



UNIVERSITY OF
CAMBRIDGE

The Psychometrics Centre

SSRMC Psychometrics Module

Classical Test Theory and Practice

Professor John Rust

<http://www.psychometrics.cam.ac.uk>

The Psychometrics Centre
Judge Business School
Trumpington Street

The Psychometrics Centre

- Educational and diagnostic tests e.g. BAS-III, WISC, CELF
- Organisational e.g. Watson-Glaser, Orpheus/OBPI, Giotto
- Statistical, IRT, machine learning and AI techniques
- Software products e.g. Concerto, Apply Magic Sauce
- Web based assessment, 'Discover My Profile'
- BPS Professional training courses
- Short courses, seminars and events (SEM in R, Mplus, ML)
- PhDs in psychometrics
- Tutorial materials on website - www.psychometrics.cam.ac.uk

The programme

- 1: Introduction to psychometrics (John Rust)
 - Psychometrics yesterday, today and tomorrow
 - How to design and build your own psychometric test
- 2: Testing in the online environment (Dr David Stillwell)
 - Testing via the internet. How to, plus do's and don'ts
 - Putting your test online
- 3. Modern Psychometrics (Dr Luning Sun)
 - How to build your own CAT using Concerto
 - Practical
- 4: Online Computer Adaptive Tests (Dr Aiden Loe)
 - Item Response Theory (IRT) models and their assumptions
 - Assessment and development of item parameters

Introduction to psychometrics

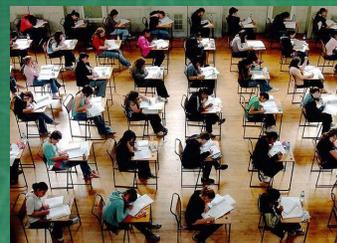
- Definition: The science of psychological assessment
- From IQ to AI: Past, present and future
- Intended and unintended consequences
- Behavioural Economics
- Online digital footprints
- 'Psychology' in Cyberspace

Psychometric Prediction

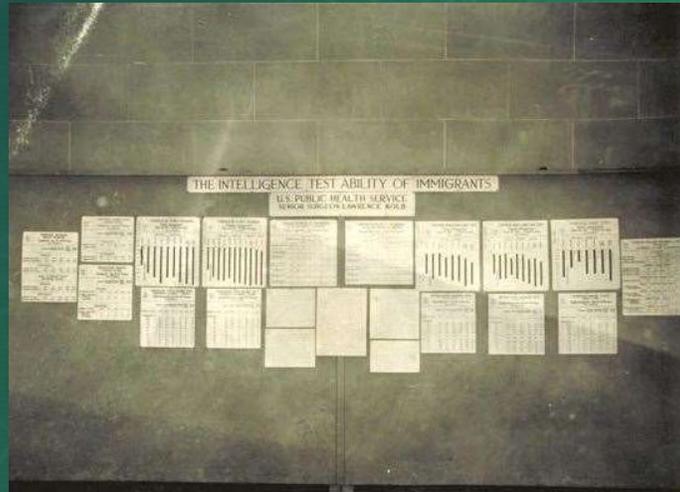


The 20th Century. Testing IQ with the best of intentions

MERITOCRACY



And sometimes the worst. Ellis Island



Eugenics and Dysgenics

- Virginia (1924) "Sterilization Act" for the "feeble-minded"
- Germany (1933) "Law for the Prevention of Genetically Diseased Offspring" (sterilization for feeble-mindedness, mental illness, blindness, deafness, physical deformity)
- Germany (1937) 'Commission Number 3' compulsory sterilization of children of mixed-race origins.
- Germany (1939) Euthanasia introduced in psychiatric hospitals for those with disabilities (broadly defined, this included homosexuality and "social deviancy")
- Soviet Union (1949), Incarceration on psychiatric diagnosis of "philosophical intoxication", "sluggish schizophrenia" (poor social adaptation), etc.

6. [REDACTED] - (W) - Resident Buncombe County
 Proceedings instituted by George H. Lawrence, SPW
 Medical history by H. C. Whems, M. D.
 Sterilization recommended by Dr. Whems

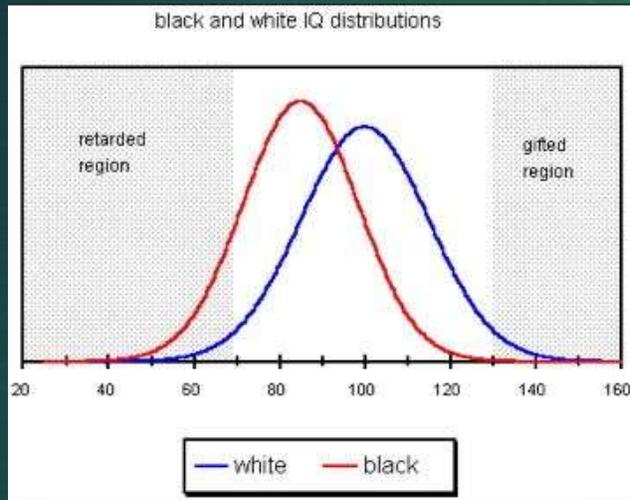
Social history:
 Married woman, [REDACTED] years of age, who is the mother of 3 children ranging in ages from [REDACTED] to [REDACTED] years. On a psychological examination given by Dorothy Hamilton she was found to have an I. Q. of 56. The family has been financially dependent for many years and there is a history of inter-marriage with Indian and Negro.

Diagnosis: Feeble-mindedness

Operation will be performed by surgeon on staff of Biltmore Hospital at time of delivery.

Consents: Consent is signed by [REDACTED], patient, and by [REDACTED], husband.

Herrnstein and Murray "The Bell Curve" (USA, 1994)



The Flynn Effect

Wechsler Intelligence Scale for Children		Average at 1949	Rate of change p.a.
WISC	1949	100.00	
WISC-R	1974	107.63	0.311
WISC-III	1991	113.00	0.322
WISC-IV	2003	117.63	0.363



Openness



Openness-to-Experience

I am always on the lookout for new ideas to explore

Conventionality

I stick with tried and tested ways of doing things

Conventionalism



Conscientious



Detail conscious

If you get the details right everything will come out fine.

Seeing the bigger picture

I prefer to leave the boring details to my staff

Disorganised



Extraversion



Extraversion

I can be the life and soul of any party

Introversion

I have my best ideas when I am left alone

Introversion



Agreeableness



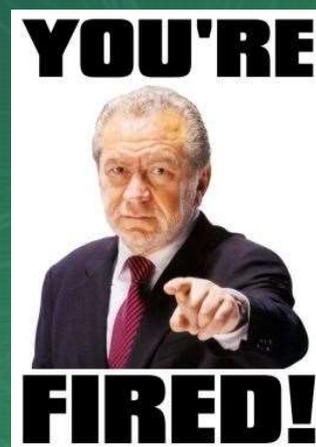
Tendermindedness

I will always try and help my friends when they need me

Disagreeable

I have a special ability to make tough decisions when needed

Toughmindedness



Neurotic



Emotional

I get anxious when I have to
make difficult decisions

Unemotional

I am more able than most to
cope with disaster

Stable



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?)

• *Rust, J. & Golombok, S. (2009) Modern Psychometrics (3rd Ed): Routledge: London*

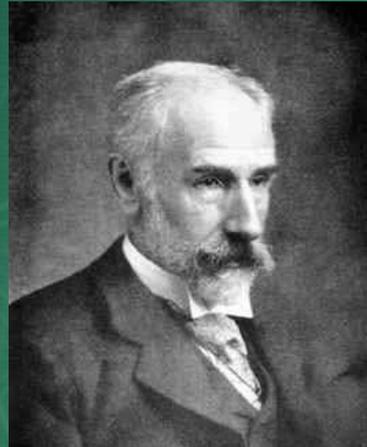
Reliability

Theory of True Scores

If we tabulate the marks given by the different examiners they will tend to be disposed after the fashion of a gendarme's hat.

I think it is intelligible to speak of the mean judgment of competent critics as the true judgment; and deviations from that mean as errors. This central figure which is, or may be supposed to be, assigned by the greatest number of equally competent judges, is to be regarded as the true value, just as the true weight of a body is determined by taking the mean of several discrepant measurements.

Francis Edgeworth, 1888



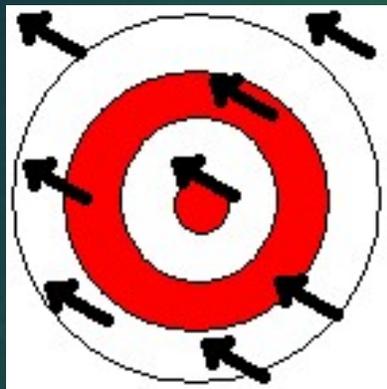
The Theory of True Scores

- Sometimes called or 'Latent Trait Theory'
- $X = T + E$
 - Where X = Observed score
 - T = True Score
 - E = Error
- Latent Variable Analysis

Measuring reliability

- Reliability is reported as a positive correlation coefficient
- The reliability of a score is a value between 0 and 1.
 - If zero, all is error, one is perfect accuracy.
- Can use it to:
 - Report the expected accuracy of our question or questionnaire
 - Improve the accuracy of our measure
 - Compare the accuracy of different forms of assessment
 - Assign a degree of confidence to a test result.

Reliability



Reliability

- Inter-rater reliability
- Test –retest reliability (stability)
- Parallel forms reliability
- Split-half reliability
 - The Spearman-Brown Formula
- Cronbach's alpha

Expected reliabilities

- | | |
|----------------------------|------|
| • Individual ability tests | 0.92 |
| • Group ability tests | 0.85 |
| • Personality scales | 0.75 |
| • Essays | 0.66 |
| • Creativity tests | 0.50 |
| • Projective tests | 0.30 |
| • Graphology/Astrology | ? |

Reliability and Validity

- Reliability is the extent to which a measurement is free from error
- Validity is the extent to which a measurement is measuring what it is purported to measure

Validity



Validity

- Face validity
- Content validity
- Concurrent validity
- Predictive validity
- Construct validity

Standardization

- Calculate means and standard deviation of norm group
- Provide norm table or conversion
 - Standard scores $z = (x - \text{mean})/s.d.$
 - Standardised scores
 - T-scores = $z*10 + 50$
 - Stanine = $z*2 + 5$ (min = 1, max = 9)
 - Sten = $z*2 + 5.5$ (min = 1, max = 10)
 - IQ format = $Z*15 + 100$

Equivalence

- Item bias
- Intrinsic test bias
- Extrinsic test bias
- Adverse impact
- Equivalence
- Differential Item Functioning

Constructing a psychometric test

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

Developing the blueprint

- Curriculum based
 - Bloom's taxonomy of educational objectives
- Job description
 - The job analysis
 - The person specification
- Theoretical
 - Ability
 - Personality

Knowledge Test Specification

		Content areas			
		Arithmetic	Geometry	Algebra	Statistics
Manifestations	Knowledge of Terms (25%)	4	4	4	4
	Understanding (25%)	4	4	4	4
	Application (25%)	4	4	4	4
	Generalisation (25%)	4	4	4	4

Personality Test Specification

		Content areas			
		Extraversion	Neuroticism	Detail	Tough-mindedness
Manifestations	High/ Positive	4	4	4	4
	High/ Negative	4	4	4	4
	Low/ Positive	4	4	4	4
	Low / Negative	4	4	4	4

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%

Software

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

Item analysis

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

TD12 Subscale 6

1. When on public transportation and in public places, I often start a conversation with people I do not know.
13. In a group or a meeting, I love it when everyone notices me.
25. I am an attractive person.
37. I am often told that I am charming.
49. I am often the centre of attention at business meetings.
61. I like to draw attention to the way I am dressed and my style.
73. I am not afraid to be the centre of attention.
85. I enjoy having a lot of people around me.
97. I seek out and love strong emotions.
109. I express my feelings and emotions very easily.

SPSS Reliability: Item-Total Statistics (Cronbach = 0.767)

Scale Mean if Item Deleted

Scale Variance if Item Deleted

Corrected Item-Total Correlation

Squared Multiple Correlation

Cronbach's Alpha if Item Deleted

1.	21.63	15.946	.264	.108	.770
13.	21.47	14.668	.511	.359	.737
25.	21.06	15.536	.429	.247	.749
37.	21.17	14.986	.467	.262	.743
49.	21.74	15.052	.515	.343	.738
61.	21.76	14.813	.431	.243	.748
73.	21.19	14.152	.561	.394	.729
85.	21.13	15.072	.445	.223	.746
97.	21.35	15.226	.400	.217	.752
109.	21.25	15.519	.320	.160	.764

Writing the handbook

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

UNIVERSITY OF CAMBRIDGE Study at Cambridge About the University Research at Cambridge Quick links Search

The Psychometrics Centre Cambridge Judge Business School Log in

Home About Us Training Research Students Projects Concerto Collaboration Services Contact Us News

Follow @CamPsych 2,858 followers
Like You and 54 others like this.
G+

Upcoming events

Certificates of Occupational Test Use, Level A and B (3 days)
Dec 09, 2017
Cambridge, UK

Mplus course in Structural Equation Modelling (3 days)
Dec 06, 2017
Cambridge, UK

IRT and CAT using Concerto (3 days)
Jan 20, 2018
Cambridge, UK

SEM in R workshop (3 or 4 days)
Jan 26, 2018
Cambridge, UK

Upcoming events >

CONCERTO Develop your own online adaptive test using our open-source platform
Find out more >

Apply Magic Sauce Translates individuals' digital footprints into detailed psychological profiles
Find out more >

Workshops and training courses