

APPSC PAPER I: GENERAL STUDIES & MENTAL ABILITY · EDUCATION (B.ED. LEVEL)

Statistics for teachers

Free study notes · 2 notes · 25 Oct 2026

Free study material from Medha AI. For notes, practice, mock tests and previous papers for your exam, start free at **app.medhaai.in**.

1. Statistics in the classroom: averages, spread and correlation of scores
2. Testing and evaluation: kinds of tests, qualities of a good test, types of evaluation and norms

1 Statistics in the classroom: averages, spread and correlation of scores

A teacher first arranges a class's scores (array, frequency table, graph), sums them up with an average (mean, median or mode), shows how spread out they are (range, quartile deviation, mean deviation, variance, standard deviation), places a pupil in the group with percentiles and standard scores, and uses correlation (-1 to +1) to see whether two sets of scores move together.

Core facts

- **Measures of central tendency** (averages): **mean, median, mode**. C.V. Good: the tendency of scores to **cluster about a point** on the scale.
- **Mean** = sum of scores $\div$ number of scores ($M = \Sigma X \div N$); the **mathematical (arithmetic) average**; uses every score; pulled by extreme scores.
- **Median** = the **middle** score of scores arranged in order: **half the scores lie above it and half below**; the **positional average**; not affected by extreme scores; = **50th percentile** = Q_2 .
- **Mode** = the score that occurs **most often**; the only average for **nominal** data (categories such as blood group, religion, favourite subject).
- **Empirical relation** (Karl Pearson; moderately skewed data): **Mode = 3 Median - 2 Mean**.
- **Combined mean** of two groups = $(n_1\bar{x}_1 + n_2\bar{x}_2) \div (n_1 + n_2)$.
- **Measures of variability** (dispersion, spread, scatter): **range, quartile deviation, mean deviation, standard deviation**.
- **Range** = **highest score - lowest score** (quickest, roughest; depends on two scores only).
- **Quartile deviation (Q)** = $(Q_3 - Q_1) \div 2$, the **semi-interquartile range**; goes with the median.
- **Mean deviation** = average of the deviations from the mean, **signs ignored**.
- **Variance** = the **average of the squared differences** of the scores from the mean (σ^2).
- **Standard deviation (SD, σ)** = **square root of the variance**; the **typical (standard) distance** of scores from the mean; the most stable measure of spread; goes with the mean. Term coined by **Karl Pearson** (1893 lecture, in print 1894).
- **Percentile (P_k)**: the score below which k% of the group falls; **percentile rank (PR)**: the percentage of the group scoring below a given score.
- **z-score** = $(X - \text{Mean}) \div \text{SD}$ (mean 0, SD 1); **T-score** = $50 + 10z$ (mean 50, SD 10).
- **Normal curve**: mean = median = mode; **68.26%** of cases within ± 1 SD, **95.44%** within ± 2 SD, **99.73%** within ± 3 SD.
- **Ogive** = the **cumulative frequency** (cumulative percentage) graph; median and percentiles are read from it.
- **Correlation** = the statistical tool for the **relationship** between two variables; coefficient from **-1 to +1**.
- **Positive** correlation: both rise together and fall together; **negative**: as one **increases** the other **decreases**; **zero**: no relationship.

- **Karl Pearson:** product-moment correlation r ; **Charles Spearman:** rank correlation ρ (ρ) = $1 - \frac{6\sum D^2}{N(N^2 - 1)}$. **Francis Galton** first gave the idea of correlation ("co-relation") and regression.

Explained

1. Organising scores before averaging

- **Array:** scores arranged in ascending or descending order.
- **Grouped frequency distribution:** scores put into **class intervals** (usually 3, 5 or 10 units wide; about 5 to 15 classes) with tallies and frequencies; total of frequencies = N .
- **Exact limits** of the class 36–39 are **35.5–39.5**; its midpoint is 37.5.

Graph	How it is drawn	Used for
Histogram	adjoining bars: base = class interval, height = frequency	shape of the distribution, mode
Frequency polygon	frequencies plotted at class midpoints, joined by lines	comparing two distributions
Ogive	cumulative % plotted at upper exact limits	median, quartiles, percentiles
Pie chart	sector angle = $(\text{part} \div \text{total}) \times 360^\circ$	shares of a whole

Example: 45 of 180 hostel pupils come from tribal areas, so their sector is $45 \div 180 \times 360^\circ = 90^\circ$.

2. Which average when

Average	Type	Best for	Weakness
Mean	mathematical	interval data, further work (SD, r); most stable from sample to sample	pulled by extreme scores
Median	positional	skewed scores, ranks and letter grades, open-ended classes ("60 and above")	ignores the size of other scores
Mode	most frequent	nominal data, "most popular" (one uniform size for most pupils)	may be absent, or two (bimodal) or more

- Two uses of an average: to **describe a group as a whole** and to **compare groups** on the same test.
- Properties of the mean: the **sum of deviations from the mean is always zero**; the **sum of squared deviations is least** when taken from the mean.
- The mode is the **quickest and roughest** average; it can be read from the peak of a histogram.
- The mean is the **centre of gravity (balance point)** of the scores: the deviations on either side balance out.
- **Correcting a wrong entry:** 30 pupils, mean 45, but one score of 72 was entered as 42. Correct total = $1350 - 42 + 72 = 1380$, so the correct mean = $1380 \div 30 = 46$.

3. Worked example: one class test (10 pupils, out of 25)

Scores: 12, 15, 9, 18, 15, 11, 20, 15, 13, 12

- **Mean** = $140 \div 10 = 14$.
- **Median**: in order 9, 11, 12, 12, 13, 15, 15, 15, 18, 20. With 10 (even) scores the median is the average of the 5th and 6th: $(13 + 15) \div 2 = 14$.
- **Mode** = **15** (three times).
- **Range** = $20 - 9 = 11$.
- **Percentile rank of 18**: 8 of the 10 pupils scored below 18, so $PR = 8 \div 10 \times 100 = 80$ (some books also add half of the pupils with the same score).

4. Median of raw scores (odd number). 17, 19, 22, 24, 26, 30, 32, 38, 40: nine scores, already in order, so the median is the **5th** score = **26**. Rule: median = the $((N + 1) \div 2)$ th score after arranging.

5. Formulas for grouped data

- Mean, long method: $M = \Sigma fx \div N$ (x = class midpoint). Short (assumed-mean) method: $M = A + (\Sigma fx' \div N) \times i$.
- Median: $Md = l + [(N/2 - F) \div f] \times i$, where l = exact lower limit of the median class, F = cumulative frequency **below** that class, f = its frequency, i = class width. Example: $l = 39.5$, $N = 50$, $F = 18$, $f = 12$, $i = 5$ gives $39.5 + (7 \div 12) \times 5 \approx 42.4$.
- Q_1 is found the same way at $N/4$, Q_3 at $3N/4$. $Q_1 = 25$ th percentile, $Q_3 = 75$ th percentile; $Q_3 - Q_1$ (interquartile range) covers the **middle 50%** of scores.
- Empirical mode: mean 40, median 38 $\rightarrow$ mode = $3 \times 38 - 2 \times 40 = 34$.

6. Combined mean. Section A: 15 pupils, mean 8; section B: 10 pupils, mean 6. Combined mean = $(15 \times 8 + 10 \times 6) \div (15 + 10) = 180 \div 25 = 7.2$. (Not the simple average 7, because the groups are of different sizes.)

7. Spread in a small set. Scores 4, 6, 8, 10, 12: mean = 8; deviations -4, -2, 0, 2, 4.

- **Mean deviation** = $(4 + 2 + 0 + 2 + 4) \div 5 = 2.4$ (signs are ignored because the signed deviations always add to zero).
- Squares 16, 4, 0, 4, 16; sum of squares = 40. **Variance** = $40 \div 5 = 8$; **SD** = $\sqrt{8} \approx 2.83$. If SD = 6, variance = 36.
- Σd^2 , the total of the squared deviations from the mean, is called the **sum of squares**; mean deviation is also called **average deviation (AD)**.
- Variance is in **squared units** (marks²); SD is in the original units, which is why SD is reported. In testing, total variance = **true variance + error variance**.

Measure	Formula	Goes with
Range	$H - L$	quick look
Quartile deviation	$(Q_3 - Q_1) \div 2$	median, open-ended classes
Mean deviation	Σd (signs ignored) $\div N$	rarely used now
Standard deviation	$\sqrt{(\Sigma d^2 \div N)}$, deviations from the mean	mean, normal data

- In a normal distribution: **QD** < **MD** < **SD** (about 0.67σ , 0.80σ and σ).
- SD = **0** only when **all scores are equal**.

8. Reading the SD. Two sections have the same mean of 60. Section A has SD 5 (pupils are alike: **homogeneous**, teach as one group); section B has SD 15 (a wide spread: **heterogeneous**, plan remedial

and enrichment groups). This is why an average alone cannot compare two classes.

9. Effect of changing every score. Adding the same number (5 grace marks) to every score raises the mean, median and mode by 5 but leaves the range and SD **unchanged**. Multiplying every score by k multiplies the mean and SD by k and the variance by k^2 (mean 12, every score $\times 2$ then $+ 3 \rightarrow$ new mean 27).

10. Skewed scores. A very easy test piles scores at the top (**negatively skewed**: mean $<$ median $<$ mode); a very hard test piles them at the bottom (**positively skewed**: mean $>$ median $>$ mode). In skewed data the median is the safer average.

11. Comparing consistency: coefficient of variation (Karl Pearson) = $(SD \div \text{mean}) \times 100$, a **relative** (unit-free) measure. Class A: mean 60, SD 12 $\rightarrow$ CV 20%; class B: mean 40, SD 10 $\rightarrow$ CV 25%. Class A is **more consistent** (lower CV). Range, Q, MD and SD are **absolute** measures, in score units.

12. Standard scores

Scale	Mean	SD
z-score	0	1
T-score	50	10
Stanine (nine bands)	5	about 2
Deviation IQ (Wechsler)	100	15

Mean 50, SD 10, pupil scores 65 $\rightarrow z = (65 - 50) \div 10 = 1.5$, $T = 65$. Standard scores have equal units; percentile units are unequal along the scale.

13. Correlation in school

Pair of variables	Likely correlation
intelligence and school achievement; height and weight	positive
hours of TV and marks; typing practice and mistakes	negative
roll number and marks; height and marks in English	about zero

- **Kinds of correlation: positive, negative, zero; linear and curvilinear; simple, partial and multiple.** ("Cumulative correlation" is not a kind.)
- Correlation needs **two sets of scores for the same pupils**; r gives the **degree and direction** of the relationship in one number.
- The **sign** shows direction; the **size** shows strength: -0.82 is stronger than $+0.60$; $+0.9$ and -0.9 are equally strong.
- **Scatter diagram**: paired scores plotted as dots. A narrow band rising to the right = high positive; falling to the right = high negative; all on a rising straight line = **+1**; a shapeless cloud = zero.
- **Pearson's r** suits interval or ratio scores with a **linear** relationship; "product-moment" = the products of paired deviations from the two means. A strong **curved** (inverted-U) relationship can still give r near 0.
- **Spearman's p** uses **ranks**; quicker for small groups (N below about 25) or data available only as ranks. D = the difference between the two ranks of each pupil. **Tied** scores get the **average** of the ranks they share (three pupils sharing 2nd, 3rd and 4th each get 3).

- **Spearman example:** ranks (1, 2, 3, 4, 5) and (2, 1, 4, 3, 5); $D: -1, 1, -1, 1, 0$; $\Sigma D^2 = 4$. $\rho = 1 - (6 \times 4) \div (5 \times 24) = 0.8$. Identical ranks $\rightarrow \rho = +1$; exactly reversed ranks $\rightarrow \rho = -1$. $N = 5$, $\Sigma D^2 = 10 \rightarrow \rho = 0.5$; $N = 6$, $\Sigma D^2 = 14 \rightarrow \rho = 0.6$.
- **r^2 (coefficient of determination):** the share of variance the two have in common: $r = 0.7 \rightarrow 49\%$.
- Usual reading of r (Garrett): 0 to ± 0.20 negligible; ± 0.20 to ± 0.40 low; ± 0.40 to ± 0.70 moderate (substantial); ± 0.70 to ± 0.90 high; ± 0.90 to ± 1.00 very high.
- **Correlation is not causation:** correlational studies ask whether two variables are **associated**; a significant r , positive or negative, only says the two **vary together**.
- Tests use correlation: **test-retest reliability** and **validity coefficients** are correlation coefficients (see EDU-ST-02).

14. Scales of measurement (S.S. Stevens, 1946; they decide which statistic to use)

Scale	Example	Average
Nominal	gender, blood group, roll number	mode
Ordinal	ranks, grades A, B, C	median
Interval (no true zero)	test scores, temperature in °C	mean
Ratio (true zero)	height, weight, time	mean (and all others)

Common traps

✗ "Median is the mathematical average." ✓ **Mean** is mathematical; **median** is positional.

✗ "Mean is right for nominal data." ✓ **Mode** is the only average for nominal data.

✗ "The score above and below which half the scores fall is the mean." ✓ It is the **median**.

✗ "SD is the square of the variance." ✓ SD is the **square root** of the variance; variance is the square of SD.

✗ "Variance is the average difference from the mean." ✓ It is the average of the **squared** differences.

✗ "Highest – lowest score is the quartile deviation." ✓ That is the **range**; $Q = (Q_3 - Q_1) \div 2$.

✗ "Norms, reliability or validity measure the spread of scores." ✓ Spread is measured by the **SD** (range, Q, MD); the others are qualities of a test.

✗ "A cumulative frequency graph is a histogram." ✓ It is an **ogive**.

✗ "Combined mean of 15 items (mean 8) and 10 items (mean 6) is 7." ✓ Weight by size: $180 \div 25 = 7.2$.

✗ "Negative correlation means no relationship" or "-0.6 is weaker than +0.6". ✓ Negative = one rises as the other **falls**; -0.6 and +0.6 are equally strong; no relationship is **zero** correlation.

✗ "A significant correlation proves X causes Y." ✓ It only shows X and Y **vary together**.

✗ "Adding 5 grace marks to all raises the SD." ✓ The SD stays the **same**; only the averages rise by 5.

✗ BRAOU's BRS text calls ± 0.40 to ± 0.70 "high". ✓ The standard (Garrett) reading is **moderate (substantial)**; high starts at ± 0.70 .

Quick revision

- Averages: mean (mathematical), median (positional, half above and half below), mode (most frequent).
- Nominal → mode; ranks, grades, skewed or open-ended → median; interval → mean.
- Median = $((N + 1) \div 2)$ th score after arranging; combined mean 15×8 and $10 \times 6 \rightarrow 7.2$.
- Mode = $3 \text{ Median} - 2 \text{ Mean}$; ogive = cumulative frequency graph.
- Range = $H - L$; $Q = (Q_3 - Q_1) \div 2$; MD ignores signs.
- Variance = mean of squared deviations; $SD = \sqrt{\text{variance}}$; SD 0 = all scores equal.
- Grace marks: averages rise, SD unchanged; $CV = SD \div \text{mean} \times 100$.
- Normal curve 68.26 / 95.44 / 99.73% within 1 / 2 / 3 SD; z mean 0 SD 1; T mean 50 SD 10.
- Correlation -1 to +1: positive together, negative opposite, zero none; not causation.
- Pearson r (product-moment), Spearman $\rho = 1 - 6\sum D^2 \div [N(N^2 - 1)]$, Galton first.

2 Testing and evaluation: kinds of tests, qualities of a good test, types of evaluation and norms

Measurement gives a number, evaluation adds a value judgement; a good test is valid, reliable, objective and usable and has norms; evaluation is placement (before teaching), formative and diagnostic (during), summative (at the end), while prognostic tests predict future success.

Core facts

- **Measurement** = describing a trait in **numbers** (quantitative only). **Evaluation** = measurement or description **plus a value judgement** (quantitative + qualitative); it is wider and continuous.
- **Test** = a tool (set of questions or tasks) that gives a sample of behaviour; **examination** = a formal, usually end-of-course, test.
- **Achievement test**: measures what a pupil **has learnt** (present attainment in a subject).
- **Aptitude test**: measures capacity to learn a skill with training; it **predicts** future success, so it is **prognostic**. An entrance test (for example B.Ed. entrance) is best an **aptitude test**.
- **Prognostic test** = a test that **forecasts** a pupil's future performance.
- **Diagnostic test**: finds a pupil's **specific weaknesses and their causes**; it is followed by **remedial teaching**.
- **Qualities of a good test**: **validity, reliability, objectivity, usability (practicability)** and **norms**; also discriminating power and comprehensiveness. **Subjectivity is not** a quality of a good test.
- **Validity** = the test measures **what it claims to measure** (the most important quality).
- **Reliability** = **consistency**: the same pupils get about the same scores when tested again.
- **Objectivity** = scoring is free from the examiner's personal judgement: **different examiners give the same score**.
- **Usability** = easy and economical to give, score and interpret (time, cost, length).
- **Norms** = the **average performance of a representative (standardisation) group**, used to interpret a raw score: **age, grade, percentile** and **standard-score** norms.
- **Percentile rank 70** = the pupil scored higher than **70% of the norm group** (not 70% marks).
- **Placement evaluation**: **before** teaching, to check entry knowledge and place pupils in the right group.
- **Formative evaluation**: **during** teaching, to give feedback and improve learning; terms formative and summative coined by **Michael Scriven (1967)**.
- **Diagnostic evaluation**: during teaching, for pupils who keep failing despite formative help; finds the **cause** of the difficulty.
- **Summative evaluation**: at the **end** of a unit, course or **programme of study**; for grading, promotion and certificates (terminal or annual examination).
- **Assessment for learning** = formative, part of teaching, the **prerequisite** for planning teaching; **assessment of learning** = summative, after teaching; **assessment as learning** = the learner **assesses himself** (self-monitoring).

- **CCE = Continuous and Comprehensive Evaluation**; required by the **RTE Act, 2009, Section 29(2)(h)**; recommended by NCF 2005.
- **Bloom's taxonomy (1956)**, cognitive domain, lowest to highest: knowledge, comprehension, application, analysis, synthesis, **evaluation**. Revised (Anderson and Krathwohl, 2001): remember, understand, apply, analyse, evaluate, **create**.

Explained

1. Measurement, assessment, evaluation

Term	What it does	Example
Measurement	assigns a number	Ravi scored 32 out of 50
Assessment	gathers evidence of learning by many tools	tests, projects, observation, portfolio
Evaluation	judges worth against a standard	Ravi's 32 is "good" and he needs help in fractions

Gronlund's summary: **evaluation = measurement (or description) + value judgement**. Measurement is a part of evaluation, not the other way round.

2. Kinds of tests

Basis	Kinds
Purpose	achievement, aptitude (prognostic), diagnostic, intelligence, personality, attitude scale, interest inventory
Who makes it	standardised (experts; tried out, has norms and a manual) vs teacher-made (informal, for one class)
Form of items	essay type vs objective type (MCQ, true-false, matching, completion) vs short answer
Reference point	norm-referenced (pupil compared with the group) vs criterion-referenced (pupil compared with a fixed standard of mastery)
Administration	individual vs group; verbal vs non-verbal (performance)
Time	speed test (many easy items, strict time) vs power test (items of rising difficulty, enough time)

- Achievement looks at the **past** (what was learnt), aptitude at the **future** (what can be learnt), diagnostic at **why** a pupil fails.
- Attitude scales: **Thurstone, Likert**; aptitude battery: **DAT** (Differential Aptitude Tests).

3. Essay type and objective type

Point	Essay type	Objective type
Scoring	subjective, varies with the examiner	objective, with a key

Point	Essay type	Objective type
Coverage	few questions, small part of syllabus	many items, wide coverage
Tests best	organisation, expression, original thinking	recall, understanding, application
Weakness	low reliability, slow to mark	guessing, costly to prepare

4. Qualities of a good test in more detail

- **Validity** is specific to a purpose: a good maths test is not a valid test of reading.
- **Reliability is necessary but not sufficient** for validity: a valid test is always reliable, but a reliable test need not be valid. A scale that always reads 2 kg too much is reliable, not valid.
- **Objectivity** raises reliability; a teacher who wants **credibility** in valuing answer scripts must be **objective** (same criteria for every script, no favours).
- **Standardisation**: the same instructions, time and scoring rules for everyone.
- **Discriminating power**: the test separates good and weak pupils.

5. Ways to estimate reliability (each gives a correlation coefficient; .70 or more is usually acceptable)

Method	How	Shows
Test-retest	same test, same group, after a gap	stability over time
Parallel (alternate) forms	two equivalent forms of the test	equivalence
Split-half	odd and even items correlated; Spearman-Brown formula for the whole test	internal consistency
Rational equivalence	Kuder-Richardson (KR-20, KR-21), Cronbach's alpha	internal consistency

- A longer test, objective scoring, clear instructions and a wide range of ability raise reliability.
- Total variance of scores = **true variance + error variance**.

6. Kinds of validity

Kind	Question it answers	Example
Content	do the items cover the syllabus and objectives?	built in with a blueprint ; key for achievement tests
Concurrent (criterion)	does it agree with a present criterion?	new test vs teacher's ratings now
Predictive (criterion)	does it forecast later success?	entrance test vs later grades
Construct	does it measure the trait in the theory?	an anxiety scale behaving as theory says

- **Face validity** = the test only **looks** right to the user; it is not proof of validity.
- A **validity coefficient** (test vs criterion) is usually lower than a reliability coefficient.

7. Norms and derived scores

Norm	Score is compared with	Example
Age norm	average score of each age	mental age 10
Grade norm	average score of each class	reads at class 6 level
Percentile norm	percentage of the group below	PR 85
Standard-score norm	distance from the mean in SD units	z, T, stanine, deviation IQ

- Percentile: $P_{25} = Q_1$, $P_{50} = \text{median}$, $P_{75} = Q_3$. Percentile norms are the most widely used and easy to explain, but percentile units are **unequal** along the scale.
- A raw score has no meaning by itself; norms or a criterion give it meaning.

8. Types of evaluation by time and purpose

Type	When	Purpose
Placement	before instruction	entry behaviour, grouping
Formative	during instruction, often	feedback, correct and improve
Diagnostic	during, when difficulties persist	find causes, plan remedial teaching
Summative	end of unit, course or programme	grade, promote, certify

- **Prognostic** use (predicting success in a course or job) runs across these: aptitude and entrance tests.
- **Assessment for learning** "closes the gap" between where the pupil is and where he should be; pupils get feedback and take part. **Assessment of learning** sums up achievement against standards at a point in time.

9. CCE and exam reform

- **Continuous**: regular assessment through the year (formative and summative tests), not one annual examination.
- **Comprehensive**: covers **scholastic** (subjects) and **co-scholastic** areas (life skills, attitudes, values, health, arts, sports).
- Aims: reduce exam stress and rote learning, find learning gaps early, give remedial help.
- RTE Act 2009 also bars any **board examination** for a child **until elementary education is complete** (Section 30(1)). NEP 2020 asks for a **holistic, 360-degree progress card**.

10. Bloom's taxonomy of educational objectives

Domain	Levels (lowest → highest)	By
Cognitive (knowing)	knowledge, comprehension, application, analysis, synthesis, evaluation	Bloom, 1956
Cognitive, revised	remember, understand, apply, analyse, evaluate, create	Anderson and Krathwohl, 2001

Domain	Levels (lowest – highest)	By
Affective (feeling)	receiving, responding, valuing, organisation, characterisation	Krathwohl, Bloom and Masia, 1964
Psychomotor (doing)	imitation, manipulation, precision, articulation, naturalisation	R.H. Dave, 1970

From the options understanding, applying, analysing, evaluating, the **highest is evaluating**.

11. Planning a test: blueprint. A **blueprint (table of specifications)** spreads the marks over **content units × objectives** (knowledge, understanding, application) and item types. It gives the test **content validity** and balance.

Common traps

✗ "Subjectivity is a characteristic of a good test." ✓ A good test is **objective**; subjectivity is a defect.

✗ "A test that predicts the future of students is an achievement test." ✓ It is a **prognostic** test (aptitude tests are prognostic).

✗ "The best B.Ed. entrance test is an achievement test." ✓ An **aptitude** test, because selection must predict success.

✗ "Diagnostic test = prognostic test." ✓ **Diagnostic** finds present weaknesses and causes; **prognostic** predicts future success. (APPSC 2019's Telugu printed రోగనిర్ధారణ పరీక్ష, "diagnosis", for prognostic; the English key is prognostic.)

✗ "Evaluation at the end of a programme is formative or diagnostic." ✓ It is **summative**.

✗ "Assessment of learning is the prerequisite for teaching." ✓ **Assessment for learning** (formative) is.

✗ "A reliable test is always valid." ✓ A **valid** test is always reliable; a reliable test may not be valid.

✗ "Percentile rank 80 means 80% marks." ✓ It means 80% of the group scored **below** the pupil.

✗ "Measurement and evaluation are the same." ✓ Evaluation = measurement + **value judgement**.

✗ "Synthesis (or create) is below evaluation in every version of Bloom." ✓ Original 1956: **evaluation** is highest; revised 2001: **create** is highest.

Quick revision

- Evaluation = measurement + value judgement.
- Good test: validity (most important), reliability, objectivity, usability, norms; not subjectivity.

- Validity = measures what it claims; reliability = consistency; objectivity = same score from any examiner.
- Achievement = what was learnt; aptitude = prognostic, predicts; diagnostic = finds causes → remedial teaching.
- Placement before, formative during, diagnostic when problems persist, summative at the end.
- Assessment for learning = formative (prerequisite for teaching); of learning = summative; as learning = self-assessment.
- Reliability: test-retest, parallel forms, split-half (Spearman-Brown), KR-20; validity: content, concurrent, predictive, construct.
- Norms: age, grade, percentile, standard score; PR = % of the group below.
- CCE: continuous + comprehensive (scholastic and co-scholastic); RTE 2009 Section 29(2)(h).
- Bloom 1956: evaluation highest; revised 2001: create highest.