

01

Psychometrics simply refers to the psychological measurements; which include the fields of psychology and education devoted to testing, measurement, assessment, and related activities -NCME(National Council on Measurement in Education, USA).

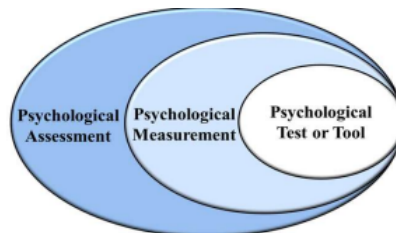
Psychometrics is a branch of psychology and education that focuses on the theory and techniques of psychological measurement. It involves the development, evaluation, and application of tools and techniques to measure mental abilities, traits, attitudes, and behaviors.

The field of **psychometrics** is often described as having **three hierarchical metric components**:

1. Testing – The Building Block: A test is a standardized set of questions or tasks used to gather specific data about a person's behavior, abilities, or characteristics. Test is the smallest unit of psychometrics. **Examples:** IQ tests, personality tests, achievement tests.

2. Measurement: Measurement refers to the process of assigning numerical values or scores to the results of a test according to certain rules. It interprets the data collected through tests to quantify psychological attributes (e.g., intelligence, anxiety, self-esteem). **Example:** Converting test performance into an IQ score or percentile.

3. Assessment – The Broadest Area: Assessment is a comprehensive process that involves collecting, integrating, and interpreting data from various sources. It includes tests, interviews, observations, background information, and more to make informed decisions. **Example:** A psychological evaluation for ADHD that includes tests, interviews with parents, and classroom observation.



Importance

Psychometrics is important because **psychological tests and measurements affect many areas of life, work, and research**. Here's why:

- 1. It affects everyone** - Whether a psychologist, researcher, teacher, or just a regular person, one is likely to encounter psychological measurements in some form.
Example - Job tests, student assessments, personality quizzes — they all use psychometric tools.
- 2. It's Used Across Many Professions** - Psychologists, teachers, and HR people use these tests to make decisions, like spotting learning problems or hiring the right person.
- 3. It measures important traits** - Psychometric tests are used to measure things like mental health (depression, anxiety), cognitive ability (IQ, memory), personality traits (introversion, openness).
- 4. Controls misuses of tests** - If someone doesn't understand how psychological tests work, they might misread or misuse the results. This could lead to wrong decisions. Psychometrics can be the educative media.
- 5. Research Advancement** - Psychometric tools enable researchers to collect quantifiable data, analyze psychological constructs, and advance theoretical knowledge in psychology and related fields.
- 6. Standardization** - It ensures that tests and assessments are consistent and standardized, so results are comparable across different individuals, groups, and settings.

7. **Objective Measurement** - Psychometrics provides a reliable and objective way to measure psychological traits, abilities, and characteristics that are otherwise difficult to quantify, such as intelligence, personality, and attitudes.
8. **Validity and Reliability** - Psychometric methods ensure that the tools used are valid (measuring what they are supposed to measure) and reliable (producing consistent results over time).
9. **Monitoring Progress and Outcomes** - It allows continuous evaluation and monitoring of changes in psychological states or abilities over time, useful in therapy, education, or employee development.

Assumptions

1. Psychological constructs exist.
2. Psychological constructs can be measured.
3. Although we can measure constructs, our measurement is not perfect.
4. There are different ways to measure any given construct.
5. Performance on tests can be generalized to non-test behaviors.
6. All assessment procedures have strengths and limitations.
7. Multiple sources of information should be part of the assessment process.
8. Assessment can provide information that helps psychologists make better professional decisions.
9. Assessments can be conducted in a fair manner.
10. Testing and assessment can benefit individuals and society as a whole.

Psychological Tests

A psychological test is "a systematic procedure for comparing the behavior of two or more people" (Cronbach, 1960).

A **psychological test** is defined as "essentially an objective and standardized measure of behavior" (Anastasi, 1982). These tests provide quantifiable insights into human behavior, personality, intelligence, aptitude, and other psychological constructs; reported as test scores.

Three core components of the definition:

1. Behavioral samples: Tests involve observing or eliciting behaviors from individuals.
2. Systematic collection: These behaviors must be gathered in an organized, consistent manner.
3. Comparison purpose: The goal is to compare behaviors either between different people or within the same person across different times or situations.

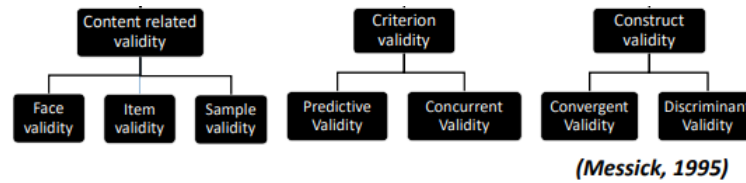
Basic Principles

1. **Validity:** Validity is the most important characteristic of a good test. It refers to the extent to which a test measures what it is supposed to measure and allows for appropriate interpretation of scores (Gay, Mills, & Airasian, 2012).

A test is valid if it accurately measures what it is supposed to measure. If it doesn't, then it is useless. To determine a test's validity, we compare its results with a real-world criterion.

For example: If a test says it can predict who will do well in medical school. To check if this test is valid, we compare the test results with actual performance in medical school. If students with high test scores also do well in medical school, the test is valid. If not, it's invalid.

Types of Validity:



2. **Reliability:** Reliability means dependability or trustworthiness. It refers to the consistency of test scores when people are retested with the same or an equivalent instrument (Hogan, 2014; Urbina, 2014). If a test is reliable, it will give the same results under the same conditions.

- **Types of Reliability:**

1. **Test-Retest Reliability:** The same test is given to the same group after a period of time to see if scores remain stable.
 2. **Parallel-Form Reliability:** This tests if two different versions of the same test give the same results. The versions should measure the same thing, but the questions are different. e.g., spelling tests.
 3. **Internal Consistency:** This tests if different parts of the same test measure the same thing consistently.
3. **Standardization & Norms:** Standardization means giving a test in the same way every time, with consistent conditions for both administering and scoring (Cohen et al., 2013). This allows comparisons between a person's scores over time and between different individuals.

3 criteria should be met for standardization -

- **Uniform Administration:** Tests should be held under a homogenous setting for all.
- **Uniform Scoring:** Must be scored carefully to ensure consistency and fairness.
- **Uniform Interpretation:** The results must be interpreted in a consistent manner for all.

Norms are average scores for specific groups, which help us compare how one person's score stacks up against others (Kaplan & Saccuzzo, 2013). Counselors should carefully check how test norms are created and consider creating their own local norms to reduce bias and prevent misuse of tests.

4. **Practicability:** The test should be feasible: not too long, quick to complete, and easy to score. Scoring should not require highly specialized personnel or complex procedures.
5. **Objectivity:** A test must be objective, meaning it should be free from personal bias so that all experts agree on the meaning of items and scoring.
 - **Objectivity of Items:** Test items should be written clearly so all test-takers interpret them the same way.
 - **Objectivity of Scoring:** The scoring method should be standardized so different examiners score the test uniformly.

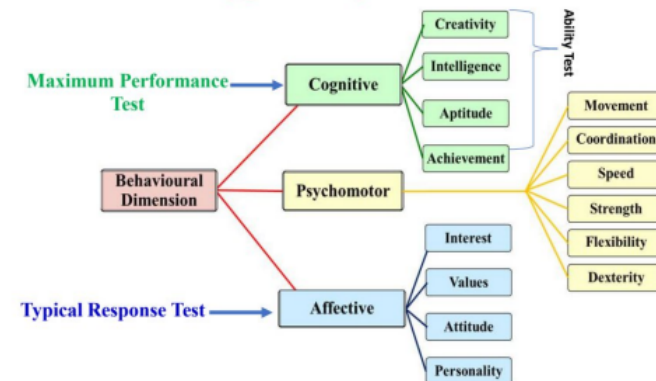
Good Characteristics এর জন্য এক্সট্রা ২টা

1. Types of Scales:

- **Nominal Scale:** Numbers or labels are used only for identification or classification. *Example:* Gender (Male = 1, Female = 2).
- **Ordinal Scale:** Subjects are ranked in order (e.g., 1st, 2nd, 3rd), but the exact differences between ranks are not measured.
- **Interval Scale:** Equal differences exist between points on the scale, allowing comparison and transformation of scores. *Example:* Temperature in Celsius.
- **Ratio Scale:** Has all properties of an interval scale plus a **true zero point**. *Example:* Height, weight, reaction time.

2. **Discriminatory Power:** Refers to how well a test can distinguish between two or more groups being assessed. It shows the accuracy with which a test or set of variables differentiates outcomes into distinct categories.

Based on Behavioral Dimension



1. Cognitive Tests/Ability Tests

Ability tests are a major category of psychological tests used to measure how well a person can perform in certain areas. These tests focus on skills, potential, and learning, and they are often used in education, career guidance, and research. Four common types are **intelligence tests**, **aptitude tests**, **achievement tests**, and **creativity tests**.

1. Intelligence Tests

Intelligence is often linked to linguistic and problem-solving skills. Many intelligence tests have an academic focus, such as the Stanford-Binet Intelligence Scale and Wechsler tests.

Examples: Stanford-Binet Intelligence Scale, Wechsler Scales.

2. Aptitude Tests

Aptitude tests are similar to intelligence tests but focus on a narrower set of skills. They measure a person's potential or ability to learn or improve in a specific task or skill through training or experience.

Two types:

- **Multi-aptitude Batteries:** Assess multiple abilities e.g., Scholastic Aptitude Test (SAT), the American College Testing (ACT) Assessment.
- **Component Ability Tests:** Assess a single skill (e.g., musical ability).

3. Achievement Tests

Achievement tests measure how much a person has learned in a subject. They can be either teacher-made or standardized:

- **Teacher-Made Tests:** Created by teachers to assess students on specific lessons. For example, a math teacher might design a test on algebra concepts taught in class.
- **Standardized Tests:** Developed by experts and used across schools or regions to compare student performance. Examples include the SAT, TOEFL, and GRE.

Teacher-made tests are flexible and customized, while standardized tests ensure consistency and comparison across different students.

4. Creativity Tests

Creativity tests focus on originality, imagination, and the ability to generate new ideas. These tests go beyond logical reasoning and look at how flexible and innovative a person's thinking is. For example, the Torrance

Tests of Creative Thinking ask participants to think of unusual uses for common objects like a brick or to complete incomplete drawings in novel ways. A creative person might say a brick could be used as a bookend, a tool for exercise, or part of an art sculpture.

2. **Psychomotor Tests**

These tests measure a person's motor skills, body coordination, and physical abilities. They are especially useful in fields like sports, defense, or jobs requiring fine motor skills. Instead of focusing on knowledge or reasoning, they look at **how well the body performs tasks**.

- **Movement & Coordination:** Tests may check balance, reaction time, or how smoothly movements are performed. For example, coordination tests are used for pilots or drivers.
- **Speed & Dexterity:** These measure how fast and accurately someone can do a task. A common example is the Purdue Pegboard Test, where a person places small pegs into holes quickly.
- **Strength & Flexibility:** Tests of grip strength, lifting ability, or stretching movements are used in physical fitness or occupational settings.

Example: In a factory job, a dexterity test might be given to see if a worker can handle small tools efficiently. In sports, agility and speed tests show how fit and responsive a player is.

3. **Affective Tests**

These tests look at the **emotional, motivational, and personality side** of human behavior. Instead of "maximum performance" (doing the best you can, like in an IQ test), affective tests focus on "typical responses" — how a person usually thinks, feels, or behaves.

- **Interest Inventories:** These measure what activities or careers a person is drawn to. For example, the Strong Interest Inventory helps students decide on career paths.
- **Values Tests:** These explore what principles or life goals are most important to someone, such as honesty, achievement, or social service.
- **Attitude Scales:** These check how people feel about issues, groups, or situations. For instance, an attitude test might measure prejudice or openness toward diversity.
- **Personality Tests:** These examine traits like introversion–extroversion, emotional stability, or openness to experience. Examples include the Minnesota Multiphasic Personality Inventory (MMPI) or the Big Five personality tests.

Example: Affective tests are often used in counseling, therapy, and organizational settings. For instance, a company might use a personality test to see if a candidate will fit well in a team.

Based on Administration

Psychological tests can also be grouped according to **how they are given (mode of administration)**. This is important because the way a test is conducted influences how comfortable people feel, how much time it takes, and how results are interpreted. Four broad types are individual tests, group tests, paper-pencil tests, and computer-assisted tests.

1. **Individual Tests**

Individual tests are given to one person at a time. The examiner works closely with the test taker, guiding them through the process. This allows the examiner to observe behavior, clarify instructions, and build rapport. Intelligence tests like the Stanford-Binet or the Wechsler Scales are examples. These are useful when detailed information is needed about a single person, but they take more time and resources.

2. **Group Tests**

Group tests are given to many people at once, often in classrooms or large halls. Instructions are given to the whole group, and test takers work on their own. Examples include school entrance exams, competitive exams, or

large-scale aptitude tests. These are quicker and more economical but give less chance to observe individual differences in detail.

3. Paper-and-Pencil Tests

Paper-pencil tests are the traditional format where people read questions on paper and write their answers with a pen or pencil. They can be objective (like multiple choice) or subjective (like essays). Most classroom exams and questionnaires are paper-pencil tests. They are easy to use but can sometimes be limited by handwriting or language skills.

4. Computer-Assisted Tests

Computer-assisted tests are conducted on computers, tablets, or even online platforms. They can be adaptive, meaning the difficulty of questions changes depending on how the test taker performs. Examples include GRE, GMAT, and many modern psychological assessments. These tests are efficient, give instant scoring, and reduce human error, but they require access to technology and may disadvantage those unfamiliar with computers.

Aspect	Individual Tests	Group Tests
Mode of Administration	One person at a time with direct examiner interaction	Many people tested simultaneously with common instructions
Observation	Examiner can closely observe behavior, effort, and response style	Limited observation of individual differences
Flexibility	Flexible pacing, examiner can clarify doubts	Fixed pace, little scope for clarification
Examples	WAIS, Stanford-Binet, Rorschach	Army Alpha & Beta, SAT, MATB
Advantages	High accuracy, detailed assessment, personalized	Economical, time-efficient, practical for large groups
Disadvantages	Time-consuming, costly, requires trained examiner	Less individual attention, risk of misinterpretation, less diagnostic depth
Best Suited For	Clinical diagnosis, research requiring depth	Educational testing, large-scale selection, aptitude screening

Based on Form/Item Content

1. Verbal Tests

- Items are presented in words, either spoken or written.
- They require reading, listening, and language skills.
- Example: A vocabulary test where you explain meanings of words.

2. Non-verbal Tests

- Items use symbols, pictures, diagrams, or figures instead of words.
- Useful for people with limited language skills or from different cultural backgrounds.
- Example: Raven's Progressive Matrices, where patterns are completed without words.

3. Performance Tests

- Involve doing an activity rather than answering questions.
- Often used to assess motor skills, problem-solving with objects, or speed.
- Example: Putting blocks together to form a design in the Wechsler Block Design subtest.

Based on Time Limit/Rate of Response

1. Speed Tests

- Contain many simple items that can be solved quickly.
- The main focus is **how fast** a person can work within a time limit.
- Example: Clerical speed and accuracy tests where you must match symbols rapidly.

2. Power Tests

- Items are fewer but more difficult.
- There may be no strict time limit, or the limit is generous.
- Aim is to measure the **level of ability** rather than speed.
- Example: Advanced mathematics or reasoning tests where accuracy matters more than quickness.

Based on Purpose

When we classify psychological tests based on **purpose**, we look at *why* the test is being given. Four common purposes are **ability testing, achievement assessment, diagnosis and screening, and neuropsychological assessment**.

1. Ability testing

Ability testing focuses on a person's potential or general capacity to perform tasks. These tests measure skills like reasoning, memory, or problem-solving. They are often used for educational placement, job selection, or research. For example, an IQ test such as the Wechsler Intelligence Scale tells us about general intellectual ability, while an aptitude test like the Differential Aptitude Test predicts how well a student might do in future learning or training.

2. Achievement assessment

Achievement assessment looks at what a person has already learned. These tests are tied to past education, training, or instruction. They are common in schools and professional certification. For example, a mathematics exam, a spelling test, or national assessments like the SAT Subject Tests all measure how much knowledge has been gained in a particular subject. Achievement tests are useful for evaluating progress, grading, or deciding promotion to the next level.

3. Diagnosis and screening

Diagnosis and screening tests are designed to identify difficulties, disorders, or special needs. They help in detecting problems early and planning treatment or support. For instance, depression inventories like the Beck Depression Inventory screen for depressive symptoms, while developmental screening tests identify learning disabilities in children. A clinical psychologist might use such tests to confirm whether a person has anxiety, ADHD, or another condition.

4. Neuropsychological assessment

Neuropsychological assessment examines how brain functioning relates to behavior and cognition. These tests are often used after brain injury, stroke, or neurological illness to understand which abilities are affected. They assess areas like memory, attention, language, and motor control. For example, the Wisconsin Card Sorting Test measures flexible thinking, while the Rey-Osterrieth Complex Figure Test checks visual memory and planning. Doctors and psychologists use these results to guide rehabilitation and treatment.

Based on Modes of Scoring

Some tests are scored manually, while others use technology.

- **Hand scoring** means the examiner checks answers and calculates results manually. This is common in small-scale testing or where personal interpretation is needed. For example, a teacher marking essay questions or a

psychologist scoring a projective test like the Thematic Apperception Test (TAT). Hand scoring allows flexibility but can be time-consuming and prone to human error.

- **Machine scoring** uses scanners, software, or computers to calculate results automatically. This is common in large-scale exams and standardized tests. For example, multiple-choice exams like GRE or IELTS are often scored by computer. Machine scoring is fast, consistent, and less biased, though it usually works best with objective-type questions.

Based on Criterion of Scoring

Here we look at whether the test has fixed, clear right-or-wrong answers or if judgment is needed.

- **Objective scoring** means there is only one correct answer, so scoring is straightforward. Multiple-choice, true/false, or matching tests are examples. For instance, in Raven's Progressive Matrices, there is a definite correct pattern to complete the figure, and everyone marking it will agree on the score. Objective scoring is reliable and leaves little room for examiner bias.
- **Subjective scoring** means answers are open to interpretation, so the scorer's judgment matters. Essays, projective tests, or interview-based assessments fall in this category. For example, in the Rorschach Inkblot Test, people describe what they see in ambiguous images, and the psychologist interprets the meaning. Similarly, essay questions in exams require evaluators to use their judgment to grade the quality of responses.

Based on Mode of Scopes/Adaptability

- **Culture-specific tests** are developed within a particular culture and reflect the language, values, and experiences of that culture. They assume that the test-taker is familiar with the cultural background of the test. For example, a vocabulary test in English assumes knowledge of English words, which may disadvantage non-native speakers. Similarly, intelligence or aptitude tests designed in Western countries often reflect Western education and life experiences. These tests work well within the same cultural group but may not be fair across cultures.
- **Culture-free tests** (also called *culture-fair tests*) are designed to minimize the influence of language, education, and cultural knowledge. They usually use nonverbal items such as patterns, shapes, or figures, so that performance depends more on reasoning ability than on cultural background. A well-known example is **Raven's Progressive Matrices**, where test-takers see a puzzle where a pattern is missing one piece, and they must choose the correct piece to complete it. This doesn't need language or cultural knowledge—it just checks their reasoning.

Mode of Interpretation - Criterion-Referenced vs. Norm-Referenced Test

Aspect	Criterion-Referenced	Norm-Referenced
Purpose	Measures what a student knows or can do .	Measures how a student compares to others .
Comparison	Against a fixed standard or skill level .	Against a norm group (other test-takers) .
Example	Driving test (need 70% to pass).	IQ test, SAT, or school ranking test.
Result Meaning	Shows if the student met the criteria .	Shows if the student is above or below average .
Focus	Mastery of skills .	Relative performance .
Use	For learning and instruction .	For ranking or selection .
Main Limitation	Doesn't show how others did.	Doesn't show what specific skills were mastered.

Mode of Standardization - Standardized vs. Non-Standardized Tests

Aspect	Standardized Tests	Non-Standardized Tests
Definition	Constructed, administered, scored, and interpreted in a prescribed, precise, and consistent manner (Kelley, 1914). Designed to minimize external influences and ensure reliability (Nunnally & Bernstein, 1994).	Do not follow strict procedures for construction, administration, scoring, or interpretation; may vary depending on context, examiner, or situation.
Development	Carefully developed and rigorously tested before general use.	Often developed informally, without extensive pre-testing.
Administration & Scoring	Consistently administered and scored across all examinees.	Administration and scoring may differ between settings or examiners.
Interpretation	Interpreted on the basis of established norms.	Interpretation depends on local context or examiner judgment.
Relation to Norm-Referencing	Standardized tests are one type of norm-referenced measure, but not all norm-referenced tests are standardized.	May or may not be norm-referenced; often criterion- or teacher-referenced.