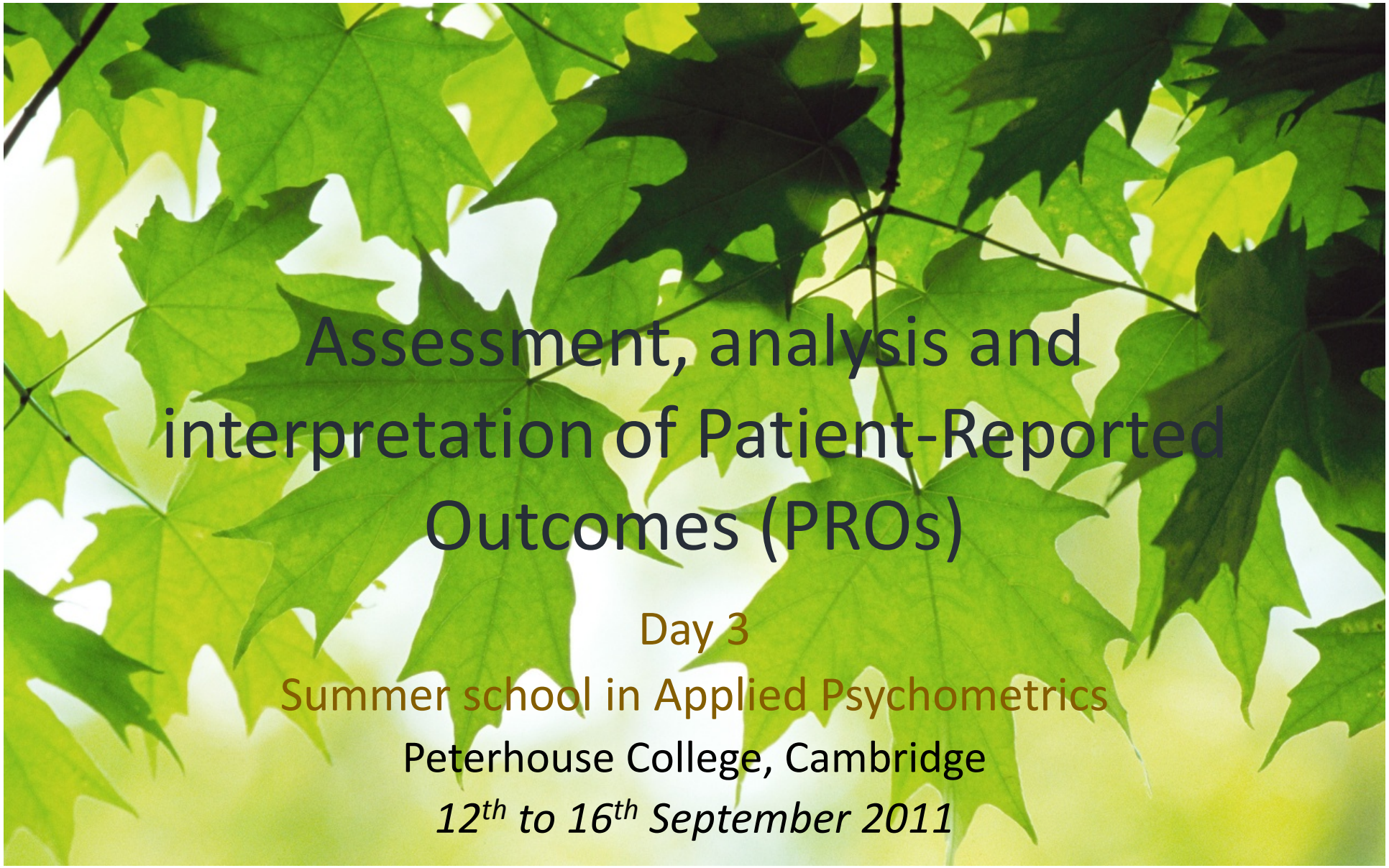


Assessment, analysis and interpretation of Patient-Reported Outcomes (PROs)

Day 3

Summer school in Applied Psychometrics

Peterhouse College, Cambridge

12th to 16th September 2011

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

This course is prepared by

Anna Brown, PhD ab936@medschl.cam.ac.uk
Jan Stochl, PhD js883@cam.ac.uk
Tim Croudace, PhD tjc39@cam.ac.uk
(University of Cambridge, department of Psychiatry)
Jan Boehnke, PhD boehnke@uni-trier.de
(University of Trier, Department of Clinical Psychology and Psychotherapy)

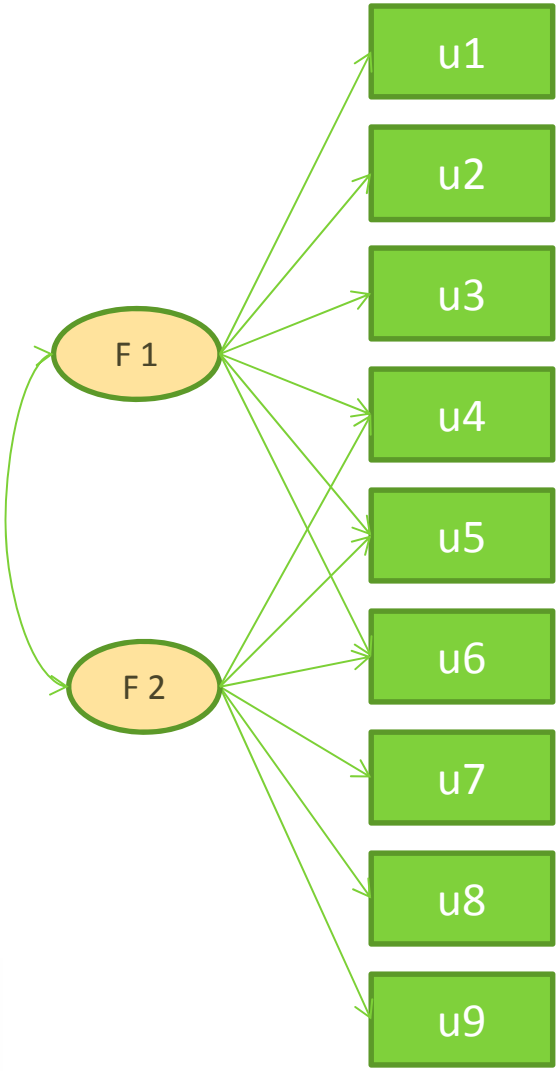
The course is funded by the ESRC RDI and hosted by
The Psychometrics Centre

Anna Brown

9. INTRODUCTION TO MULTIDIMENSIONAL IRT MODELS

Multidimensional IRT

- Sometimes items are indicators of more than one constructs
- Not merely a fancy extension of unidimensional theory
- More commonly required than you might think
  - Often things can be done unidimensionally, but really better off with MIRT
  - Sometimes things cannot be done without it



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

The item responses

Notation: u_{ij} – response of examinee *j* to item *i*

- Test items most often assume categorical response
  - Questionnaires can produce binary responses (yes-no, agree – disagree)
  - Or ordered categorical (ordinal) responses
    - Might have 3, 4, 5, 7 or even 9 rating categories
    - Rating scales can be symmetrical (agree-disagree) and not (never-always)
  - Sometimes choice alternatives are purely nominal categories

The latent trait(s)

Notation: "theta" $\theta \in (-\infty, +\infty)$

- The latent traits are simply labels used to describe what the set of questionnaire items measures
  - Validation studies are required to determine what a questionnaire measures.
- Latent traits can be broadly or narrowly defined variables related to QoL
  - No reason to think of trait as fixed over time.

Dominance response process

- In many QoL models it is hypothesised that probability of a keyed response increases monotonically as the latent trait increases
- Good model for most tests
- Underlying multidimensional factor model

$$u_i^* = \mu_i + \lambda_{1i}\theta_1 + \lambda_{2i}\theta_2 + \ldots + \lambda_{Ti}\theta_T + \varepsilon_i$$

  - u* is the underlying (unobserved) response tendency
- Assumptions as per factor analytic model; standardised parameters; var(u*)=1

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Threshold process

- The observed response u_i relates to the unobserved response tendency through a threshold process.
- For instance, there is one threshold when two response alternatives are used (0 for non-keyed response and 1 for keyed response):

$$u_i = \begin{cases} 1, & \text{if} \quad u_i^* \geq \tau_i \\ 0, & \text{if} \quad u_i^* < \tau_i \end{cases}$$

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Normal-ogive model

$$\mathrm{P}\left(u_i\right) = \mathrm{Pr}\left(u_i = 1 \middle| \boldsymbol{\theta}\right) = \Phi\left(\frac{-\tau_i + \lambda_{1i}\theta_1 + \ldots + \lambda_{Ti}\theta_T}{\sqrt{\psi_i^2}}\right)$$

- Familiar cumulative normal distribution function (can be looked up in tables)
- Maths is horrible so models with logistic links eventually became more popular (though their IRFs are virtually indistinguishable from normal ogive)

$$\mathrm{L}\left(x\right) = 1 / \left(1 + e^{-x}\right)$$

- Because u* is unobserved (only its dichotomisation is observed), unique item variance cannot be identified

Intercept/slope parameterization

- It is customary to parameterize IRT models (e.g. McDonald, 1999) by letting

$$\alpha_i = \frac{-\tau_i}{\sqrt{\psi_i^2}} \qquad \beta_{ij} = \frac{\lambda_{ij}}{\sqrt{\psi_i^2}}$$

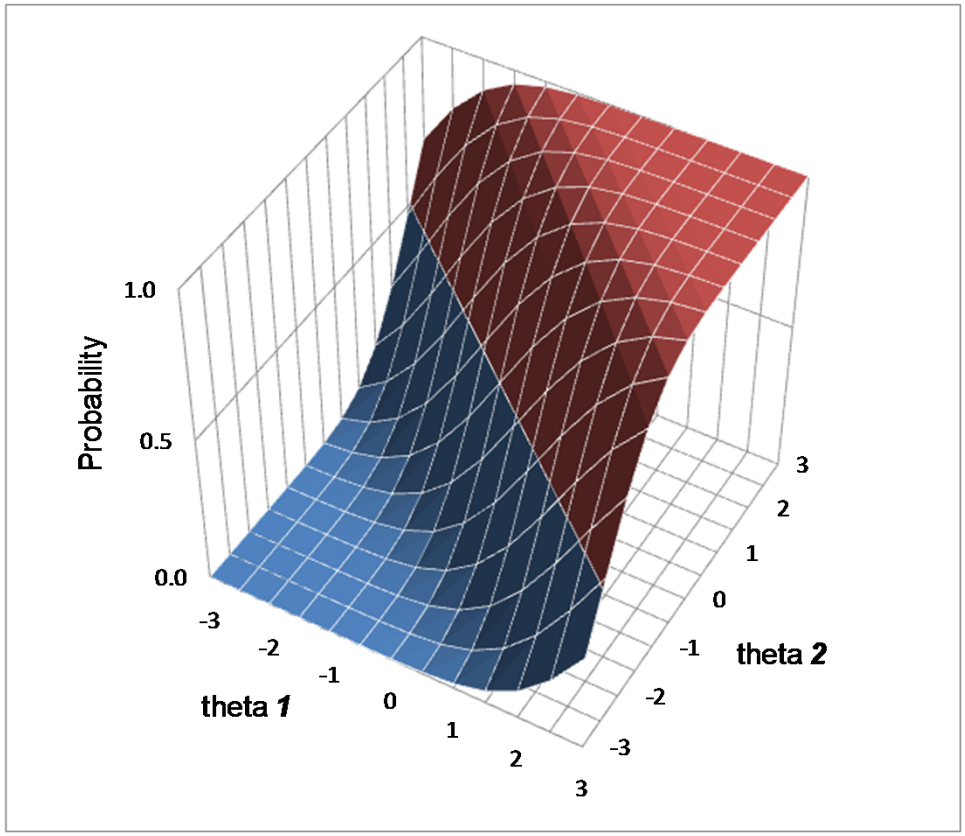
- Now, the item residual variance is assumed equal 1 and the item response function can be written in an intercept / slope form as

$$P_i = \mathrm{P}\left(u_i \middle| \boldsymbol{\theta}\right) = \Phi\left(\alpha_i + \beta_{i1}\theta_1 + \ldots + \beta_{iT}\theta_T\right)$$

  - Discrimination/difficulty parameterization comes from ability testing IRT tradition and only works for one-dimensional models

$$P_i = \mathrm{P}\left(u_i \middle| \theta\right) = \Phi\left(a_i\left(\theta - b_i\right)\right)$$

Two-dimensional case



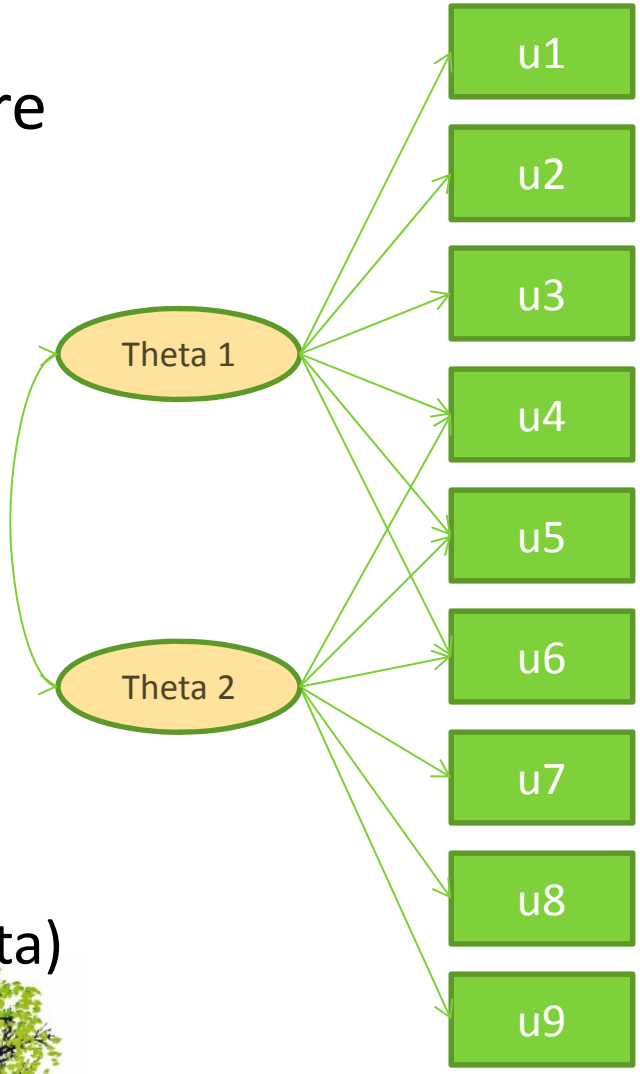
Item parameters:

Beta 1 = 0.86
Beta 2 = 1
Alpha = 0

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Fitting a multidimensional IRT model

- Items could be indicators of 1 or more constructs
- CFA with categorical variables
- Local independence is assumed
- FIML or limited info estimators are available
  - FIML is computationally heavy with several dimensions
  - FIML produces loglikelihood
  - Limited info estimators produce chi-square (problematic with most IRT data)



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Illustration – NSHD Dataset

- Responses from the ongoing Medical Research Council National Survey of Health and Development (NSHD), also known as the British 1946 birth cohort.
- Wave of interviewing undertaken in 1999 when the participants were aged 53 (Wadsworth et al, 2003).
- A total of N=2091 respondents (1422 men and 1479 women) provided answers to the GHQ-28.

GHQ Instrument

- The 28-item version of Goldberg's General Health Questionnaire (GHQ-28; Goldberg, 1972)
  - Developed as a screening questionnaire for detecting non-psychotic psychiatric disorders in community settings and non-psychiatric clinical settings
- Respondents are asked to think about their health in general and any medical complaints they have had over *the past few weeks.*
- Rating scale with 4 alternatives
  - slightly different for each item, in phrasing and verbal anchors
- Example question

"Have you recently lost much sleep over worry?"

(Not at all - No more than usual - Rather more than usual - Much more than usual)

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

GHQ - a priori structure

- Designed to measure 4 *a priory* facets of mental health variation (measured with 7 items each)
  1. Somatic Symptoms,
  2. Social Dysfunction,
  3. Anxiety / Insomnia,
  4. Severe Depression / Hopelessness.
- Also, the general *psychological distress* factor can be measured

EFA of NSHD Dataset

- Collapsed the top two response categories, effectively resulting in the item coding **0-1-2-2**
- Exploratory analysis of polychoric correlations
  - five eigenvalues greater than one (12.9, 2.6, 2.2, 1.5 and 1.1)
  - Four correlated factors: χ^2 = 3917 (*df* = 272); RMSEA = 0.068, CFA = 0.959
- Target rotation to a hypothesized solution with 4 correlated constructs yields well-behaved solution
  - Correlations range from 0.33 to 0.56
- Ratio of the first to the second eigenvalue also suggests presence of a general factor (psychological distress)

Facet 1 – somatic symptoms

Item	Wording	Factor 1 Somatic Symptoms	Factor 2 Anxiety	Factor 3 Social Dysfunction	Factor 4 Depression
S1	good health	0.53		0.39	
S2	need good tonic	0.62	0.36		
S3	run down	0.72	0.32		
S4	felt ill	0.69			
S5	pains in head	0.97			
S6	pressure in head	0.97			
S7	hot and cold spells	0.39			

Facet 2 – anxiety/insomnia

Item	Wording	Factor 1 Somatic Symptoms	Factor 2 Anxiety	Factor 3 Social Dysfunction	Factor 4 Depression
A1	lost sleep over worry		0.79		
A2	difficulty staying asleep		0.67		
A3	constant strain		0.73		
A4	edgy		0.58		
A5	panicky/scared		0.49		0.33
A6	everything on top of me		0.69		
A7	nervous		0.59		0.36

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Facet 3 – social dysfunction

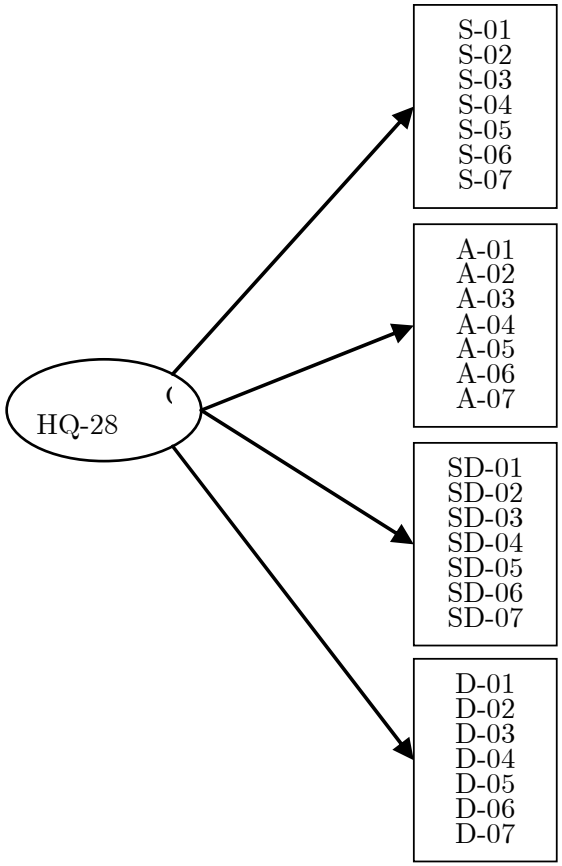
Item	Wording	Factor 1 Somatic Symptoms	Factor 2 Anxiety	Factor 3 Social Dysfunction	Factor 4 Depression
SD1	manage to keep busy			0.52	
SD2	take longer time			0.55	
SD3	doing things well			0.95	
SD4	satisfied with tasks			0.86	
SD5	play a useful part			0.76	
SD6	making decisions			0.61	
SD7	enjoy daily activities		0.35	0.41	

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Facet 4 – (severe) depression

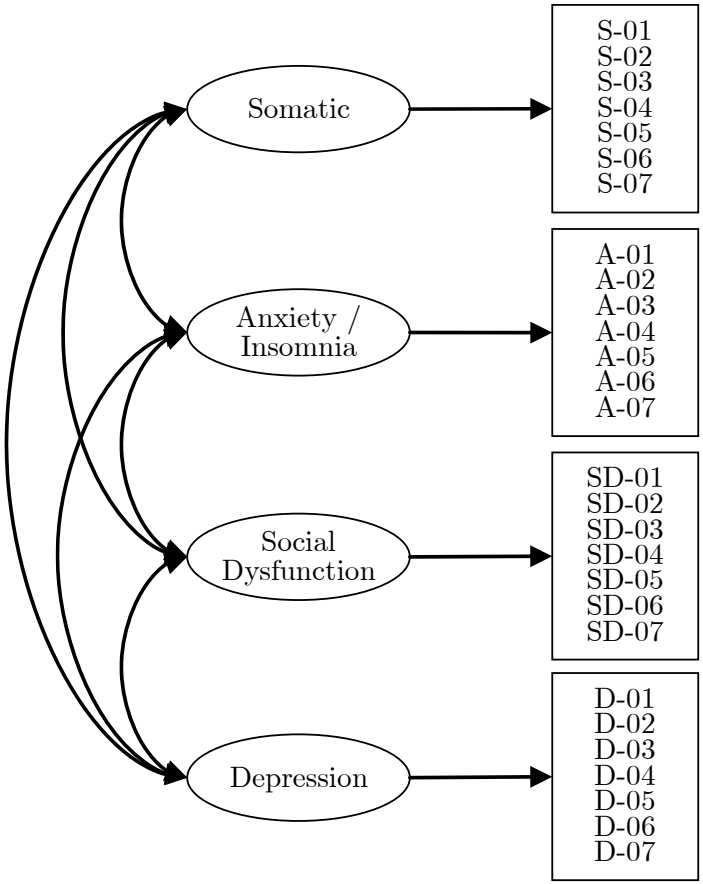
Item	Wording	Factor 1 Somatic Symptoms	Factor 2 Anxiety	Factor 3 Social Dysfunction	Factor 4 Depression
D1	worthless				0.61
D2	hopeless				0.72
D3	not worth living				0.79
D4	thoughts of suicide				0.95
D5	nerves too bad				0.53
D6	wishing dead				0.86
D7	suicidal ideas				0.92

1. Unidimensional model



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

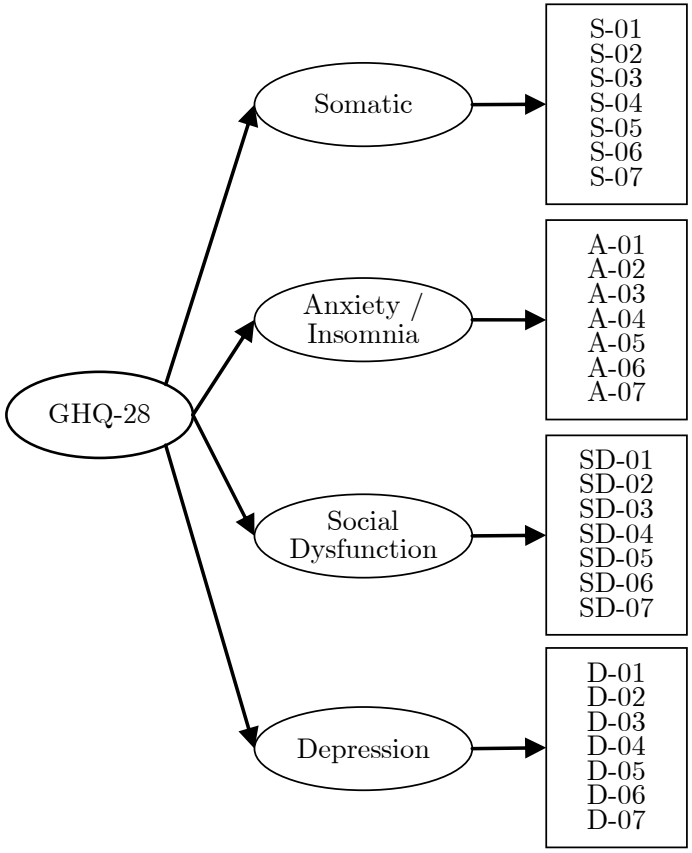
2. Correlated traits model



Unidimensional or multidimensional IRT model?

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

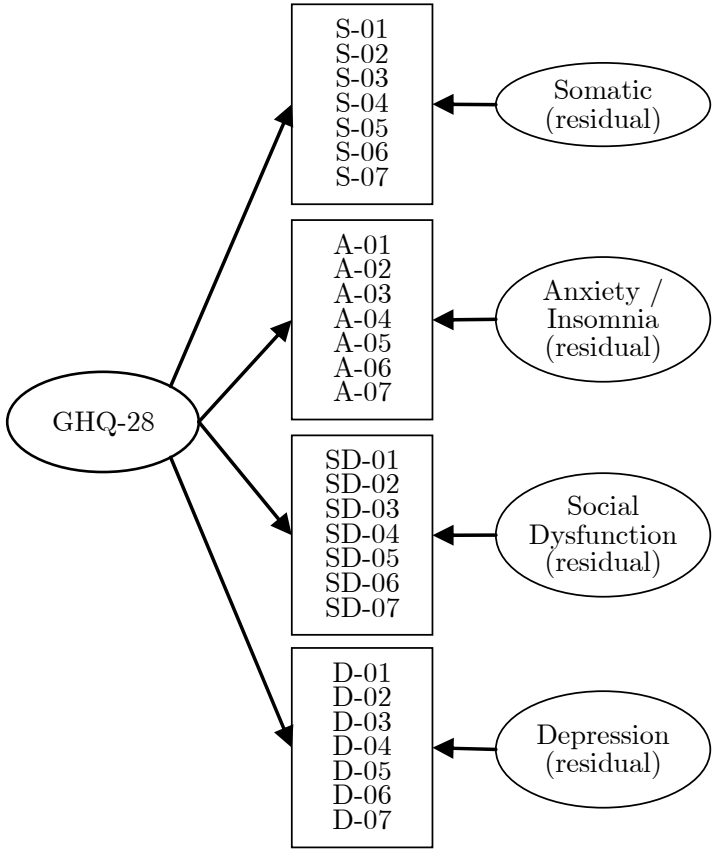
3. Hierarchical model



Unidimensional or multidimensional IRT model?

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

4. Bifactor model



Unidimensional or multidimensional IRT model?

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Practical

1. We are going to test the GHQ alternative models 1-4 for *simulated* data using Mplus
   - Specify and estimate the four alternative models
   - Record model fit (chi-square, RMSEA, CFI, TLI)
   - Examine residuals – what do they tell us about the different models?
   - Which model is the most appropriate for the data and what substantive explanation for its appropriateness would you give?

GHQ models - Goodness of fit

Model	Chi-square	df	RMSEA	CFI	TLI
Unidimensional	15,314	350	.121	.831	.818
Hierarchical	5,920	346	.075	.937	.931
Correlated traits	5,870	344	.074	.938	.932
Bifactor	4,142	323	.064	.957	.950

Models were estimated using the limited information DWLS estimator (based on tetrachoric correlations) in M*plus* software (version 6.11).

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Why use MIRT?

- More often than not multiple factors underlie responses to questionnaire items (and often by design)
- If this multi-factor influence on items is ignored, critical assumptions of IRT models will be violated
  - Local independence will not hold
- Results will be compromised
  - Item parameters
  - Accuracy of estimation of person's scores
  - Assessment of that accuracy (reliability overestimated)