Stages of Tool Construction

- Stage I-Planning The Test
- Stage II Test Construction
- Stage III- Review, Try Out, Item Analysis

Stage IV

Standardization

Stage I
Planning The Test

1.	2.	3.
• Identification of Purpose, • Objectives Scope	• Types & format of the questions • Mode of Scoring	• Preparation of Blue Print

Planning

Make a list of teaching objectives.

Classify them according to domain & level.

Determine the relative importance of each objective.

Devise & set up learning experience which will accomplish the objectives,

Select evaluation procedure appropriate to the domain & level .

Planning

Nature of test must reflect it's purpose

&

Conditions under which it will be administered.

General Objectives

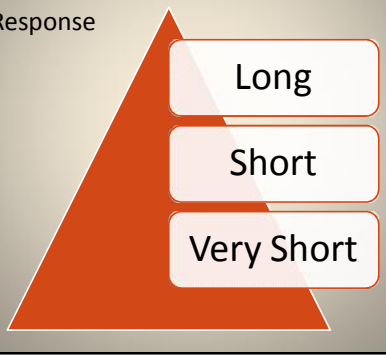
- Developing effective thinking
- Developing useful work-habit
- Learning social attitude
- Acquiring significant interest
- Developing appreciation for art & literature
- Developing Social sensitivity
- Developing social & personal adjustment
- Developing a philosophy of Life

Specific Objectives

- Knowledge of facts
- Technical Terms
- Learning problems &methods
- Capacity to generalize, confirm & reject hypothesis
- Capacity to understand cause & effect relationship
- Development of scientific attitude

Types & format of Questions

Free Response



Fixed Response

Recall Type(Supply)

Fill in The Blanks

Recognition Type(Selection)

1-Double Choice-True False

Yes/No

2-Multiple Choice

3-Matching Type

Types of Test

- 1- Achievement
- 2.Diagnostic
- 3.Aptitude
- 4.Proficiency
- 5.Placement
- 6.Admission

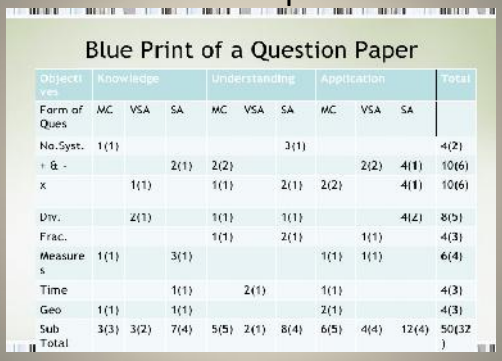
Source Material

- Text Books
- Note Books
- Questions at the end of each chapter
- Class Notes
- Class Discussions

Arrangement of Items

- Equal Difficulty Plan
- Increasing difficulty Plan
- Spiral Plan
- Logical Combination
- Heading Sub- heading

Blue Print Preparation



Blue Print of a Question Paper

Objectives	Knowledge			Understanding			Application			Total
Form of Ques	MC	VSA	SA	MC	VSA	SA	MC	VSA	SA	
No.Syst.	1(1)					3(1)				4(2)
+ & -			2(1)	2(2)				2(2)	4(1)	10(6)
x		1(1)		1(1)		2(1)	2(2)		4(1)	10(6)
Div.		2(1)		1(1)		1(1)			4(2)	8(5)
Frac.				1(1)		2(1)		1(1)		4(3)
Measures	1(1)		3(1)				1(1)	1(1)		6(4)
Time			1(1)		2(1)		1(1)			4(3)
Geo	1(1)		1(1)				2(1)			4(3)
Sub Total	3(3)	3(2)	7(4)	5(5)	2(1)	8(4)	6(5)	4(4)	12(4)	50(32)

Stage II
Test Construction

Writing of Items

- Test may include more than one type of Items.
- 50% difficulty Value
- No of items should be three times more then final number
- The items should be phrased so that the content rather than the form of statement will determine the answer. May, sometimes,generally.

- An item should be so worded that its whole content functions in determining the answer rather than only a part of it.
- All items of particular type should be placed together in the test.
- Regular sequence should be avoided(every third ans correct)
- Directions should be written in clear language.
- Each Item should have equal mark.
- No guess work.

Knowledge	Understanding	Application
define	Explain, describe	solve
identify	Interpret, paraphrase	apply
describe	Summarize ,classify	illustrate
label	Compare ,differentiate	modify
list	Discuss ,distinguish	calculate
name	Predict ,associate	change
state	Contrast ,convert	Experiment ,relate
match	demonstrate	demonstrate
recognize	Estimate ,express	construct
select	Identify ,indicate	compute
memorize	select	interpret
recall	translate	choose

Stage III

- Review , Try Out & Item Analysis

Review of Items

Try Out

Normally distributed sample should be selected.

Proper supervision is required.

Instructions should be clear & unambiguous.

- Pre Try out- Small Group of 15-20 people
- S= R- $\frac{W}{N-1}$

Actual Try Out 400 people

Item Analysis

Item analysis is a process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole.

Item analysis is especially valuable in improving items which will be used again in later tests, but it can also be used to eliminate ambiguous or misleading items in a single test administration.

item analysis is valuable for increasing instructors' skills in test construction, and identifying specific areas of course content which need greater emphasis or clarity.

Item Analysis has two parts

- Item Difficulty
- Item Discrimination

Item Difficulty

For items with one correct alternative worth a single point, the item difficulty is simply the percentage of students who answer an item correctly. In this case, it is also equal to the item mean.

- The item difficulty index ranges from 0 to 100; the higher the value, the easier the question.
- It can be assessed by two methods
- 1. Method of Judgment
- 2. Empirical Method

Method of Judgment

- On the basis of expert opinion.
- Rank the items in their increasing order of difficulty
- Total Time taken by examinees.

Empirical Method

- Also called statistical method.
- It determines the index on the basis of responses of all examinees.
- It determines the index on the basis of responses of only a portion of examinees.

Item Difficulty Index

- Is defined as the percentage or proportion of test takers who correctly answer the item.

$$P = \frac{\text{number of examinees correctly answering the item}}{\text{number of examinees}}$$

- For example, in a class of 30 students, if 20 students get the item correct and 10 are incorrect, the item difficulty index is 20/30 =0.67
- Range from 0 to 1

Index of Difficulty

- the proportion of the total group who got the item wrong

FORMULA:

$$P = \frac{R_u + RL}{T} \times 100$$

RU – The number in the upper group who answered the item correctly
RL – The number in the lower group who answered the item correctly
T – The total number who tried the item

Level of Difficulty

Index Range	Difficulty Level
0.00-0.20	Very Difficult
0.21-0.40	Difficult
0.41-0.60	Average/Moderately Difficult
0.61-0.80	Easy
0.81-1.00	Very Easy

Item Discrimination

- The single best measure of the effectiveness of an item is its ability to separate students who vary in their degree of knowledge of the material tested, and their ability to use it.

- The maximum item discrimination difference is 100%. This would occur if all those in the upper group answered correctly and all those in the lower group answered incorrectly .

- Zero discrimination occurs when equal numbers in both groups answer correctly. Negative discrimination, a highly undesirable condition, occurs when more students in the lower group then the upper group answer correctly.

Inter-Item Correlations

- The **inter-item correlation matrix** is another important component of item analysis.
- This matrix displays the correlation of each item with every other item.

- Usually each item is coded as dichotomous (incorrect = 0, correct = 1), and the resulting matrix is composed of **phi coefficients**, that are interpreted much like the Pearson product-moment correlation coefficients.

- This matrix provides important information about a test's internal consistency, and what could be done to improve it. Ideally each item should be correlated highly with the other items measuring the same construct. Items that do not correlate with the other items measuring the same construct can be dropped without reducing the test's reliability.

Item-Total Correlations

- Point-biserial or item-total correlations assess the usefulness of an item as a measure of individual differences in knowledge, ability, or personality characteristic.
- Here each dichotomous test item (incorrect = 0; correct =1) is correlated with the person's total test score.
- Interpretation of the item-total correlation is similar to that of the D statistic. A modest positive correlation illustrates 2 things:

- 1) That the item in question is measuring the same construct as the test;
- 2) The item is successfully discriminating between those who perform well and those who perform poorly.
- In reality, item-total correlation provides the same sort of information that the D statistics does

The formula for discrimination index(D.I)

D.I = (R.H - R.L)/ (N.H or N.L)

- R.H – rightly answered in highest group
- R.L - rightly answered in lowest group
- N.H – no of examinees in highest group
- N.L - no of examinees in lowest group

Levels of Discrimination

Ebel (1986) and Hetzel (1997)

Index Range	Discrimination Level
0.19 and below	Poor item, should be eliminated or needed to be revised
0.20- 0.29	Marginal item, needs some revision
0.30-0.39	Reasonably good item but possibly for improvement
0.40 and above	Very good item

Test Revision

Distracter Analysis

- In distracter analysis, we are no longer interested in how test takers select the correct answer, but how the distracters were able to function effectively by drawing the test takers away from the correct answer.
- The number of times each distracter is selected is noted in order to determine the effectiveness of the distracter. We would expect that the distracter is selected by enough candidates for it to be a viable distracter.

Stage IV

Standardization

Final Try Out- on large Population
Reliability - Test retest
Split half
Equivalences(Parallel)
Internal Consistency(Kuder Richardson)
Validity - Face
Content (Curriculum)
Construct(factorial)
Criterion Related(Concurrent)

Development of Norms

Age Norms
Grade Norms
Percentile Norms

Objectivity

Usability

START YOUR
ACHIEVEMENT TEST PREPARATION

"Knowing is not enough; we must apply. Wishing is not enough; we must do." –
Johann Wolfgang Von Goethe
(German Writer)

Dr.Arpana Godbole
Department of Education
University of Lucknow