Item Response Theory and Rasch Models

I. Review of Basic Concepts

II. Item characteristic curves

III. One-Parameter Logistic Models

IV. Two-Parameter Logistic Models

V. Other IRT Models

VI. Differential item functioning

I. Basic Concepts

Item response theory (IRT; Lord, 1953; 1980; Rasch, 1960) is a widely used approach to psychometric analysis

Primarily used for ability or knowledge tests with binary items (correct/incorrect), but can be used with ordinal responses and in other contexts

Used for item analysis, scale development/evaluation, and investigation of test bias

Specialized software, such as BILOG-MG, MULTILOG, PARSCALE, IRTPRO, and ltm in R

I. Basic Concepts

IRT is usually described as an alternative to Classical Test Theory (CTT)

Can be thought of a refinement and one focused on ability or knowledge tests with discrete (usually binary) items

Sometimes referred to as “latent trait theory,” the IRT approach quantifies the relationship between the ability and the response to an item (“latent trait”, “ability”, and “proficiency” are all used interchangeably")

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage.

I. Basic Concepts

CTT can be said to be "group dependent" because difficulty is assumed to be the same for all, whereas IRT focuses on difficulty as a function of the overall ability

Overall difficulty of test also affects the item difficulty

Error variance may differ for different levels of ability, which CTT does not account for

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage.

II. Item characteristic curves

Remember *item difficulty* is the proportion of respondents who get an item correct (so really, easiness of the item)

II. Item characteristic curves

In IRT, the idea of discrimination is used in a more continuum sense

Item characteristic curves (ICCs; or *item response function*) are graphic representations of the probability the item is correct for values of the full score (or, in a personality context, the trait)

The ICC is a plot of the probability that an item is correct by the latent trait, represented in equations by the Greek letter theta, θ

II. Item characteristic curves

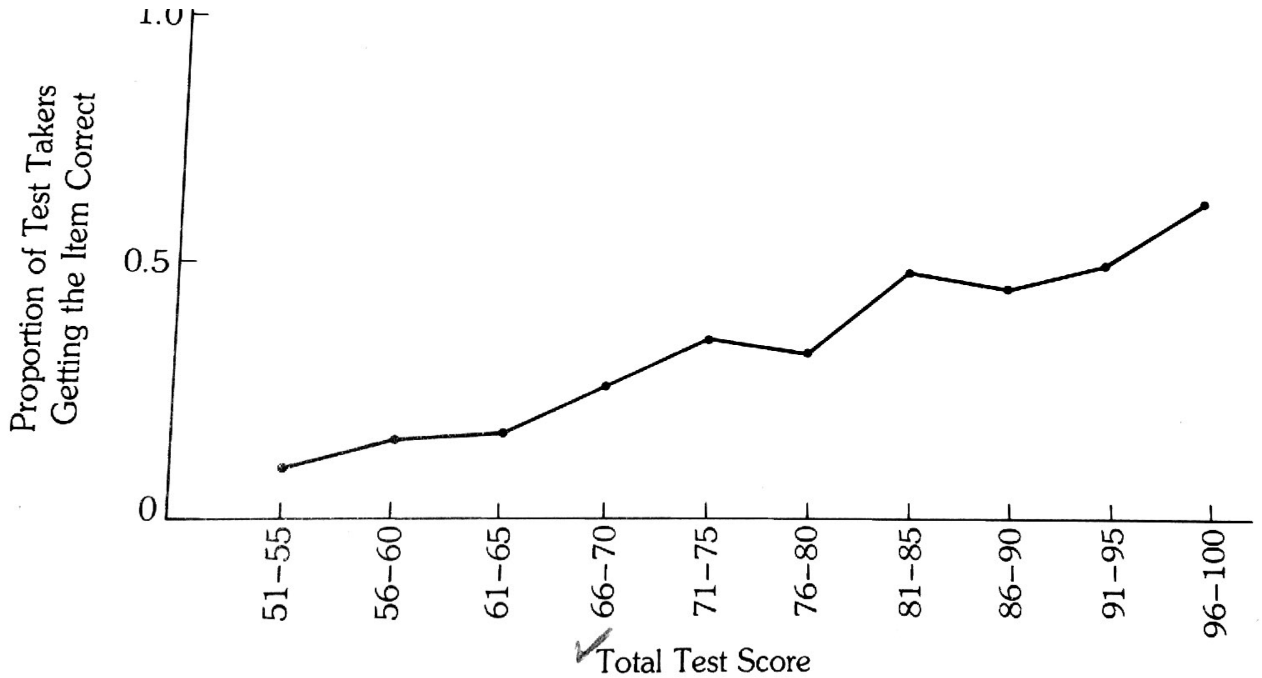


Figure 6-2. Item characteristic curve for a "good" test item. The proportion of test takers getting the item correct increases as a function of total test score.

Kaplan, R.M., & Saccuzzo, D.P. Psychological testing: Principles, applications, and issues. Belmont, CA: Wadsworth

II. Item characteristic curves

The core of IRT models is the prediction of the response by the ability (trait) as depicted by the item characteristic curve

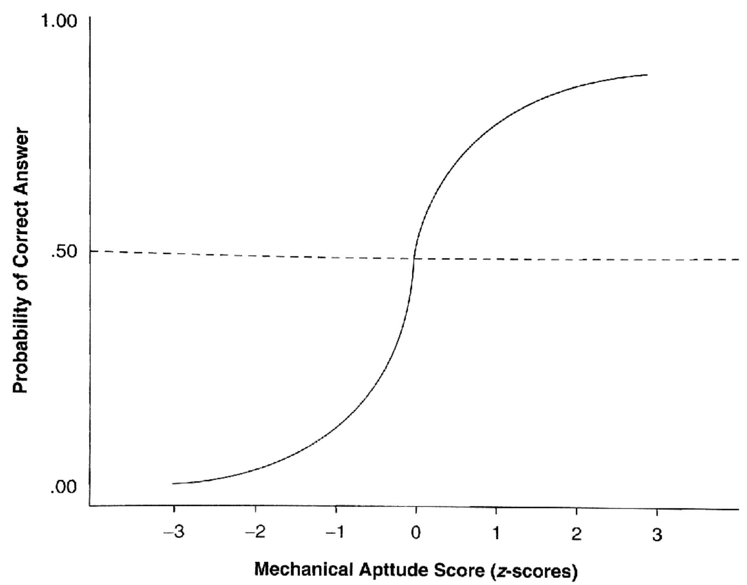


Figure 11.1 ICC Curve for a Test Item

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage.

III. One-Parameter Logistic Models

In the basic IRT model, the probability of a correct response, $\mathrm{P}(X_{is} = 1)$, is predicted by the ability, θ_s using logistic regression

The subscript i is for item and the subscript s is for subject (respondent)

The one-parameter logistic model (1PL) or the equivalent Rasch model (Rasch, 1960) is a logistic regression model in a slightly altered form

III. One-Parameter Logistic Models

Remember the logistic regression model

$$\ln\left(\frac{p}{1-p}\right) = a + bx$$

III. One-Parameter Logistic Models

For the IRT model, will use the logistic regression equation but will substitute the new notation

$$\ln\left(\frac{P_i}{1-P_i}\right) = \theta_x - \beta_i$$

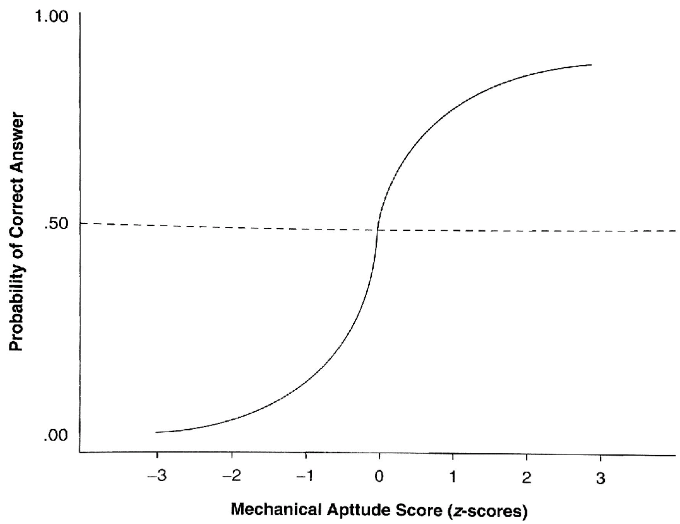
θ_x is the trait value and β_i is the item difficulty

This is parallel to the logistic equation, except θ_x replaces the x from the logistic regression equation, the slope b is assumed to be 1, β_i replaces the intercept, a, and the sign of the intercept is arbitrary change to negative (which creates subtraction)

III. One-Parameter Logistic Models

Now take another look at the mechanical ability example from the text

$$P(X_{is} = 1)$$



$$\theta_s$$

III. One-parameter Logistic Models

The ICC models the probability of the correct response, $\mathrm{P}(X_{is} = 1)$, as a function of the ability, θ_s

So to get back to the probability, we have do "undo" the natural log using the exponent function

$$P\left(X_{is} = 1\right) = \frac{e^{\theta_s - \beta_i}}{1 + e^{\theta_s - \beta_i}}$$

IV. Two-parameter Logistic Models

Remember item discrimination provides information about whether those who do well on the test are more likely to get the item correct

- Item discrimination index:

$$D = p_{high} - p_{low}$$

The *two-parameter* IRT model (or two-parameter logistic model, 2PL) takes into account the item difficulty and discrimination

IV. Two-parameter Logistic Models

The *two-parameter* IRT model (or two-parameter logistic model, 2PL) takes into account the item difficulty and discrimination

But discrimination is a continuous function that is captured by the logistic regression slope for the relationship between ability and probability the item is correct

IV. Two-parameter Logistic Models

The two-parameter logistic model (2PL) extends the 1PL model by adding the discrimination parameter, α_i

$$\ln\left(\frac{P_i}{1-P_i}\right) = \alpha_i\left(\theta_x - \beta_i\right)$$

This model is also a logistic regression, somewhat disguised, if you take into account the negative intercept tradition and that the term $\alpha_i\beta_i$ together represents the intercept

$$\begin{aligned}\ln\left(\frac{P_i}{1-P_i}\right) &= \alpha_i\theta_x - \alpha_i\beta_i \\ &= \alpha_i\theta_x + \left(-\alpha_i\beta_i\right) \\ &= bx + a\end{aligned}$$

IV. Two-parameter Logistic Models

We have to use the exponent function to get back to a predicted probability for a certain ability score

This time the probability is a function of the difficulty β_i and the relationship between the trait and probability, the discrimination parameter, α_i

$$P\left(X_{is}=1\right)=\frac{e^{\left(\alpha_i\left(\theta_s-\beta_i\right)\right)}}{1+e^{\left(\alpha_i\left(\theta_s-\beta_i\right)\right)}}$$

IV. Two-parameter Logistic Models

Often used to compare multiple items—items may differ in overall difficulty or discrimination

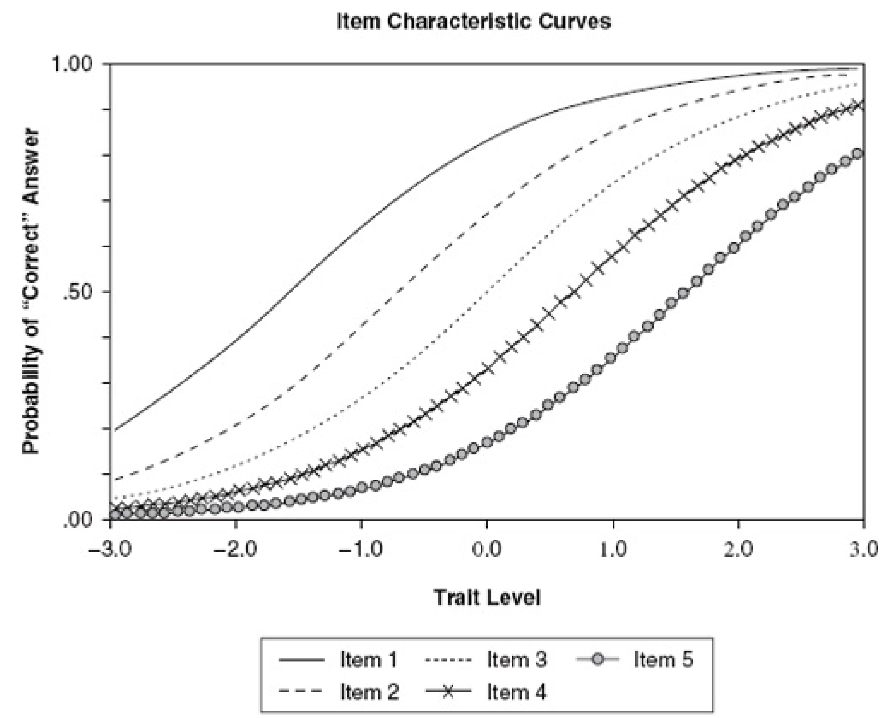


Figure 14.1 Item Characteristic Curves

Furr & Bacharach (2017) Psychometrics: An Introduction, Second Edition. Sage.

V. Other IRT Models

Three-parameter logistic model (3PL) includes the difficulty (β) and discrimination (α) parameters but also a guessing parameter, c

Used to take into account base rate of chance for guessing

- with true/false, guessing is 50% chance
- with multiple choice with 4 categories, guessing is 25%
- With multiple choice with 5 categories, guessing is 20%

V. Other IRT Models

Graded response model is IRT model for measures with items that have more than two response options (polytomous; Likert level of agreement)

Based on an *ordinal* logistic regression model

Book uses probability of response notation for certain category j, to represent probability value is at one level compared to the lower levels:

$$P(X_{is} \geq j \mid \theta_s, \beta_{ij}, \alpha_{,})$$

V. Other IRT Models

Graded response model ICCs

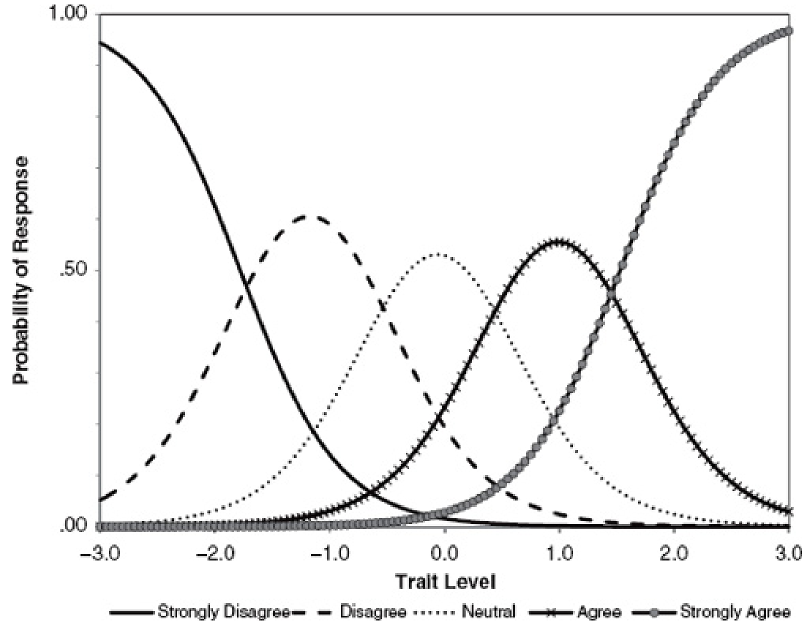


Figure 14.2 An Item Characteristic Curve, Based on the Graded Response Model

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage.

Portland State UNIVERSITY

V. Other IRT Models

Computer Adaptive Testing (CAT) is a use of IRT modeling

During the test, questions are selected based on the test taker’s assessed ability, θ. Questions are selected to maximize discrimination and ability score is updated afte each question.

V. Other IRT Models

Advantage CAT is that it can produce more precise test results for those at the lower and higher ends of the ability range compared with non-CAT methods

Fewer items need to be completed by test takers

Disadvantages are that very large pool of items needed and the prior IRT information must be well-studied

VI. Differential Item Functioning

Widely used IRT tool is differential item functioning (DIF)

DIF is used to investigate item bias

DIF may refer the values for either difficulty or discrimination or both differing across groups

Logistic regression used to test for these differences

VI. Differential Item Functioning

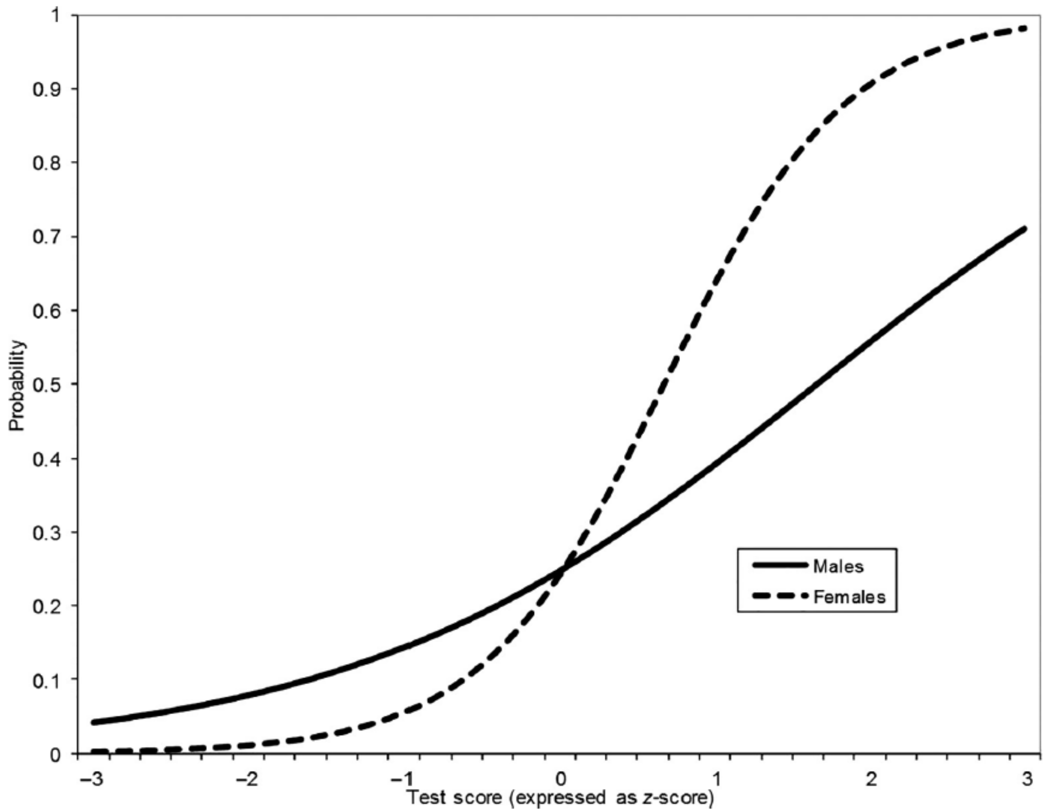


FIGURE 16.2. Plot showing nonuniform DIF for men and women.

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.