Week 1 – Introduction to psychological testing

Fundamentals of psychological testing

All tests have:
- Standardised procedures
- Meaningfully described outcomes
- Norms and standards
- Made up of items

Types of tests
- **Intelligence:** ability
- **Aptitude:** capacity/potential; specific task or skill
- **Achievement:** previous learning/accomplishment; your degree
- **Creativity:** novel or original thinking
- **Personality:** traits/behaviours
- **Interest inventory:** preference for activities/choice; occupational testing
- **Behavioural procedures:** describes or count behaviour
- **Neuropsychological:** cognitive, sensory, perceptual or motor control

Why we use psychological testing:
- Classification
- Diagnosis & treatment planning
- Self-knowledge (decisions about actions & understanding choices)
- Program evaluation (impact of intervention, e.g. NAPLAN)
- Research

Responsibilities
- *Test developers and publishers:* test construction; evaluation standards
- *Test administration:* who should be using the test
- *Test takers:* consider the impact on individuals
- *Society:* individual differences acknowledge; but systematise society (e.g. who is schizophrenic?)
- *Other parties:* don't always have the same intentions as psychologists

Factors affecting testing
- Test characteristics
- Standardisation
- Psychometric properties
- Test-taker characteristics
- Test-administrator characteristics
- Familiarisation
- Rapport
- Anxiety & motivation
- Purpose/reason for test

HISTORY OF PSYCHOLOGICAL TESTING
- China (206BC) – developed most thoroughly for jobs within the public office
- Britain (early 19th century) – class distinctions for social order; intelligence = social position
- France (late 19th century) – an individuals worth is determined by their ability/merit; tests categorised individuals; Alfred Binet developed 'mental age' and IQ score
- World War 1 – pencil and paper tests; required reading ability (army alpha); intelligence in illiterate adults (army beta)
- World War 2 – group intelligence was reaffirmed; gave rise to the role of clinical psychologist
- 20th century – Wechsler-Bellevue Scale; Personality testing – projective (i.e. TAT and Rorschach); structured (MMPI)

PSYCHOLOGICAL TESTING VS. ASSESSMENT

	Testing	Assessment
Objective		
Definition	Gauge an individuals ability Scored and categorised	Answer a referral question Solve a problem Uses multiple tools of evaluation
Example: Childhood behavioural problems	Will measure child's behaviour	Will ascertain a diagnosis through: interviews; rating behaviour; observations; measuring cognitive ability; examining history/social factors; testing
Process		
Definition	Administer and score in accordance to specific rules (manual for use)	Consider processes beyond the score and how individuals got a score
Example: Vocational testing	Will administer and score a measure	Will select tests (or part of tests) with consideration of the individual factors Interviews; rating behaviour; observations; measuring cognitive ability; examine history and social factors
Evaluator role		
Definition	Should have no influence	Key to process Selection of tools and formulation of conclusions
Example: Clinical testing	Administer for standardised depression inventory	Formulates a diagnosis based on: interviews; observing behaviour/presentation; measuring behaviour; examining history and social factors
Outcome		
Definition	The final score(s)	Answer to referral question
Example: Geriatric testing	Indicates level of functioning	Decision regarding ability to live independently or requirement of assisted living

ASSUMPTIONS OF PSYCHOLOGICAL TESTING AND ASSESSMENT
- Psychological traits and states exist
- Psychological traits and states can be quantified and measured
- Test-related behaviour predicts non-test-related behaviour
- Tests and other measurement techniques have strengths and weaknesses
- Various sources of error are part of the measurement process
- Testing can be conducted in a fair and unbiased manner
- Testing and assessment benefits society

NORMS AND BASIC STATISTICS FOR TESTING
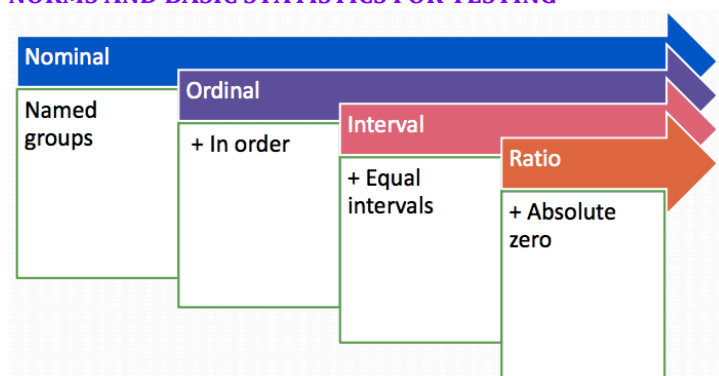
Examples:
Nominal: Lab Class 1, Lab Class 2 → we don't know anything about them other than their names
Ordinal: 1st, 2nd, 3rd in a race → doesn't tell the distance between people; Likert scale: strongly agree, agree, somewhat agree, etc.
Interval: IQ; Temperature in Celsius → can't say that 10º is twice as hot as 5º; 0º ≠ absence of heat
Ratio: height, weight → 10kg is twice as heavy as 5kg; 0kg = absence of weight

Basic statistics
- **Inferential statistics:** a statistical procedure that allows one to make inferences about large groups by examining a smaller sample
- **Percentile ranks:** What % of scores fall below a particular score → N = total cases; B = number of cases below a score of interest; X_i = score of interest
- **Quartiles:** divide frequency distributions into equal fourths
- **Deciles:** divide frequency distributions into 10 equal groups
- **Mean:** sum of all scores, divided by the total number of scores
- **Standard deviation:** average deviation around the mean
- **Z score:** a distribution with a mean of 0 and a SD of 1
- **T score:** a distribution with a mean of 50 and a SD of 10
- **Variance:** how much scores within a distribution differ among themselves
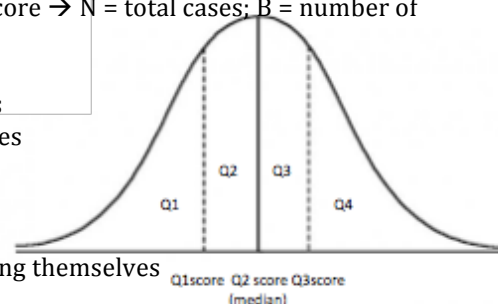
Norms
- Z-scores, percentiles, quartiles and means are all versions of norms
- Used to give info about a population based on the observations of a standardised sample
- Norm-referenced tests: NAPLAN (age-related performance based on peers of the same age); 'Tracking' of new born babies on certain characteristics; culture free intelligence tests
- **Criterion-referenced tests:** mastery of specific skill, e.g. driving test, exam – everyone's on the same playing field regardless of age, gender, culture; there's a certain criteria you have to meet
- **Within-group norming:** comparing an individual's test score only with members of his/her own racial group

Issues with norms
- Initial sample must be large enough to adequately represent the population
- Need to be regularly updated to accurately reflect the population
- Criterion-referenced tests are sometimes based on arbitrary cut-points