

Measurement Theory


Psychometric Theory

- ▶ Well developed theory and methodology for developing assessments
- ▶ Always applied to psychometric tests – but relevant to any assessment
- ▶ Complex statistical techniques – but we will focus on underlying concepts.

Scoring psychometric tests

- ▶ Map behaviour onto numbers
 - ▶ E.g number of correct responses
 - ▶ Summing ratings across items

Measurement Exercise
How tall am I?



Why were the measures not all the same?
▶ What were the sources of error?

Sources of inaccuracy

- ▶ Unstable characteristics
 - ▶ Changes over time
 - ▶ Changes moment to moment
- ▶ Poor measures
 - ▶ Badly written, ambiguous items
 - ▶ Measure too short
 - ▶ Inconsistent raters
- ▶ Disruption to measurement process
 - ▶ Respondent can't concentrate
 - ▶ Can't read questions
 - ▶ Respondents are all the same

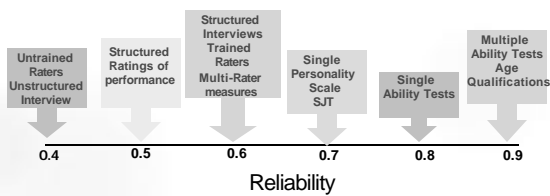
All measures have error

$$\text{Observed score} = \text{"True Score"} + \text{Error}$$

Reliability

- ▶ Reliability is the Accuracy or Precision of measurement
- ▶ True Score Variation / Observed Score Variation
- ▶ Expressed as a number between 0 and 1
- ▶ Values > 0.6 for a useful measure
 > 0.75 for selection purposes

Typical Reliability



Standard Error of Measurement

- ▶ Any measurement has an error band
- ▶ SEM shows error band for psychometric scores
- ▶ 95% of people have a 'true score' within 2 SEM of their observed score
- ▶ Scores within two sem of each other cannot be considered different

Validity

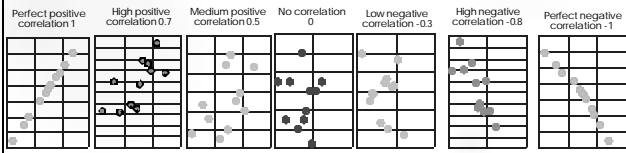
- ▶ Does the test measure what it says it does?
- ▶ Evidence
 - ▶ Predictions regarding test scores are confirmed e.g. scores are related to job performance
 - ▶ Scores relate to other measures in predictable ways e.g. different measures of ability correlate, measures of personality do not correlate with ability

Types of validity

- ▶ Faith Validity
- ▶ Face Validity
- ▶ Content Validity
- ▶ Criterion Related Validity
 - ▶ Concurrent
 - ▶ Predictive
- ▶ Construct
 - ▶ Correlations with other measures
 - ▶ Relationships between sub-scores
 - ▶ Other hypothesised relationships
- ▶ Consequential Validity

Correlation

- ▶ Indicates strength of relationship between 2 variables
- ▶ Varies from -1 to 1
- ▶ Don't forget to look for negative correlations

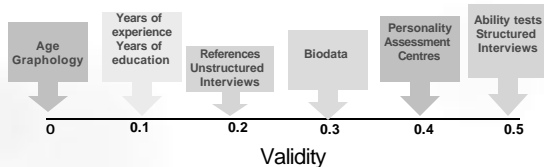


Construct validity example: OPQ32 and MBTI

Extraversion-Introversion	Outgoing (-0.65) Socially Confident (-0.43) Affiliative (-0.37) Emotionally Controlled (-0.36)
Sensing-Intuitive	Conventional (-0.47) Innovative (0.36) Detail Conscious (-0.36) Rule Following (-0.34)
Thinking-Feeling	Caring (0.44) Competitive (-0.39) Emotionally Controlled (-0.34) Evaluative (-0.32)
Judging-Perceiving	Detail Conscious (-0.52) Conscientious (-0.43) Rule Following (-0.40) Forward Thinking (-0.34)

N=141

Criterion Related Validity



Robertson and Smith, 2001

Personality validity

- ▶ Depends on role
 - ▶ What behaviours are required for success
- ▶ Depends on criterion
 - ▶ Overall performance
 - ▶ Rating
 - ▶ Hard metrics e.g. sales figures
 - ▶ Aspect of performance
- ▶ Depends on culture
 - ▶ What type of behaviour is valued
- ▶ Depends on situation
 - ▶ What is scope to influence outcomes with behaviour

Predictors of Leadership

	Business	Government	Military
Extraversion	✓	✓	✓
Emotional Stability	✓	✓	✓
Openness	✓	x	x
Conscientious	x	✓	✓

Judge et al
2002

What is overall job performance?

- ▶ Quality of work
- ▶ Quantity of work
- ▶ Sales performance
- ▶ Absence of errors
- ▶ Attendance
- ▶ Liked by customers
- ▶ Job knowledge

What
personality
traits would
predict
each of
these?

Meta-analytic validity coefficients

Personality Trait	With relevant job facet	With other job facets	With Overall Performance
Agreeableness	0.09	-0.02	-0.05
Extraversion	0.12	0.03	0.02
Openness	0.09	0.01	0.01
Conscientiousness	0.12	0.00	0.05
Emotional Stability	0.08	0.00	-0.02

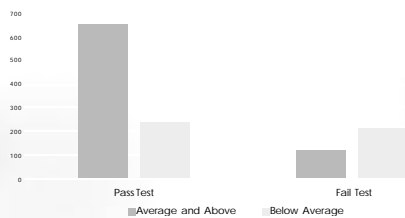
n = 2087 - 2193, k = 10, uncorrected correlations, Bartram et al.

Predicting with multiple narrower traits

Job Performance Facet	Multiple R	Number of traits
Analysis and Problem Solving	0.48	4
Communicating	0.35	1
Interpersonal	0.47	4
Leadership	0.43	3
Drive	0.38	4
Strategic	0.43	3
Marketing	0.40	2

n = 2087 - 2193, k = 10, uncorrected correlations, Bartram et al.

Impact of validity of 0.59



For every 100 people selected using test
10 fewer below average performers

Impact of validity depends on

- ▶ Validity coefficient
 - ▶ Higher coefficients have greater impact
- ▶ Base rate for success
 - ▶ More impact when base rate lower
- ▶ Selection Ratio
 - ▶ More impact when smaller proportion of applicants chosen
- ▶ Cut score
 - ▶ More impact with higher cut score
