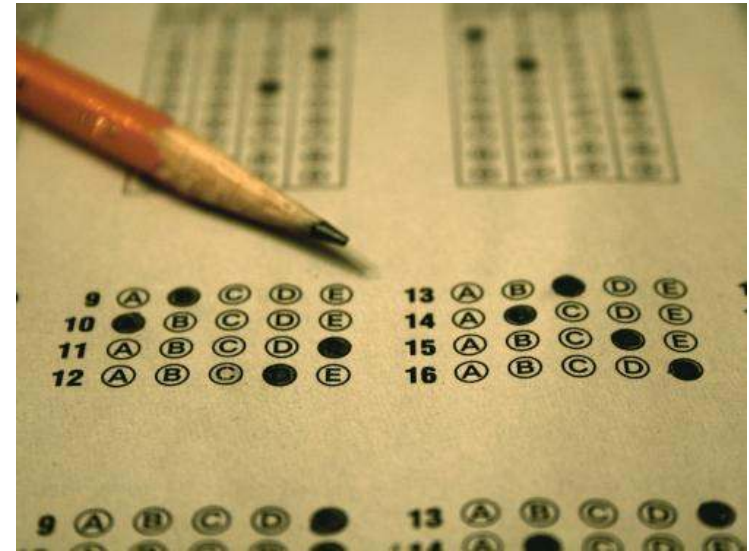


Psychological Test Construction Workshop

Jenny Symonds
jes81@cam.ac.uk

Psychological tests

- The measurement of specific psychological phenomenon in individuals and in groups.
Not simply a 'descriptive' survey.



- Psychological tests use psychometrics (the science of psychological experiments) and therefore **have psychometric properties** (aspects of the test that measure phenomena and can be tested for validity and reliability – i.e. scales, items, images)
- Occur on paper, online or as computer programmes

Examples of psychological tests

IQ tests	IQ (Binet 1905)
Achievement tests	British Ability Scales
Cognitive tests	Cognitive ability tests (CATs) Concept formation (Vygotsky) Spatial and verbal reasoning
Personality trait tests	The Big Five (Costa and McCrae 1986) Myer Briggs Type Indicator (MBTI)
Personality functioning (neuroticism, extroversion etc)	Junior Eysenck Personality Inventory (Eysenck 1965)
Self construct tests	Self-perception (Harter 1985) Self-esteem (Rosenburg & Simmons 1972)
Tests of psychological motivation and self-efficacy	Self regulation Self efficacy (Bandura 1986)

Test Administration

X

Single Administration

Cross-sectional

i.e. a single test of self-esteem

X O X

Pre-test post-test experimental design

Longitudinal

i.e. test cognitive ability, apply a teaching intervention, then retest cognitive ability

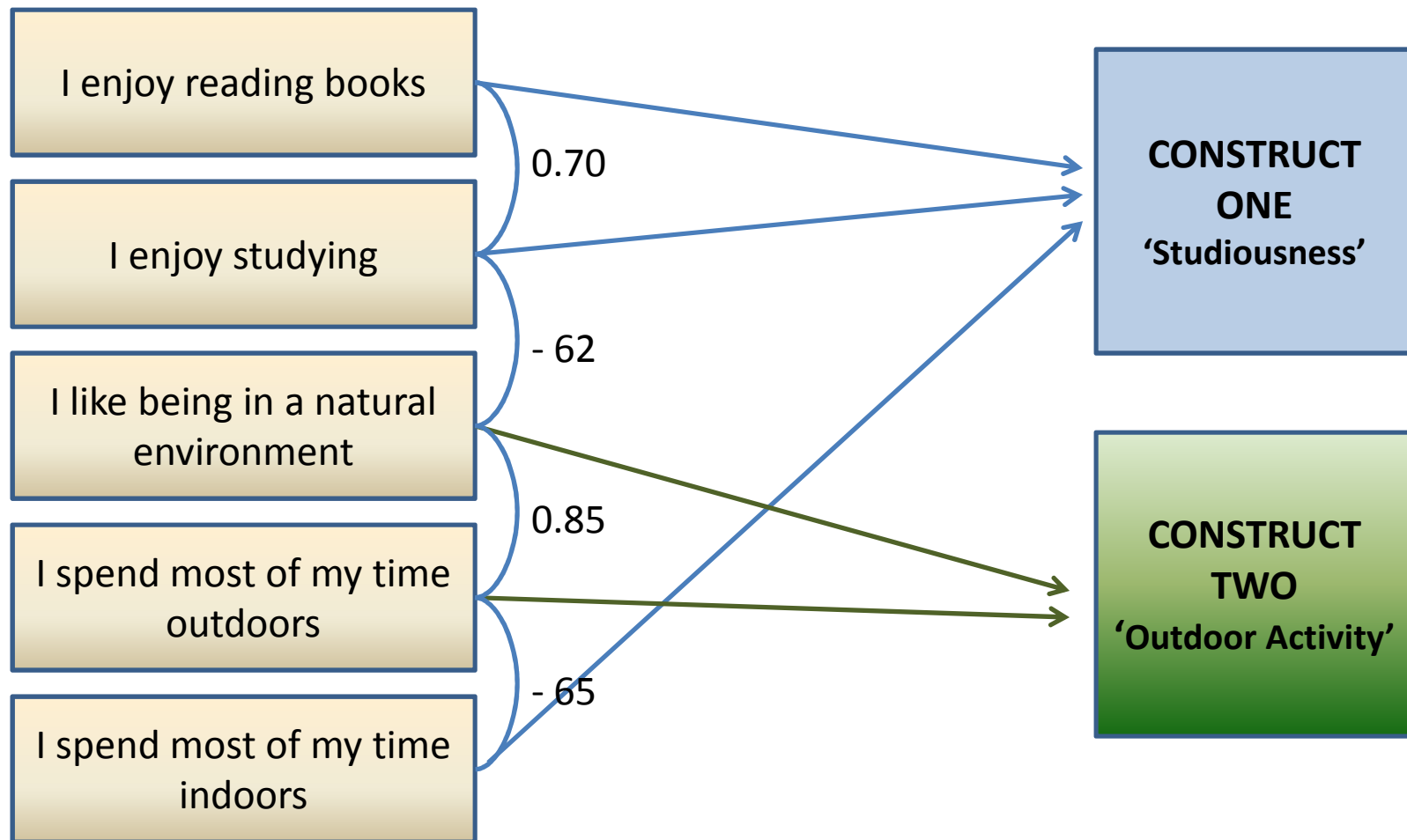
X X X

Multiple administration (in 'waves' across time) Longitudinal

i.e. give the same test at different time points (terms 1, 2 and 3 of the school year) to measure developmental change

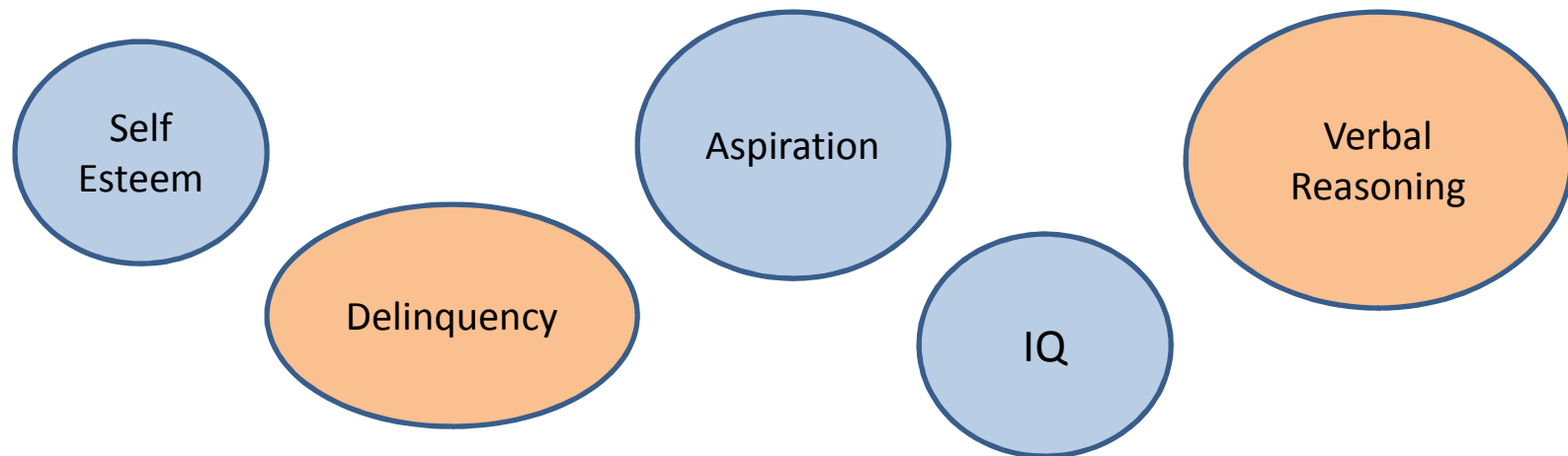
How are Tests Constructed?

Items are tested for their correlation with each other (**factor analysis** - as below) or for their applicability to a specific population of people (**criterion-keyed method**)



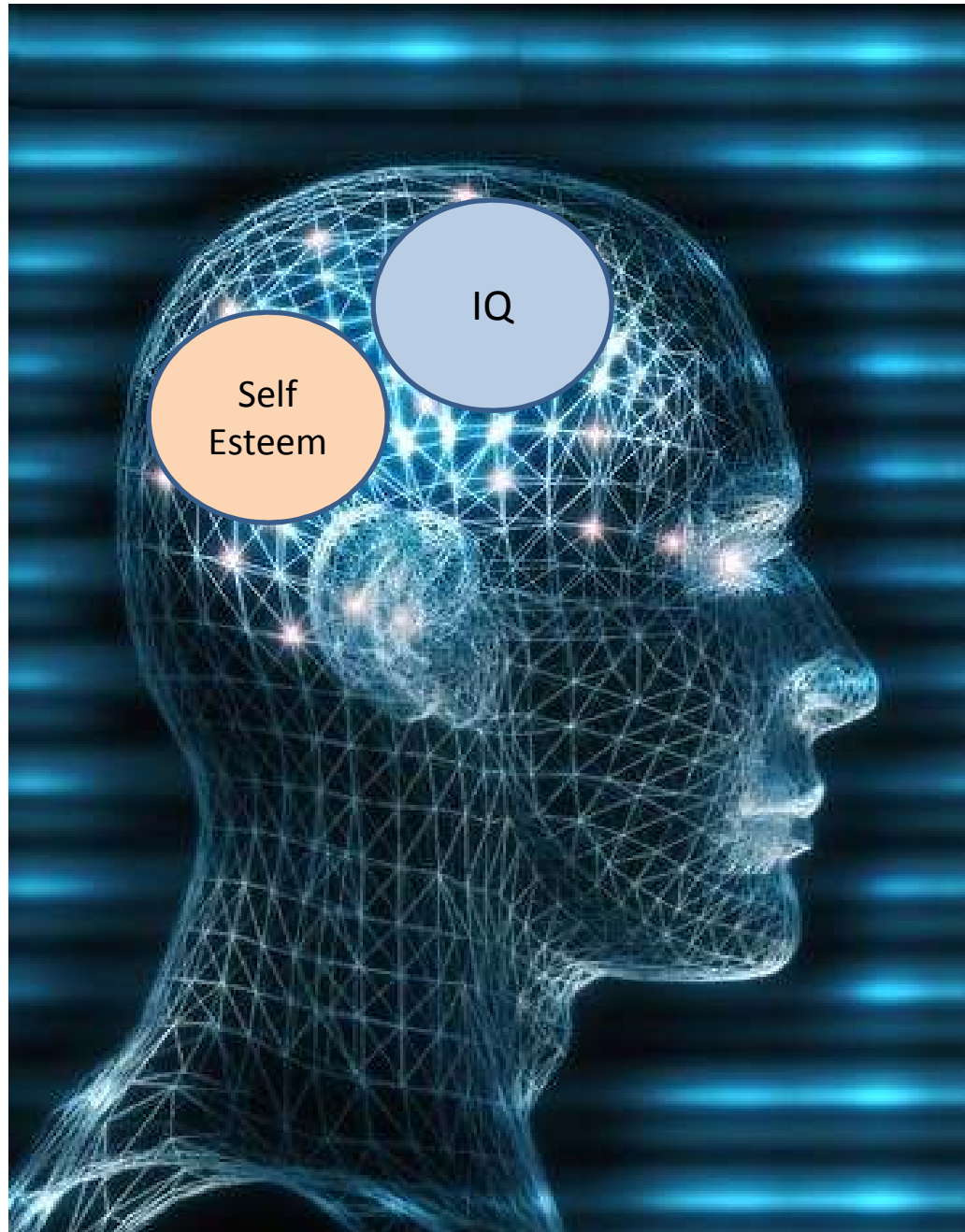
Psychological constructs

- You might be familiar with these commonly used constructs:



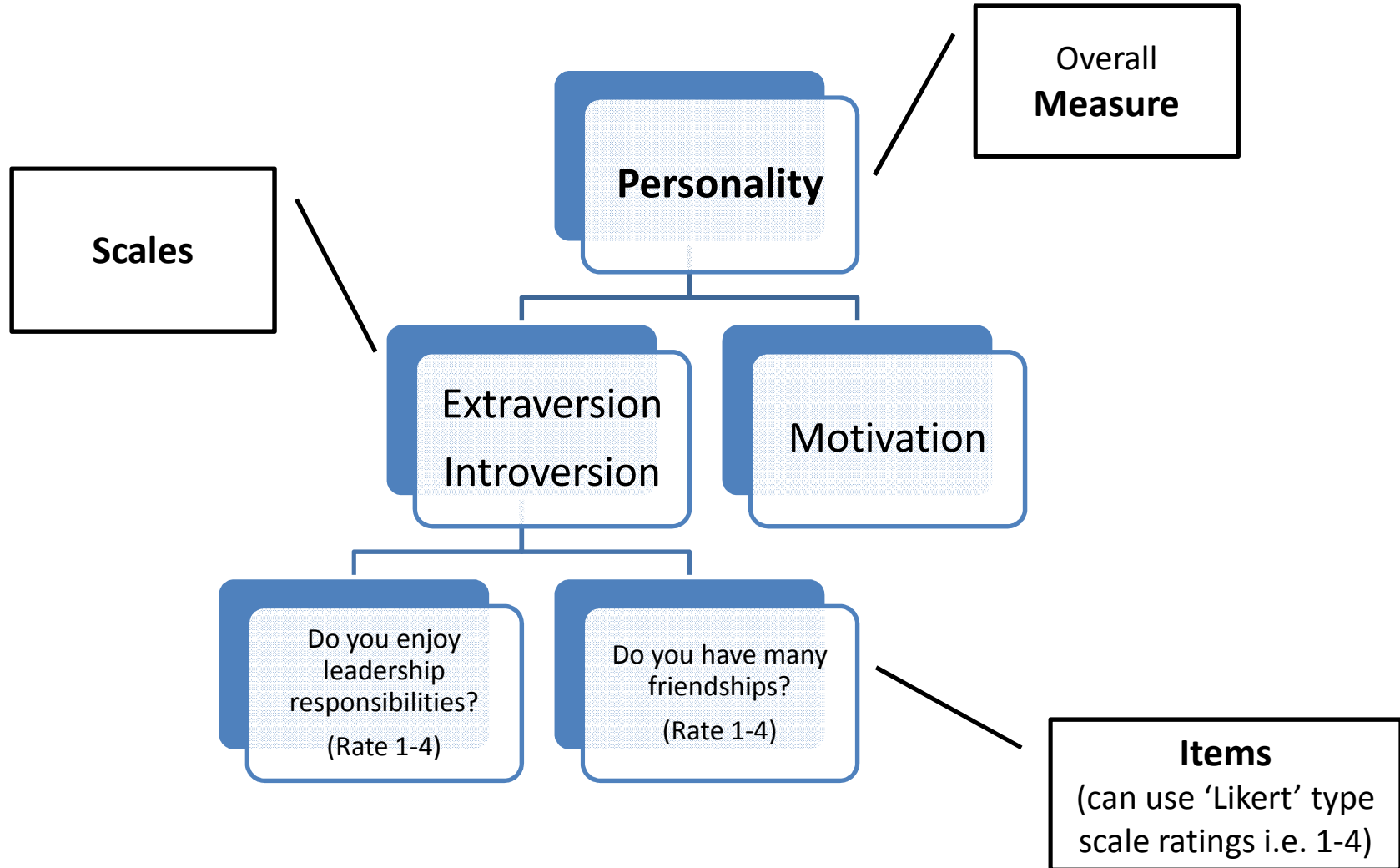
Question: do you think that psychological constructs are objective measures?

Note the difference between 'personal' (Kelly 1955) and psychological or 'hypothetical' (Coolican 2004) constructs. The latter are construed by psychologists for use in research and practice. The former refer to thought constructs existing in the mind, i.e. aspirations. This definition includes our personal conceptions of psychological constructs!

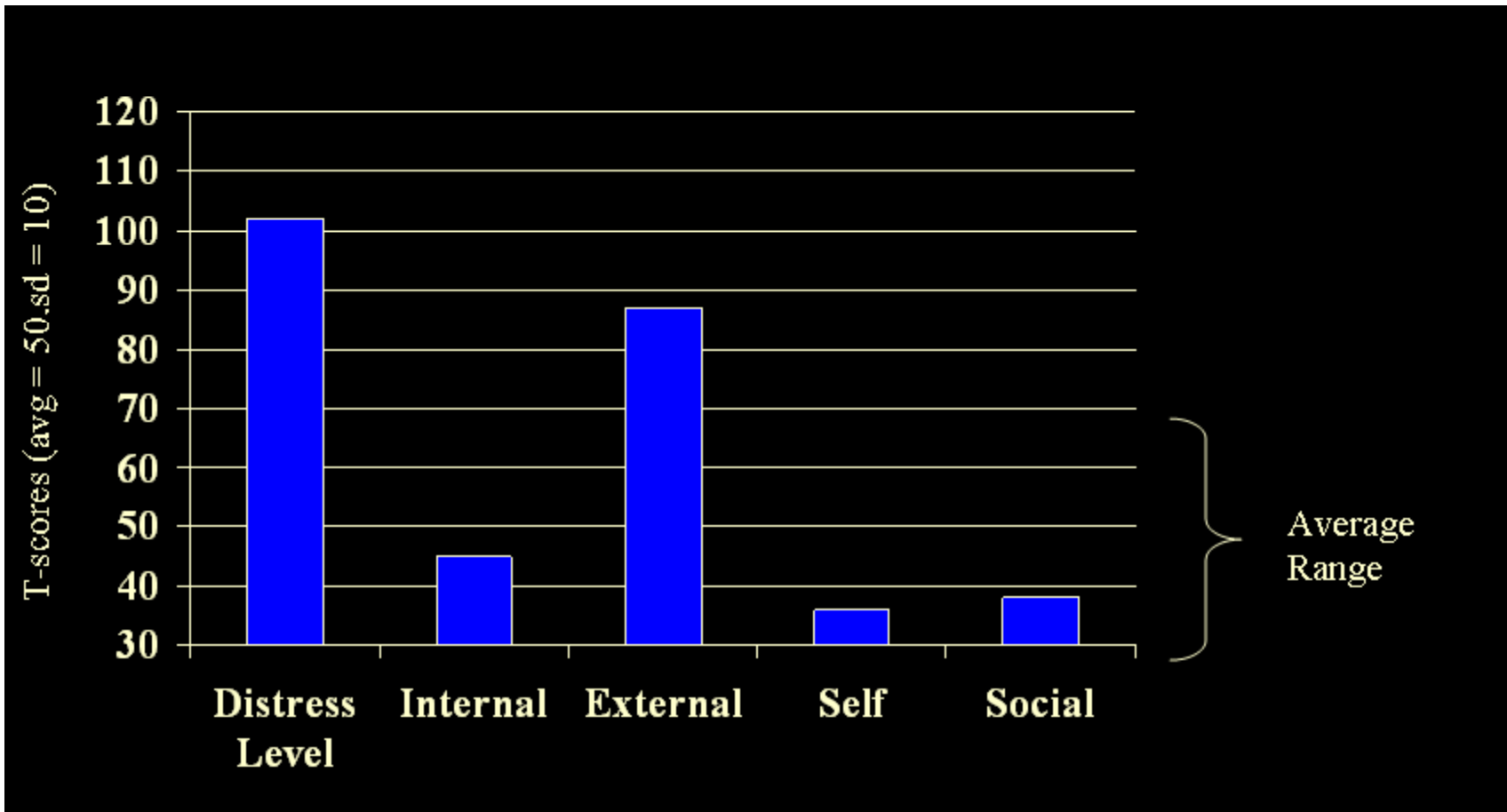


**Question: are
psychological
constructs
objective
measures?**

Examples of Test Design



Test Results



Evaluating the tests

Freedom from bias

Wording bias	Manipulative (leading) words or phrases <i>i.e. 'Describe your aspirations about farming...' (= assumes there are some)</i>
Item bias	An item that measures differently in different groups <i>i.e. 'self identity' conceived differently by Japanese and English origin subjects</i>
Ordering bias	When the order of items affects participants' responses <i>i.e. measure of 'dream job' directly before a measure of 'job satisfaction'</i>
History	Context specific phenomena that effects responses <i>i.e. prior training in psychology</i>

Testing the tests

Validity

Face validity	How well the test appears to be measuring what it intends to measure (i.e. a picture test measures spatial reasoning)
Predictive validity	How well results meet expectations e.g. people with depressive symptoms should score higher on neuroticism in the Big Five than people who are not depressed
Construct validity	How well the test measures each construct that it was designed to measure Cronbach's Alpha scores rate the strength of the scales
Content validity	How well the test measures <i>all aspects</i> of the main focus of measurement. e.g. affective <i>and</i> behavioural dimensions of <u>self-esteem</u> e.g. academic <i>and</i> social dimensions of <u>self-concept</u>

Testing the tests

Reliability

Test-retest reliability	The test is given at two different times and checked for how similar the responses are across time
Parallel forms reliability	Two different types of the same test are administered and their scores are correlated
Split half reliability	Test responses are halved. Random correlations are done between the two halves of the test then compared
Cronbach's alpha (0.70 + considered an accurate measure)	Compares the variance of scores on individual items with the covariance (variance between items) within a measure (e.g. a psychological construct). The variance of the item should be similar to the covariance between items.
Inter-rater reliability	Two or more coders analyse the same open-ended data. Their results are compared.

Question: What other ways are there to determine whether a test is reliable across time and population?

Ethics in psychological testing

- **Purpose**

Be ethical about why and how the test is given

- **Authenticity**

The test must reflect 'real life' psychology

- **Generalisability**

When reporting results, be realistic about who these can be extended to

- **Subjectivity**

Be honest about how much personal judgement is included in the design of your test and results analysis

Where to find tests

- The Faculty library
- The Psychometrics Centre
- Cantab (for cognition)
- Internet
- Lecturers and Professors



Acknowledgements

- *Dr Dénes Szucs*
- *Dr Ruth Kershner*
- *Dr Linda Hargreaves*
- *Jane Webb*
- *Dr Jason Rentfrow*
- *Prof. John Rust*

Critiquing a test

Purpose

What does the test measure overall?

Who is it for?

What assumptions are these purposes based on?

Design

What are the individual constructs that it measures?

What logical or theoretical assumptions underpin these constructs?

What is the empirical basis is there for the constructs, i.e. how were they developed, and what evidence is there of this?

Bias, validity and reliability

Examine the test for item, wording, ordering biases

How would you test the test for content, construct and predictive validity? Make a plan for one of these.

Overall – what are the main flaws and strengths of this test?