PRINCIPLES OF PSYCHOLOGICAL ASSESSMENT SUMMARY NOTES

Table of Contents

WEEK 1 – INTRODUCTION

Historical development of tests – recognises scientists associated with each important development and can illustrate the influence of society on testing development

Historical development of tests	Scientist and time period	Key ideas
Categorisation	Mid 19th century **Esquirol** 1838	Described degrees of intellectual deficiencies on a continuum from normalcy to idiocy using a language ability test.
	Seguin 1866	Attempted to train intellectually impaired using sensory and muscle training. He developed a tool to measure the effects of training – *Sequin form board* – a performance test
		Both important in developing tests for intellect to distinguish between those who had intellectual deficits and those who were emotionally disturbed.
Early experimentalists	Mid 19th century **Wundt** 1879	Standardised procedures, hallmark of tests
Individual differences	**J.M. Cattell** 1890	Developed tests of sensory ability, muscle strength, memory and reaction time. Criticised due to little correspondence between scores on test and scores on scholastic aptitude. First time term "mental test" used in published paper.
	Ebbinghaus 1897	Tested school children in arithmetic, memory span and sentence completion (language ability). He found the best correspondence with scholastic achievement. Similar tests are used today. Measured complex cognitive behaviour rather than simple abilities -> better correspondence w/ scholastic aptitude.
	Galton **Kraepelin** 1895	Assessed psychiatric patients' ability to perform arithmetic problems to measure individual differences in intellect.
Intelligence testing	Early 20th century **Binet** (1905, 1908, 1911) – Binet-Simon Scale	30 problems increasing in difficulty involving judgement, comprehension and reasoning. 1908 increased number of items and grouped tests based on age that 90% could pass (i.e. 3y.o. tests, 90% of 3y.o. could pass) – score indicated *mental age*. First instance of normative data which compared ability to normal population/sample.
	Terman 1916 Stanford-Binet Scale	First development of a standardised score of intelligence/individualised test. First use of "IQ" score: *mental age / chronological age x100*. IQ=100: mental and chronological age the same; <100: mental < chron.; >100: mental < chron.

	Kuhlman-Binet scale 1912 **Wechsler** 1939 Modern day tests of intelligence	Further revision. Extension of items on test – can be used in children as young as 3 months Individual tests of adult intelligence -> developed into modern day tests (Wechsler scales). *Deviation IQ* -> individual score – mean score / SD. • Wechsler Adult Intelligence Scale (**WAIS-IV**) • Wechsler Intelligence Scale for Children (**WISC-V**) • Stanford-Binet (**SB-5**)
Group testing	World war I – 1917 **Otis** and **Yerkes**	Army Alpha and Army Beta efficiently tests and allows army recruiting of high intellect. The alpha was a routine test and the beta was a test for illiterate and foreign-born recruits. Showed that psychological tests must account for societal influence of how those in psychiatric institutions were dealt with and how they distinguished between those with emotional disturbance and intellectual deficit. Precursors of today's group tests
Personality testing	**Kraepelin** 1892 – Free association **Woodworth** WWI – Personal data sheet **Rorschach** 1921 – inkblots **Hathaway** and **McKinley** 1942 – Minnesota test Modern day tests of personality:	Measure of psychiatric illness. List of words associated with a given word. Self-report test of personality – those with emotional disturbance were excluded from the army. Allowed efficient testing for the needs of society. Laid the basis for modern tests, especially self-report. Projective test applied to psychiatric patients to see differences in interpretation. Objectivity and validity questioned. Extensive test that was an indicator of differences on personality domains. Redeveloped and used today. Minnesota Multiphasic Personality Inventory (MMPI). • Sixteen PF – 5th edition • MMPI; MMPI – 2 – RF • NEO-PI-3 -> based on 5 factor model of personality.

Use of tests – can define a psychological test and can explain why a procedure is or is not a psychological test.

- A **psychological test** is a *standardised measure* of a *sample of behaviour* -> we can't measure entirety of personality, so we make inferences using this sample.
    - Differences in scores are taken as an indicator of real differences, rather than due to measurement error.

The main types of psychological tests are *intelligence/ability* – maximum performance – and *personality* – typical performance.
Both test types measure individual differences.

What areas of psychological practice are psychological tests used in?

Psychological tests are used to provide scores that enhance decision making in a variety of contexts and is used to classify individuals as either emotionally disturbed or intellectually defective.

- **Clinical** – diagnosis and severity of symptoms and impact on functioning of those with psychological disorders. Also used to measure progress throughout treatment
- **Schools** – measure attainment of knowledge (exams), used for selection into special ability classes
- **Industry and business** – selection of staff
- **Counselling** – measure emotional wellbeing or interpersonal relationships. Enhance self-understanding and personal development.
- Research – measure individual differences to advance understanding of psychological constructs. Central to theory development.

Where can tests be found (sources)?

Databases are available at MQ library (limited to postgraduate students, https://libguides.mq.edu.au/psychology):

- PsycInfo
- PsycTests

Some reputable internet sites include:

- https://www.apa.org/science/programs/testing/find-tests
- https://libraries.uta.edu/tmdb/
- https://buros.org

Restricted availability of tests – why is the use of psychological tests restricted?

- To prevent familiarity resulting in invalidity

- To prevent abuse

Appropriate and valid administration of psychological tests requires training and thus access is limited to those with the proper knowledge. Also, if tests were freely available, people could practice their responses to items, thus invalidating the results by familiarity.

WEEK 2 – NORMS AND RELIABILITY

Test scores

Raw score: the score of an individual on a test. This is not meaningful and needs to be compared to a standard in order to compare with other scores:

- Criterion referencing – specify a standard that needs to be met
- Norm referencing – compared with scores from a large representative sample (norms)

We need a frame of reference/standard to compare raw scores with to be meaningful and see how it differs from others' scores.

Derived scores (norms): allow us to ascertain an individual's position relative to a standardisation (or normative) sample. Provide comparable measures that permit a comparison across different tests and between individuals.

Type of derived score	About	Advantages	Disadvantages
Percentiles	Percentile score (P) = percentage of people in the sample who fall *below* a particular raw score • ((n < x)/n) x100 – number of people with score below x divided by the size of sample. Special points: • P50: the median, 50% of sample below • P25: first quartile, 25% below • P75: third quartile, 75% below	Easy to compute, readily understood, universally applicable (applied to any test score)	Inequality of units especially at extremes. The increase between scores 1SD apart is not consistent The space between points on percentile scale is not the same as space between raw scores.
Standard scores	**Linear z** (most common form) (z) = (x – M)/SD • z = raw score – mean / SD Allows comparison of scores for individual across tests/on different distributions - retains exact numerical relationship between raw scores compared to others. z ranges from +/- 3 in a normal distribution. • Creates meaning to compare scores	Good for statistical analysis Allows comparison across distributions	
	Normalised standard scores: linear standard scores will be comparable only if they come from similar distributions (normal distribution). Raw score → percentile → normal curve frequency table → normalised z score		Scores have to be from similar distributions (can be fixed by normalising distribution)
	Deviation IQ: another linear transformation of z score • Deviation IQ = z x 15 + 100 A standard score with a mean of 100 and SD of 15	Ensures comparability of IQs	
Age scores	**Age scores**: only use with properties showing a monotonic increase with age, e.g. height, weight, intelligence during childhood. • IQ = mental age / chronological age x 100 Original IQ score was an age score. As age increases, SD varies.		Not used as often as scores have to have monotonic variation age

Specificity of norms

Norms are specific to the population from which they are derived – they are not absolute, universal or permanent. Need to be careful of this when comparing that they are representative.

Many psychological tests have normative samples made up of individuals from WEIRD backgrounds. There may not be Australian norms for psychological tests. It is the responsibility of the psychologist to understand the nature of the normative sample and to interpret test scores for an individual in light of any limitations in the normative sample.

Judging the quality of a test

A test provides an objective and standardised measure of a sample of behaviour, via uniformity of:

- **Administration** – using the same materials, instructions, time limits, environment etc.
- **Scoring**
- **Interpretation** of test scores