

# **Lecture 07: Psychometrics**

## **Scale Reliability and Validity**

Dr. Gordon Wright

November 20, 2023

### **This week**

- Social Psychology submission due
- 2 parts to the lab
  - Ethics II - the Goldsmiths Portal
  - Reliability and Validity recap materials

### **Any Questions?**

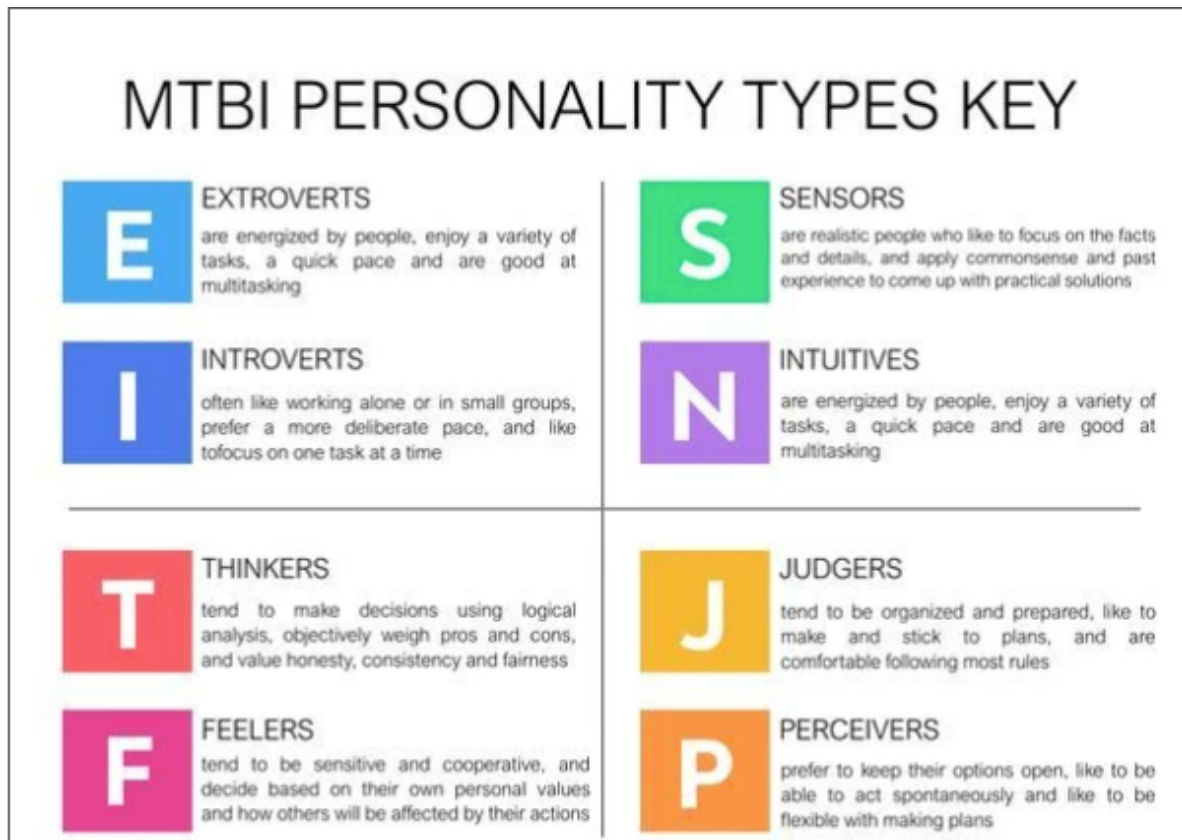
### **Key topics today**

### **How do we measure or assess psychological concepts and constructs?**

### **Psychometrics; the science of psychological assessment.**

General reader: Breakwell, Smith & Wright (2012) – Chapter 7 (available via reading list free online)

## What Myers-Briggs type are you?



## Myers-Briggs...

- Based on Jung's non-scientific ideas about personality
- The four dimensions are binary. But most characteristics are normally distributed
- Very poor test-retest reliability.
- Almost no research support.
- Company behind the test CPP makes \$20 million a year from it. Has little incentive to start from scratch!

<https://www.vox.com/2014/7/15/5881947/myersbriggs-personality-test-meaningless>

## **What is psychometrics?**

- Meaning from Greek origin: 'measuring the soul'
- Psychometrics is the field of study concerned with the theory and technique of psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits
- Refers to all areas of psychology concerned with psychological measurement (methods of testing and substantive findings)
- Two major research tasks:
  - (i) the construction of instruments and procedures for measurement;
  - (ii) the development and refinement of theoretical approaches to measurement

## **A brief history of psychometrics**

- Charles Darwin's (1809–1882) Origin of the Species impacts scientific thinking in 19 th century
- Evolution (anthropology) combined with quantification (allure of numbers)
- Francis Galton (1822–1911) builds on cousin Darwin's ideas with measurement and statistics

## **A brief history of psychometrics**

- Galton developed the theory underpinning correlation and regression
- Used this theory to try to explain the heritability of human ability and achievement (amongst many other things)
- Developed a lab and tests for many concepts e.g. prayer, boredom, beauty

### **What is a psychometric test?**

- Sample of affect, behaviour, cognition etc
- Obtained under standardized conditions
- Scored using rules that provide allow for comparison of individuals
- Ideally, we would like:
  - Multiple samples
  - Multiple situations (contexts, several occasions)
  - Multiple methods

### **But you can't always get what you want...**

Often, must measure individuals on

- One occasion
- Timed/ restricted conditions

So must use efficient methods

- Many opportunities (multiple choice tests)
- Objective scoring (no judgment involved)
- Adaptive item selection

### **Differences between a psychometric test and a general survey**

- Scientific rationale
- Careful item development and test construction
- Objective
- Standardised
- Instructions
- Scoring procedure
- Reliable
- Valid

### **Clinical uses of psychometric tests**

- Describe current functioning
- Further investigate impressions from less formal evaluation approaches
- Identify therapeutic needs
- Aid in differential diagnosis of disorder
- Monitor treatment over time to monitor success and identify new treatment needs
- Provide empathetic feedback

### **Occupational uses of psychometric tests**

- Initial hiring
- Job selection
- Team development
- Career counseling
- Training readiness
- Succession planning
- Performance assessment
- Promotion

### **Educational uses of psychometric tests**

- Counseling
- School exams
- University entrance exams
- Course exams
- Learning disabilities

### **Types of psychometric tests**

Maximum performance test (can do)

- Intelligence tests (basic reasoning ability common to a variety of intellectual tasks)
- Attainment tests (mastery tests, e.g., your exams, certification testing)

Typical performance test (will do)

- Personality tests (ways of thinking, feeling and behaving)
- Careers and interests tests

- Different answer demands: effort versus candid truth
- Context dependent

### Examples of maximum performance items (ability)

Odd one out

Tree, Man, Paper, Mouse

Next in sequence

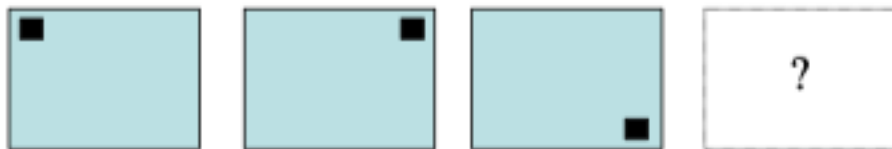
1, 1, 2, 3, 5, 8...

### Spatial reasoning

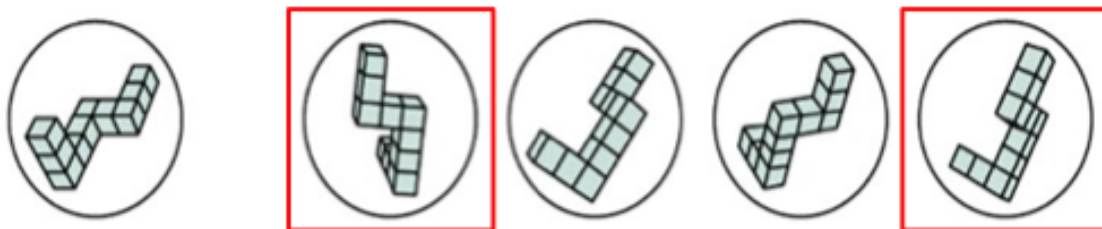
First 3 form a series,

Which comes next A, B or C ?

### Stimulus



### Image rotation task



### Examples of typical performance items

Rate on a scale from 1 to 5 how true this is of you

(Costa & McCrae, 1992, Big Five)

Once I find the right way to do something, I stick to it

Dichotomous yes/ no answers

(Eysenck & Eysenck, 1976, Giant 3):

I am the life of a party

Forced choice

(Zuckerman, 1979, Sensation Seeking Scale)

A: I like "wild" uninhibited parties

B: I prefer quiet parties with good conversation

### Properties of Psychometric tests

Properties of psychometric tests

Two important properties of psychometric tests

Reliability

-The consistency with which a test measures the construct

Validity

-The degree to which a test actually measures what it claims to measure "accuracy"

### Essential properties: Validity

A test is valid if it assesses what it claims to measure

The validity of an assessment strategy is the extent to which the strategy yields a reasonably accurate estimation of the characteristic or phenomenon in question.

Many steps to achieve validity (including concurrent validity, predictive validity, construct validity and face validity)

## Essential properties: Reliability

Test retest reliability

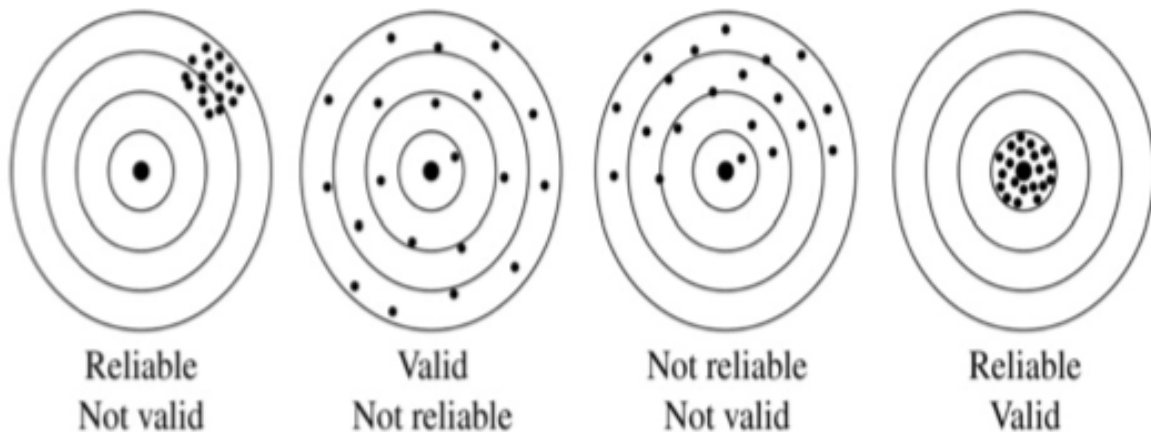
- Rule of thumb  $r$  between the two test times, 3 months apart  $> 0.7$  (just under 50% agreement)
- Test re-test reliability is not perfect – never reaches 1: beware real changes!

Internal consistency reliability

- Internal consistency is the degree to which all items are measuring the same construct
- Cronbach's Alpha should be greater than .70 for scales with items  $> 10$

## Reliability and Validity

I like to think of them as Consistency and Accuracy



## Different types of tests - raters

Behavioral observation (observer-rated)

- People scored according to behaviors observed by a rater
- Used frequently in work and clinical settings (e.g. Performance appraisal)

Self-report

- Subjects indicate their level of agreement or preference concerning statements reflecting attitudes or behaviors



- Response distortion is a problem (e.g. faking a personality test)

## Standardizing psychometric test scores

The raw score on many psychometric tests is based on an arbitrary scale

To give the scores meaning, we compare a person's scores to a meaningful comparison group

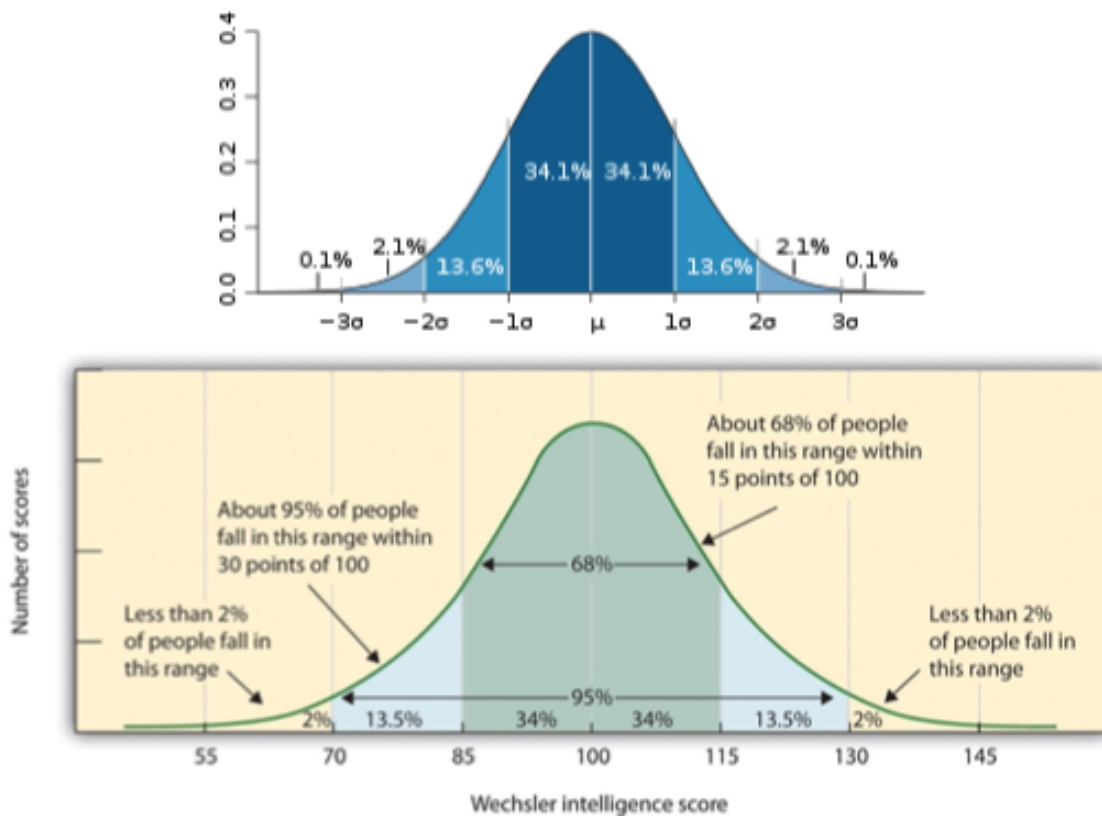
Statistical basis: Normal distribution

Most human traits approximate to normal curve

-Largest number of cases cluster in centre

-Area under curve can be closely specified from mean and standard

## Intelligence



**Next week...**

We might go about developing our own Psychometric Test.. if you want.

**References**