Latent variables

- The Theory of True Scores
  - Sometimes called or ‘Latent Trait Theory’
- \( X = T + E \)
  - Where \( X \) = Observed score
  - \( T \) = True Score
  - \( E \) = Error
- Latent Variable Analysis (Factor Analysis)
The Psychometric Principles
Maximizing the quality of assessment

- Reliability (freedom from error)
- Validity ('... what is says on the tin')
- Standardisation (compared with what?)
- Equivalence (is it biased?)

Constructing a psychometric test

- Defining the purpose
- Designing the blueprint
- The pilot study
- Item analysis
- Obtaining reliability and validity
- Writing the handbook
Appropriate latent traits

- Personality rather than ability – e.g. “Tendency to feel ....”
- Single traits with about 32 pilot items
- Reduce to 20 through item analysis
- Examples: Apathy, Bewilderment, Collywobbles, Disgruntlement, Embarrassment, feeling like a Fraud, Guilt, Hatred, Humility, Indignation, Jealousy, Loneliness, Love, Morbid Curiosity, Nostalgia, Overwhelmed, Perversity, Road Rage, Self-Pity, Triumph, uncertainty, Vengefulness, Wanderlust
Designing the blueprint

- Curriculum based
  - Bloom’s taxonomy of educational objectives

- Job description
  - The job analysis
  - The person specification

- Theoretical
  - Ability
  - Personality
# Knowledge Test Specification

## Manifestations

<table>
<thead>
<tr>
<th>Manifestations</th>
<th>Arithmetic</th>
<th>Geometry</th>
<th>Algebra</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of Terms (25%)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Understanding (25%)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Application (25%)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Generalisation (25%)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Personality Test Specification

<table>
<thead>
<tr>
<th>Manifestations</th>
<th>Extraversion</th>
<th>Neuroticism</th>
<th>Detail</th>
<th>Tough-mindedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High/Positive</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>High/Negative</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Low/Positive</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Low/Negative</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
The pilot study

- Pre-piloting
  - Are the correct items correct?
  - Are the distractors incorrect?
  - Are any items offensive or likely to be biased?
- The sample and sample size
- Data collection
- Data entry
Psychometric Item Characteristics

Classical
- Difficulty (or Facility) Value (or Index)
- Discrimination Index
- Correction for Guessing

Item Response Theory
- Difficulty (or Rasch) Parameter (1st)
- Discrimination Parameter (2nd)
- Guessing Parameter (3rd)
- Inattention Parameter (4th)
Using difficulty (p) and discrimination (r) indices

- p should be between 20% and 80%
- Adjusted Item-Total Correlation (point or rank biserial) > 0.2
- Remember the test specification!
Classical Item Reduction

- Record form analysis
  - Non-responses
  - Altered items
  - Comments
- Delete extreme items
- Delete items with poor discrimination
- Retain the balance of the test
  - test specification
  - Positive and negative items
- Aim to reduce items by 50%
Inter-rater reliability
Test–retest reliability (stability)
Parallel forms reliability
Split-half reliability
  - The Spearman-Brown Correction
Cronbach’s alpha
Expected levels of reliability

- WAIS: 0.92
- Ability tests: 0.82
- Personality test: 0.75
- Essay marks: 0.60
- Supervisors ratings: 0.40
- Projective tests: 0.30
- Creativity tests: 0.25
- Astrology/graphology: ?
p should be between 20% and 80%

Adjusted Item-Total Correlation (point or rank biserial) > 0.2

Remember the test specification!
Software

- R
- Excel (with Real Statistic or other Add On)
- Mplus
- SPSS
  - Analysis
    - Scale
      - Reliability analysis
      - Statistics
      - Item
      - Scale if item deleted
Validity

- Face validity
- Content validity
- Predictive validity
- Construct validity
- Differential validity
- Consequential validity
Calculate means and standard deviation of norm group

Provide norm table or conversion

- Standard scores \( z = \frac{x - \text{mean}}{\text{s.d.}} \)
- Standardised scores
  - T-scores = \( z \times 10 + 50 \)
  - Stanine = \( z \times 2 + 5 \) (min = 1, max = 9)
  - Sten = \( z \times 2 + 5.5 \) (min = 1, max = 10)
  - IQ format = \( z \times 15 + 100 \)
Freedom from bias

- Item bias
- Intrinsic test bias
- Extrinsic test bias
- Adverse impact
- Equivalence
- Differential Item Functioning
Writing the handbook

- Include copyright notice
- Include the scoring key and instructions
- Give evidence of reliability and validity
- Provide norms