would provide the test constructor with really plausible alternatives.)

3 At times, a test-item or a subtest constructed to measure a particular ability failed to do so during the pre-tryout. At that time, the test-item or the test was modified.

The pre-tryout version was administered on 100 students in June, 1970. On the basis of its experience, the try-out version, having 584 items was prepared and administered on 370 students in June, 1971. Item analysis was done in one month and the final version having 400 items was administered in July and August, 1971. The norms and correlations have been calculated on the data of this final administration.

(d) Description of the Abilities measured by the Sub-Tests included in the Try-out Version

Table : 4.2

Test No.	Items	Ability tested by it	Nature
1.1	90	Ability to recognize synonyms.	Multiple choice.
1.2(a)	75	Ability to recognize antonyms.	Matching.
1.2(b)	10	Ability to recognize antonyms.	Multiple choice.
1.3	5	Ability to recognize incorrect use of words.	Multiple choice.
1.4	15	Ability to recognize the correct Sandhi.	Multiple choice.
1.5(a)	10	Ability to recognize the correct meaning of an idiom.	Multiple choice.
1.5(b)	20	Ability to recognize the correct meaning of an idiom.	Matching.
1.6(a)	5	Ability to recognize the correct meaning of a proverb.	Multiple choice.
1.6(b)	20	Ability to recognize proverbs with similar meaning.	Matching
1.6(c)	3	Ability to recognize proverbs with opposite meaning.	Multiple choice.
2.1(a)	32	Ability to recognize the correct structure word.	Multiple choice.
2.1(b)	5	Ability to recognize a structurally correct sentence.	Multiple choice.

Table : 4.2
(Contd.)

Test No.	Items	Ability tested by it	Nature
2.1(c)	15	Ability to recognize a structurally correct sentence.	Multiple choice.
2.2(a)	11	Ability to arrange jumbled sentences in order, from the view-point of coherent presentation.	Multiple choice.
2.2(b)	11	Ability to arrange jumbled sentences in order, from the view-point of coherent presentation.	Multiple choice.
2.2(c)	11	Ability to arrange jumbled sentences in order, from the view-point of coherent presentation.	Multiple choice.
2.2(d)	10	Ability to arrange jumbled sentences in order, from the view-point of coherent presentation.	Multiple choice.
2.2(e)	11	Ability to arrange jumbled sentences in order, from the view-point of coherent presentation.	Multiple choice.
3.1(a)	150	Ability to recognize correct spelling of words.	Multiple choice.
3.1(b)	10	Ability to recognize the correct letter to make the correct spelling of words.	Multiple choice.

Table : 4.2

(Contd.)

Test No.	Items	Ability tested by it	Nature
3.1(c)	10	Ability to recognize the correct letter to make the correct spelling of words.	Multiple choice.
3.2	10	Ability to recognize the correct punctuation mark.	Multiple choice.
3.3	2	Ability to recognize the correct juncture where a new paragraph should begin.	Multiple choice.
4.1(a)	5	Ability to recognize the correct response to a stimulus question testing comprehension.	Multiple choice.
4.1(b)	11	Ability to recognize the correct response to a stimulus question testing comprehension.	Multiple choice and Matching.
4.1(c)	4	Ability to recognize the correct response to a stimulus question testing comprehension.	Multiple choice.
4.2(a)	6	Ability to summarize a prose passage.	Multiple choice.
4.2(b)	4	Ability to summarize a prose passage.	Multiple choice.
4.3(a)	4	Ability to comprehend a prose passage without the content material in hand for reference while answering.	Multiple choice.

Table : 4.2
(Contd.)

Test No.	Items	Ability tested by it	Nature
4.3(b)	9	Ability to comprehend a prose passage without the content material in hand for reference while answering.	Multiple choice and Matching.
	584		

(e) Arrangement of Items in the Sub-Tests

After having selected the types of tests, after having fixed up the number of items to be included in each type, and after having constructed the items, the testmaker has to arrange them before administering them. Two methods for the arrangement of items are prevalent in psychometric practice - the Discrete and the Spiral Omnibus.

In this test, the discrete method is selected. A sub-test contains items that have the same form and measure the same component ability. Within that sub-test the items are arranged according to ascending difficulty value. The second sub-test contains items that measure another component ability; and so on. For each sub-test, there are different instructions and illustrations.

Each part of this test battery is in fact a group of sub-tests; norms are calculated for each part, and not for each individual sub-test; for example norms are calculated for Part I (Vocabulary), Part II (Sentence Construction) etc., but they are not calculated for the sub-test 1.1(a), or 1.1(b) that is synonyms, antonyms etc. Time limit is fixed for the whole Part I or the whole Part II and not for individual sub-tests. Thus it would seem that the method selected for the organization of items in this investigation is a synthesis or combination of both the discrete and the spiral methods.

(f) The Separate Answer Sheet

The subjects for this test were hardly going to be confused by separate answer sheets, especially after detailed and careful instructions. Having separate answer sheets has many advantages such as:

(a) Hole-board scoring (punch card scoring)

(b) Economy in costs, because the same test booklets can be used over and over again.

So it was decided to have separate answer sheets.

Thus the test was constructed; items were arranged. The next stage was that of try-out. It is described in the chapter that follows.

REFERENCES

1. Anastasi Ann — Psychological Testing, New York, Macmillan, 1954. (P.22)

2. Freeman F. S., — Theory and Practice of Psychological Testing, New York, Holt, Rinehart and Winston, 1962. (P.4)

3. Lindquist E. F. (ed.), — Educational Measurement, Washington, American Council of Education, 1951. (Pp.161, 172, 185)