Test - Is an instrument used in capturing learning outcome

ECB 2405:

TEST & MEASUREMENT

Test

- A test is an instrument used in capturing learning outcome.
- A test can be conducted at the end of every lesson.
- A test therefore is formative.

Examination.

- A more comprehensive test that comes at the end of a programme.
- It is a summative evaluation, usually high stake.

Measurement.

- Is quantification of an outcome.
- It is a professional value.

Assessment.

- Is describing the status of a phenomenon without placing a score.

Types of Evaluation.

- Are of two types namely:
  - Programme
  - Student.

(1) Programme evaluation.

- Tax payer finance programme ie Education & health centre.
- The work that their tax is doing must be evaluated to find out whether work done is worth money given out.
- Ie in Education, tax payer would want to see whether there is quality in education.

(2) Student evaluation.

- The purpose is to determine level of performance of student in the programme.
- Evaluation is categorized into two
  a) Formative
  b) Summative.

Summative has the following features:
- Placement
- It is also diagnastic.

Properties of a good test

- Must have the following:
  i) Reliable (ii) Valid (iii) Objective.
  iv) Discriminative.

i) Reliable → An exam must produce a pattern which is repeated.

ii) Valid → Exam must measure what it was made to measure.

iii) Discriminative - should sample out those students who have achieved objectives from those

Measurement

who have not.

Assignment.

Reasons for evaluating learners.

1. Measurement and diagnosis - It helps identify what a student knows and what they don't know.

2) Formative function (Guidance & Improvement) - The evaluation during the learning process. Judgement provides feedback to both the student & the teacher to guide ongoing instruction & learning.

3) Summative function (Judgement & certification) - It is used to determine the extent to which the learning objectives have been achieved.

4) Accountability function - effectiveness of the programme

5) Motivational function - good grades

6) Selection function - selecting individuals in their specific programme

(4) Accountability function. - It assesses the effectiveness of educational programs, policies or even entire institution. eg (5) from tests of different schools.

(5) Motivation function. - motivator to students if they achieve good grades. and encourage them to put more effort.

(6) Selection function.
- Sometimes evaluation is used to select individuals for specific programs, opportunities or positions, based on their abilities or qualifications

Functions.

a) Measurement & diagnosis - improvement
b) Formative function (Guidance & Improvement)
c) Summative function (Judgement) - objectives achieved.
d) Accountability function - effectiveness of the programme
e) Motivational function - good grades, putting in more effort.
f) Selective function - in various programs

*Evaluation - is a systematic process of determining the worth, significance, or condition of something, typically by careful appraisal and study.
- Student learning & teaching.

Statistics – science that uses numbers to collect data/facts, organizes the data, analyze it and draws conclusion.

STATISTICS.

- Is a science that uses numbers or numerals in collecting data/ideas/facts, organizes the data, analyze it and draws a conclusion.

Z-Score (STANDARD SCORE)

- It is a technique found in science, used in equalizing scores from different subjects, so as to find out which subjects performed better, i.e.

A madam did 3 exams. and results shown below:

	% (x)	Mean	std. deviation
History	80	75	5
Chemistry	50	45	2
R.E	90	80	5

To z score, $z = \frac{X - \bar{X}}{std}$

History $= \frac{80-75}{5} = 1$

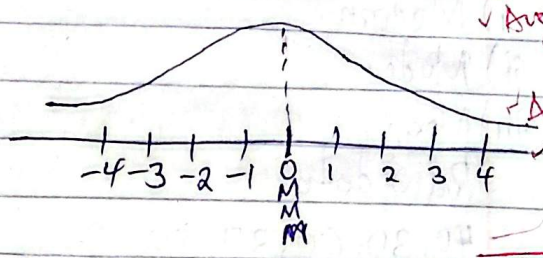
Chemistry $= \frac{50-45}{2} = 2.5$

R.E $= \frac{90-80}{5} = 2.$

Importance of z-score.

- Used in constructing normal curve and other curves.

Normal curve



- A normal curve is bell-shaped.
- It is symmetrical about the mean.
- Area under normal curve is 100% or 1.
- Is asymptotic,
- Mean, mode, median are all equal.

Mean, mode, median are equal!

Other curves

Positive skew (Many students failed)

Many students failed.

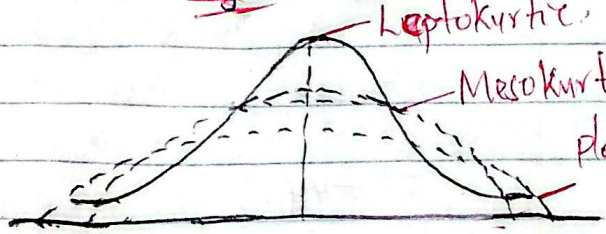
Negative skew (Many students scored highly)

scored highly

KURTOSIS

- A family of curves.

Kurtosis – is a family of curves.



$\frac{84}{5} \times 2$
$= 20$

- Techniques telling us how scores are concentrated at the centre. $\bar{x} =$

MEASURE OF CENTRAL TENDENCY

- There are techniques that tell us i.e how score are concentrated at the centre.
- These measures are three;
  i) Median.
  ii) Mode.
  iii) Mean.

Raw data.
40, 30, 60, 20, 70
↓
Organize the data.
20, 30, 40, 60, 70

GROUP SCORES.

mode is more frequent item

Class	f	$\frac{fx}{}$	fx	cf
80-89	2	84.5	169	2
70-79	9	74.5	670.5	11
60-69	24	64.5	1548	35
50-59	19	54.5	1035.5	54
40-49	9	44.5	400.5	63
30-39	9	34.5	310.5	72
20-29	17	24.5	416.5	89
10-19	8	14.5	116	97
0-9	3	4.5	13.5	100
	$\Sigma f = 100$		$\Sigma fx = 4680$	

Median = 40

Median of items in an odd distribution is;

$\frac{n+1}{2} = \frac{5+1}{2} = 3rd$ item

$= 40$.

What if the items in even distribution

20, 30, 40, 60, 70, 90

$\rightarrow \frac{6}{2} = 3^{rd}, 4^{th}$.

$3^{rd} = 40$

$4^{th} = 60$

$= \frac{40+60}{2} = 50$

Mean ($\bar{x}$) $= \frac{\Sigma fx}{\Sigma f} = \frac{4650}{100}$

$= 46.8$

Median $= L + \frac{(\frac{N}{2} - Cf_m)}{f_m} i$

Median $= L + \frac{(\frac{N}{2} - Cf_m)}{f} i$

Median $= 49.5 + \frac{(50-35)}{19} 10$

$= 57.39$

Mode $= 64.5$.

The distribution is skewed.

Position score. $= L + \frac{\text{effect } X}{f}$ i;

Mean.

Average mark.

$\bar{x} = \frac{20+30+40+60+70}{5}$

$= \frac{220}{5}$.

$= 44$.

Position score $= L + \frac{\text{effect } X}{f}$ i;

Median = $L + \left(\frac{N}{2} - cf_{am}\right) \times i$ / f_m

Mean deviation $\bar{x} = \frac{\Sigma fd}{\Sigma f}$ ⇒ $d = x - \bar{x}$

QUIZ 2.

~~Class~~	f	C	x	fx
35.5-40.5	2	36-40	38	76
30.5-35.5	3	31-35	33	99
25.5-30.5	4	26-30	28	112
20.5-25.5	6	21-25	23	138
15.5-20.5	5	16-20	18	90
10.5-15.5	3	11-15	13	39
5.5-10.5	1	6-10	8	8
	$\Sigma f = 24$			$\Sigma fx = 562$

$\bar{x} = \frac{\Sigma fx}{\Sigma f}$

$\bar{x} = \frac{562}{24}$

$\bar{x} = 23.42$

Modal class = 20.5 - 25.5

$f = 6$.

Median = $L + \left(\frac{N}{2} - cf_{am}\right) \times i$ / f_m

Median class = $\frac{N}{2} = \frac{24}{2} = 12$

= 21-25 class

$= 20.5 + \left(\frac{24}{2} - 9\right) \times 5$ / 6

$= 20.5 + \frac{3 \times 5}{6}$

$= 20.5 + \frac{15}{6}$

Median = 23

Measures of variability — techniques that tell us how scores are distributed.
- ✓ Range
- ✓ Mean deviation
- ✓ Quartiles
- ✓ variance
- ✓ Std deviation

MEASURES OF VARIABILITY

- These are techniques that tell us how scores are distributed.
- ✓ Range
- ✓ Mean deviation
- ✓ Quartiles — $\frac{Q_3 - Q_1}{2}$
- ✓ Variance. $\delta^2 = \frac{\Sigma fd^2}{\Sigma f}$
- ✓ standard deviation.

BLOOM'S TAXONOMY.

Levels of Thinking.

1) Knowledge (Remembering) Items

Terms / Questions / Keyword.

- ✓ List, Outline.

2) Comprehension Items:

- Terms / Keyword / Questions
- Explain, Why

3) Application Level.

Term / Keyword. ⇒ Compute mean & std

4) Analysis Level

Terms / Keyword / Questions.

- Breaking down into its component parts.

5) Synthesis level

- to bring together, combining all parts.

eg. What is the cost of the whole project.

- ✓ Create, design, develop.

6) Evaluation. - To give judgement based on reasoning.

Terms: Compare & Contrast, evaluate.

Table of specification.
Item | Level of thinking.

✓ each of the objectives emphasized.
✓ content validity

Table of Specification.

Item	Level of Thinking

Items are balanced or setting is not biased.

Importance of table of specification.

✓ 1) Ensures that the items are balanced, or the setting is not biased. (ensure)
✓ 2) Each objectives is emphasized
✓ 3) It ensures content validity

What is a standardized exam? -

- Types of Items.
- Types of questions.
- ✓ Objective Item/questions. (restrict from the given)
- ✓ Essay questions

1) Objective questions

✓ Restrict on from given question
✓ True or false.
✓ Cover large part of the syllabus learning.
✓ matching questions.

Disadvantages.

1) When not repeated it may not be reliable.

2) Essay types of questions

* Restricted essay - the candidate must provide a brief response of one or two paragraph.
✓ words used are explain, identify, comment on

* Extended essay - a candidate must provide a comprehensive response.
- Words used here include:
✓ Discuss ✓ Justify ✓ Evaluate ✓ Compare ✓ Analyze

Limitation.

1) Poor content sampling.
2) Subjective nature of essay.
3) The candidate does not know really what is needed.
4) Candidates don't think critically.
5) Candidates who don't know just write for the sake.

Advantages.

1) Encourages creativity.
2) Motivates a candidate to study alot
3) Easy to set.
4) It reveals the extent at which the learner has internalised. learning.

VALIDITY & RELIABILITY.

* Validity deals with finding out, and what was intended (the exam.) to measure.
- Truthfulness of the questions,
- The number of validity are as follows;
a) Face validity - is based on judgement of the learner.

b) Content validity - test covers the entire area.
c) Construct validity. - Whether rules were followed when constructing questions.
d) Criterion related validity. ← Concurrent, Predictive.

INTERNAL & EXTERNAL EXAMS

INTERNAL

- Administered by the teacher to his own students.
- They are set from what has been learnt and therefore lacks content validity.
- Marking is subjective due to influence of some students.
- Does not give a true picture of your students.

Advantages

1. Questions are familiar to the learner and the learner can easily predict what to be set.
2. Questions based on what has been covered.
3. Candidates experience less stress because they can easily remember some question attributes.

Disadvantages

1. Teacher is likely to be subjective in marking.
2. Content validity is poor.
3. Maybe below the standard of normative group.

EXTERNAL EXAMS

- Usually managed by examining bodies to a large population of candidates. Therefore, they are standardized exams.

Advantages

1) Examiners are likely to be objective ✱ Gives the exam a greater reliability than the internal exams
2) Candidates are likely to go through the whole syllabus as they cannot easily predict questions.
3) Possess a universal standard
4) Have a better validity.

Disadvantages

1) Teachers and learners are likely to work under pressure as they compete with others.
2) Learners may cram facts instead of understanding facts.

MEASURES OF VARIABILITY

Also called Measures of dispersion.

Example

4, 6, 8, 9, 3

Range

3, 4, 6, 8, 9

= 9 − 3 = 6. (Highest − Lowest)

Mean deviation.

Mean, $\bar{x} = \frac{3+4+6+8+9}{5} = \frac{30}{5} = 6$.

Classes	f	Midpoint x	fx	d $(x-\bar{x})$	fd	fd^2
35-39	2	37	74	14.64	29.28	428.36
30-34	3	32	96	9.64	28.92	278.79
25-29	6	27	162	4.64	27.84	129.18
20-24	8	22	176	−0.36	−2.88	1.04
15-19	4	17	68	−5.36	−21.44	114.93
10-14	3	12	36	−10.36	−31.08	321.99
5-9	2	7	14	−15.36	−30.72	471.86
	$\Sigma f = 28$		626			1746.44

$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{626}{28} = 22.36$

x	f	fx	d $(x-\bar{x})$	d^2	fd^2
9	1	9	3	9	9
8	1	8	2	4	4
6	1	6	0	0	0
4	1	4	−2	4	4
3	1	3	−3	9	9
		30	10		26

Mean deviation $= \frac{\Sigma f|x-\bar{x}|}{\Sigma f}$

$= \frac{10}{5} = 2$

Variance $= \frac{\Sigma fd^2}{\Sigma f} = \frac{26}{5}$

$= 5.2$

Standard deviation.

$= \sqrt{\text{variance}}$

$= \sqrt{5.2}$

$= 2.280$

Variance $(s^2) = \frac{\Sigma fd^2}{\Sigma f} = \frac{1746.44}{28} = 62.37$

Standard deviation $= \sqrt{\text{variance}}$

$= \sqrt{62.37}$

$= 7.897$

Quartiles

Class	f	cf
80-89	2	100
70-79	9	98
60-69	24	89
50-59	19	65
40-49	9	46
30-39	9	37
20-29	17	28
10-19	8	11
0-9	3	3

1st quartile score (Q^1)

2nd " " (Q^2)

3rd " " (Q^3)

Position score $= L + \frac{\frac{n}{4} - cf(x)}{f}$

Position score $= L + \frac{\frac{3n}{4} - cf(x)}{f}$

Position $\frac{1}{4} \times 100 = 25^{th}$ position score.
$\frac{2}{4} \times 100 = 50^{th}$ score.
$\frac{3}{4} \times 100 = 75^{th}$ score.

Position score $= 19.5 + \frac{14 \times 10}{17}$

$= 19.5 + 8.235$

$= 27.74$

Deciles $= \frac{n}{10}$

Percentiles $= \frac{n}{100}$

Eighth decile score $= \frac{8}{10} \times 100 = 80^{th}$ score.

ITEM ANALYSIS

- It is applying statistical techniques to the assessment of an individual test.
- It is also the process of examining response of test.
- Applied both in the classroom, final exam and external exam.
- The following factors are considered.

i) The difficulty of the item. (How many student got it right)

ii) Discrimination index (does it distinguish fast learner & slow learner.

iii) Does it bring out clearly performance btwn a fast and slow learner.

Why Item analysis.

a) It provides quantitative evidence to support difficulty and to support discrimination test used.

b) Enable teachers to identify the problem of each item.

Steps of conducting item analysis.

i) Arrange the scored marks from highest to lowest.

ii) Form groups in this manner.

Top - 27% $\Rightarrow \frac{27}{100} \times 150 = 13.5$

Bottom - 27% $\simeq 14$.

iii) For each item count the no. of candidates who choose each alternative

iv) In the case of true or false.

Question 9.

	Alternatives						
	A	B	C	D	E	O	E
UP	0	0	20	0	0	0	20
LP	4	2	8	3	3	0	20

Compute the difficult index for item

soln

Difficult index (D) $= \frac{UP + LP}{UP + LP}$ ← Those who got ← Total exami

$= \frac{20 + 8}{20 + 20} = \frac{28}{40} = 0.7$

Difficulty index, $D = \frac{UP + LP}{UP + LP \leftarrow \text{Total examinees, all}}$ — those who got

Discrimination index.

$DA = \frac{UP - LP}{\text{Total in one group}} = \frac{20-8}{20} = 0.6$

$= 20-8 / 20 = 60\% = 0.6 = 60\%$

Item discrimination index:

- Is a method that gives us the value at which the values are related.

$DA = \frac{UP - LP}{\text{Total in one group}} = \frac{20-8}{20} = 0.6$

- For true or false items.

$ID = \frac{R \times 100\%}{\text{Total}}$

R - Candidates who got right

Total = 200

Right response = 98.

$D = \frac{98 \times 100\%}{200}$

Question 5.

Alternatives	A	B	C	D	E	O	E.
UP	5	84	0	11	0.0	100	
LP	14	70	4	12	0	0	100

- Response E is not a good one it should be omitted.
- Such question is not sampling out properly a fast learner from a slow learner.
- Such question should be omitted.
- A response of 0.4 and above is high in discrimination.
- 0.3 - 0.39 is fair but not good. 0.2 - 0.29 is a minimal value and can be corrected and below 0.2 should be rejected.

CORRELATION

- Is a method that gives us the value at which the values are related.
- ✓ Regression
- ✓ Spearman Rank order correlation.
- ✓ Pearson correlation moment.

Spearman Rank Correlation

	X	y	d	d^2
1	50	45	1	1
2	49	50	1	1
3	30	25	0	0
4	11	10	1	1
5	10	15	1	1
				4

$n = 5$.

$= 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$

$= 1 - \frac{6 \times 4}{5(5^2-1)} = 1 - \frac{24}{5(25-1)} = 1 - \frac{24}{5 \times 24}$

$= 1 - \frac{24}{120} = 0.8$.

Pearson correlation regression coefficient.

$r_{xy} = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N \Sigma x^2 - (\Sigma x)^2][N(\Sigma y^2 - (\Sigma y)^2]}}$

x	y	xy	x^2	y^2
50	45	2250	2500	2025
49	50	2450	2401	2500
30	25	750	900	625
11	10	110	121	100
10	15	150	100	225
150	145	5710	6022	5478

Spearman Rank Correlation

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

Pearson Correlation Coefficient.

$$r_{xy} = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N(\Sigma x^2) - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

$N = 5$

$$r_{xy} = \frac{5(5710 - 21750)}{\sqrt{[5\times6022 - 22500) - (5\times5475 - 25025)}}$$

$$= \frac{6800}{\sqrt{7610\times6350 - 6951}} = \frac{6800}{}$$

$= 0.978. \Rightarrow 97.8\%$

$\Rightarrow$ A very strong relationship.

Revision

Q1 Njoki did 3 exams and the results are shown below.

	%	Mean	Standard deviation
History	80	75	5
Biology	50	45	2
Physics	90	80	5

a) In which subject did she perform better.

Soln

History, $Z_H = \frac{x - \bar{x}}{d} = \frac{80-75}{5} = 1$

Biology, $Z_B = \frac{x - \bar{x}}{d} = \frac{50-45}{2} = 2.5.$

Physics, $Z_P = \frac{x - \bar{x}}{d} = \frac{90-80}{5} = 2$

Biology did better with a Z-score of 2.5.

b) Importance of Zscore.

→ Used in constructing normal curves and other curves.

2) Given the data below.

12, 16, 8, 6, 10, 14, 8, 6.

Soln

Calculate the mean.

Soln

6, 8, 10, 12, 14, 16 ($|x - \bar{x}|$)

x	f	fx	d	d^2	fd^2
6	2	12	4	16	32
8	2	16	2	4	8
10	1	10	0	0	0
12	1	12	2	4	4
14	1	14	4	16	16
16	1	16	6	36	36

$\Sigma f = 8$ $\Sigma fx = 80$ $\Sigma fd^2 = 96$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{80}{8} = 10$$

ii) Variance

$d = |x - \bar{x}|$

$$\frac{\Sigma fd^2}{\Sigma f} = \frac{96}{8} = 12$$

iii) Standard deviation.

$\sqrt{12} = 3.46.$

iv) Range $= H - L = 16 - 6. = 10$

v) Median $= \frac{n+1}{2} = \frac{8+1}{2} = 4.5$

$= \frac{4\text{th} + 5\text{th}}{2} =$

6, 6, 8, 8, 10, 12, 14, 16.

$\frac{8+10}{2} = 9$

Q3.

Class	f	x	fx	$\lvert x-\bar{x}\rvert^d$	d^2	fd^2	c.f
60-62	8	61	488	6.53	42.6409	341.1272	8
63-65	10	64	640	3.53	12.4609	124.1609	18
⇒66-68	46	67	3082	0.53	0.2809	12.9214	64
69-71	30	70	2100	2.47	6.1009	183.027	94
72-74	7	73	511	5.47	29.9209	209.4463	101
	$\Sigma f=101$		$\Sigma fx=6821$			$\Sigma fd^2=871.1309$	

Calculate; Mean, mode, median, variance, standard deviation.

Soln

i) Mean $= \frac{\Sigma fx}{\Sigma f} = \frac{6821}{101} = 67.53$

ii) Mode - 66-68.

iii) Median $= L_m + \frac{\left(\frac{N}{2} - C.f_{a.m}\right)}{f_m} \times i$

$\frac{1}{2} \times 101 = 50.5$

$= 65.5 + \frac{\left(\frac{101}{2} - 18\right)}{46} \times 3$

$= 65.5 + \left(\frac{50.5 - 18}{46}\right) \times 3.$

$= 65.5 + \frac{32.5}{46} \times 3.$

$= 65.5 + 2.1196.$

$= 67.62.$

iv) Variance $= \frac{\Sigma fd^2}{\Sigma f} = \frac{871.1309}{101}$

$= 8.625$

v) Standard deviation $\sqrt{\text{variance}}$

$= \sqrt{8.625}.$

$= 2.93$

Quartile

$Q_1 = \frac{N}{4} = \frac{101}{4} = 25.25.$

$Q_2 = \frac{N}{2} = \frac{101}{2} = 50.5$

$Q_3 = \frac{3N}{4} = \frac{101 \times 3}{4} = 75.75.$

Decile $= \frac{n}{10}$

Percentile $= \frac{n}{100}$

Interquartile range $= Q_3 - Q_1$

Quartile deviation (Semi-interquartile range) $= \frac{Q_3 - Q_1}{2}$

Coefficient of quartile deviation.

$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Responses	A	B	C*	D	E
Upper group	6	5	9	0	20
Lower group	8	4	10	0	22
					42.

i) Item difficulty index, $ID = \frac{UP + LP}{\Sigma[UP + LP]}$

$= \frac{9+10}{42.} = \frac{19}{42} = 0.45$

ii) Item discrimination index.

$= \frac{UP - LP}{\Sigma UP}$

$= \frac{9-10}{20} = \frac{-1}{20} = -0.05$

iii) Comment on the distractors.

⇒ The question should be rejected since it was not able to point out the fast & slow learners

✓Response D was a bad one because no one scored.
✓0.45 - high in discrimination
✓- 0.105 - Question should be rejected.

X-trs of a good test.

✓Reliable pattern.
✓Valid.
✓Objectives.
✓Discrimination.

Qns

What is measurement.

- Is the quantification of an outcome.
- Is the process of assigning numerical values to a student's knowledge, skills, abilities or attitudes.

Key concepts

(1) Reliability - consistency of the measurement.
2) Validity - Accuracy of the measurement.
(3) Types of Measurement scales.
✓Nominal, ordinal, interval & Ratio scales.
4) Assessment - a broader term that includes measurement but also involves interpreting the data to make judgements.

Levels of Measurement.

① Nominal level. basic level.
- Numbers are used simply as labels or categories to identify or classify things, e.g. classifying students by major. like 1 = Psychology 2 = Sociology etc. - No interval order or ranking.
Operation; You can only count frequencies. ② and find the mode.

② Ordinal level.
- The numbers at this level indicate a rank order or sequence, eg greater than etc. eg. Ranking students in performance. like (1st place, 2nd - 3rd) or in survey responses like Agree, Neutral, Disagree.
- Operation; (You can count frequencies, find mode and also determine the median (the middle level)

③ Interval level.
- Has both order & equal intervals btw values. e.g: Temp in °C or F. (0°C doesn't mean no temp).
- Difference btw any two consecutive values is the same.
- Operations; You can perform all operations from the ordinal level plus you can add & subtract values.

(iv) Ratio level
- Is the highest & most informative level of measurement.
- It has all the properties of interval level plus true, meaningful zero point. e.g. Height, weight, no. of correct answers on a test, etc.
- Operations; You can perform all mathematical operations & addition, subtraction, multiplication & division.

How to ascertain Validity & Reliability.

Validity - degree to which a test measures what it claims or intends.

How to ascertain
(i) Content validity.
(ii) Criterion validity - relate to an external criterion
- Concurrent validity - compare test scores with another well established test administered at the same time.
- Predictive validity - determine how well test scores predict future performance.
(iii) Construct validity.
(iv) Face validity

Reliability
- Consistency, stability & dependability of test scores.

How to ascertain
(a) Test - Retest Reliability.
- Measures stability over time.
- High correlation indicates high reliability.

b) Split-Half Reliability - examines internal consistency.
- Divide test into two equivalent halves

c) Parallel-Forms Reliability.
- Assesses consistency across different versions of same test.

d) Inter-rater reliability
- Used where judgement or scoring is subjective.
- Has two or more raters score the same responses.

Relationship btwn Validity & Reliability
- A test can be reliable but not valid.
- A test cannot be valid unless it's reliable.

Essay Test is largely subjective. Explain 5 ways through which you can enhance objectivity while scoring essay tests.

(1) Use a detailed scoring rubric (Marking scheme) - before scoring begins. - Minimizes personal bias & ensures uniform scoring.

(2) Anonymity of scripts - Score scripts without knowing the identity of the candidates.

(3) Training & Standardization of examiners.

(4) Use of multiple markers. - two or more examiners

(5) Scoring One question at a time. - Mark the same question for all candidates at a time.

Item analysis

- Is the process of examining response of a test.
- Is the systematic statistical process used to evaluate the quality of an individual test of items.

Procedure for Item Analysis in a Norm-referenced Test

Step 1: Administer the test.
2: Score the test. - total scores.
3: Rank the candidates.
4) Form Upper & Lower Groups.
5) Calculate the Item difficulty Index (p-value)

$p = \frac{\text{No of students answering correctly}}{\text{Total no. of candidates}}$

* p - proportion of candidates
* High - p → item is easy.
* Low p → item is difficult.
- Acceptable range: 0.30 - 0.70.

6) Calculate Item Discrimination Index (D)

$D = \frac{U - L}{N}$.

7) Interpret Results.
8: Improve the test - revise, relace or eliminate weak items.

Skewness - tells more about the asymmetry of distribution. eg.
Zero skewness (normal) distribution.
Positive skewness (Right skewed) - long tail to the right.
-ve skewness. (left skewed) long tail to the left.

Kurtosis. - tells you about peaked ness or flatness of a distribution & weight of the tails.
- Mesokurtic: normal
- Leptokurtic - very peaked with heavy tails.
- Platykurtic - very leptokurtic - very peaked with heavy tails.
- Platykurtic - flat peak with light tails.

Z = 0.15 - Given area.
Area to left.
Area to right = Z
Bivariate-
ors.

Cumulative frequency table.

Class interval	frequency	c.f.

Histogram
x-axis - plot class intervals
y-axis - ↑ frequencies
Draw. bars (no gaps)

Frequency polygon
✓ Find midpoints of each class interval.
✓ Plot the midpoints against their frequencies.
✓ Join the points with straight lines.
✓ Extend the polygon to touch the x-axis at both ends.

Comments
✓ Check on scores.

Conclusion.
Skewness & kurtosis.

Educational objectives
Blooms of taxonomy.

① Cognitive domain (Knowledge & Intellectual skills)
- Mental abilities and intellectual development.
- Focuses on what learners know, understand & can think through.

Levels.
(1) Knowledge - recall
(2) Comprehension - understanding
(3) Application - using knowledge in new situations
(4) Analysis - breaking information into parts.
(5) Synthesis - combining ideas to form something new. eg designing lesson plan
(6) Evaluation - make judgements based on criteria.

② Affective Domain (Values/Attitudes)
- Concerns emotions, feeling, attitudes, & interests & values
- Focuses on how learners feel about learning and how they internalize values.

Levels.
(a) Receiving - attention
b) Responding - active participation.
c) Valuing - showing commitment.
d) Organization.
e) Characterization. - values consistently guide behavior.

③ Psychomotor Domain. (Skills/ Physical Abilities.)
- Physical movement, coordination & motor skills.
- Emphasizes learning by doing.

Levels
a) Imitation - copying an action.
b) Manipulation - perform without
c) Precision - accuracy & control
d) Articulation - coordinating several skills smoothly.
e) Naturalization - automate, effortless performance

* Standardized exam - test that is designed, administered, scored, & interpreted in the same way for everyone who takes it.

ie. Everyone gets the same test under same conditions.

Features.

- Uniform content.
- ✓ Standard instructions.
- ✓ " timing.
- ✓ Objective scoring.
- ✓ Norms or benchmarks. comparison.

Example

- ✓ National exams.

Reasons for use.

a) To compare students fairly.
b) Measure achievement, aptitude or ability.
c) To help with selection, placement or certification.

Standard error of Measurement.

- Is an estimate of the amount of error inherent in an individual's score on a test.

Functions of measurement.

1. Diagnosing learning needs.
2. Monitoring student progress
3. Assigning grades.
4. Evaluating teaching effectiveness.
5. Providing feedback to students.

Purpose of Test & Measurement in an Educational setting.

a) To assess student learning.
b) Inform instructional decisions.
c) Evaluate program effectiveness.
d) Provide accountability &
e) facilitate student placement.

Measurement techniques

1) Standardized tests.
2) Teacher made tests.
3) Performance assessments,
4) Portfolios. - compile student work over time

Qualities of an educational programme.

(1) Relevance - needs of society, learners & job market.
2) Effectiveness - achieving stated objectives
3) Efficiency - utilizing resources (time, money, personnel). wisely to achieve its goals.
4) Equity - fair and accessible to all students.

Steps in Test Specification.

(1) Define the Purpose & Objectives of the Test.
(2) Identify the Target Population. age group
(3) Determine content domain and scope.
(4) Specify the test structure & format. multiple choice
5) Outline the Weighting & Scoring Scheme.
6) Establish standards for Item Analysis

7) Consider technical qualities (Reliability & Validity).
8) Develop a Table of Specifications (Blue-print) - the intended balance of the sheet.