



# Teoria de Resposta ao Item e Rasch

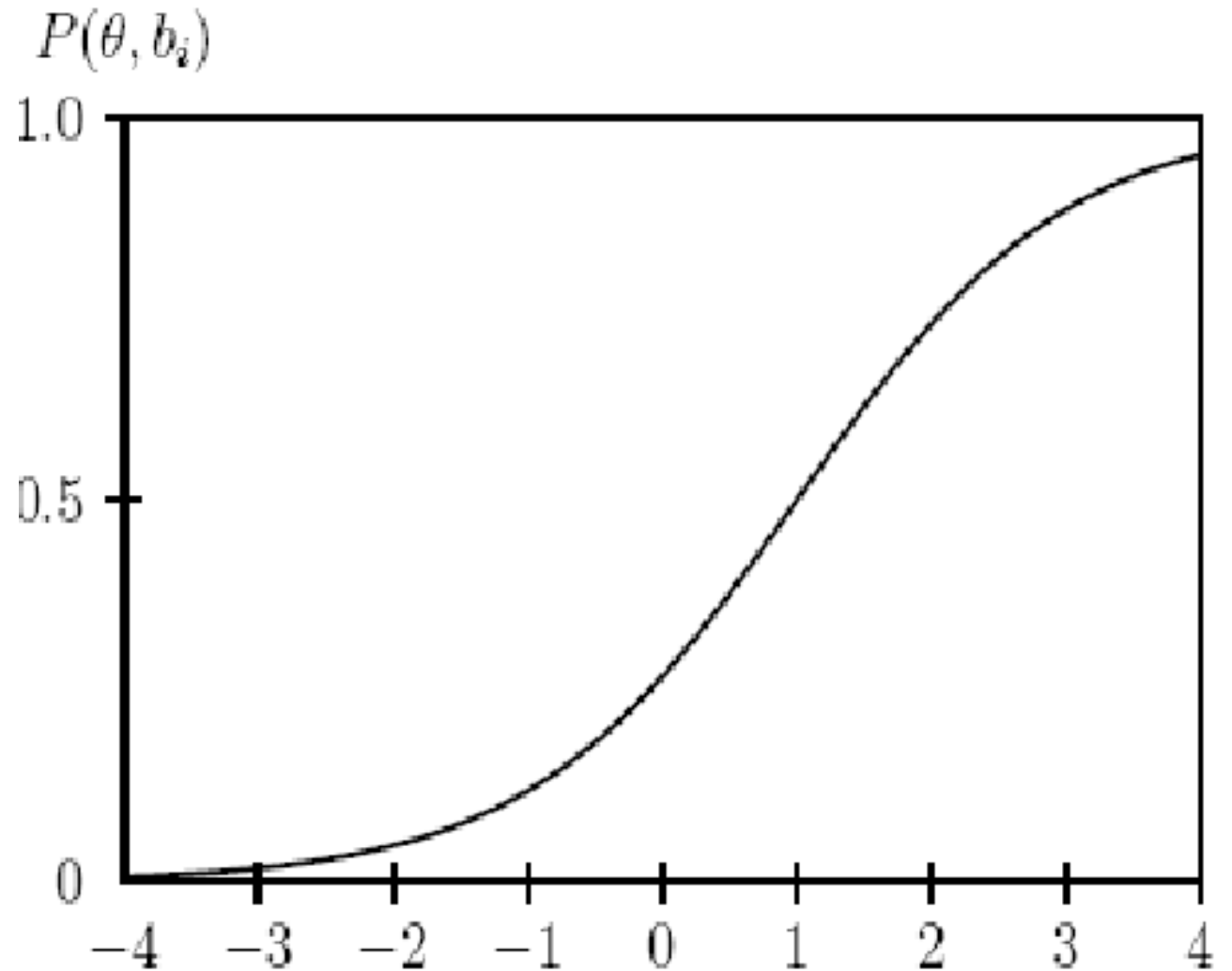
Psicometria

PPG Psicologia PUCRS

Prof. Dr. Wagner de Lara Machado

# Psicometria Moderna (1980)

- Desenvolvimento da Teoria de Resposta ao Item (TRI)
- Pesquisas em:
  - Vieses dos testes
  - Vieses de resposta
  - Vieses cognitivos
  - Validade e fidedignidade dos testes
  - Equiparação de escores
  - Construção e banco de itens
  - Testagem adaptativa



# Teoria Clássica dos Testes (TCT)

- <http://pepsic.bvsalud.org/pdf/avp/v11n2/v11n2a15.pdf>

*Avaliação Psicológica, 2012, 11(2), pp. 297-307 297*

## **PSICOMETRIA: FUNDAMENTOS MATEMÁTICOS DA TEORIA CLÁSSICA DOS TESTES**

*Ricardo Primi<sup>1</sup> – Universidade São Francisco, Itatiba, Brasil*

Assim, o escore observado pode ser definido como:

$$X_i = T_i + e_i$$

O escore verdadeiro ( $T_i$ ) pode ser concebido teoricamente de duas maneiras: (a) uma medida da variável em análise, sob condições ideais, usando um instrumento perfeito ou (b) a média de um conjunto de infinitas medidas da mesma variável, no mesmo sujeito, quando estas são independentes, usando um instrumento imperfeito – com erros de medidas. Ferguson (1981) estabelece essa definição como:

$$T_i = \lim_{k \rightarrow \infty} \frac{\sum_{j=1}^k X_i}{K}$$

# Modelos de Variáveis Latentes (MVL)

- <http://www.scielo.br/pdf/prc/v27n4/0102-7972-prc-27-04-00670.pdf>

ISSN 1678-7153

*Psychology/Psicologia Reflexão e Crítica*, 27(4), 670-678. – DOI: 10.1590/1678-7153.201427407

---

## Effects of Statistical Models and Items Difficulties on Making Trait-Level Inferences: A Simulation Study

*Efeitos dos Modelos Estatísticos e da Dificuldade dos Itens na Avaliação do Nível de Traço Latente dos Indivíduos: Um Estudo de Simulação de Dados*

Nelson Hauck Filho<sup>\*</sup>, <sup>a</sup>, Wagner de Lara Machado<sup>b</sup> & Bruno Figueiredo Damásio<sup>c</sup>

<sup>a</sup>*Universidade São Francisco, Itatiba, SP, Brasil,*

<sup>b</sup>*Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil*

& <sup>c</sup>*Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil*

---

## Análise fatorial

$$X_{ij} = v_j + a_{j1}F_{i1} + a_{j2}F_{i2} + \dots + a_{jm}F_{im} + \varepsilon_j$$

$X$  = escore observado em um item ou teste

$v$  = intercepto, geralmente fixado em zero

$a$  = carga fatorial

$F$  = escore fatorial

## Teoria de resposta ao item

$$P(U = u | \theta) = f(\theta, \gamma)$$

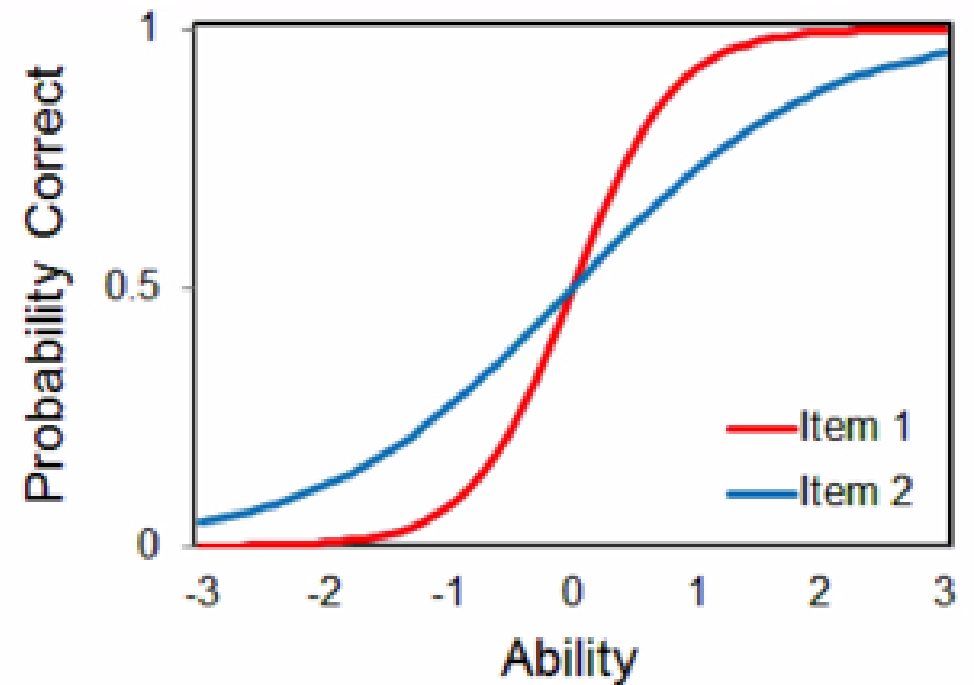
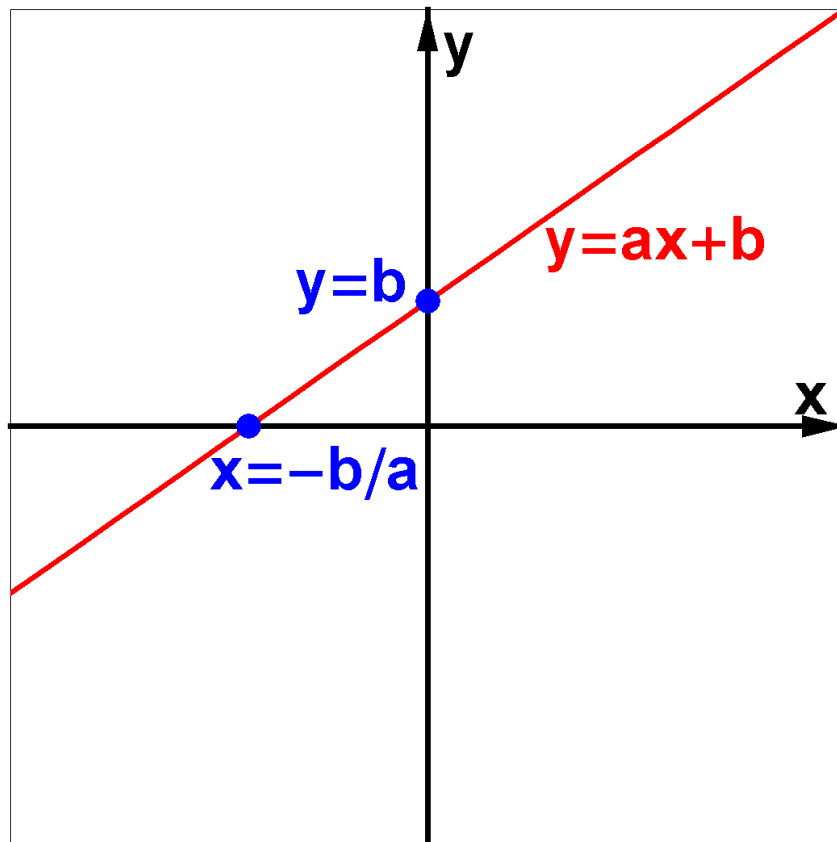
$P(U=u | \theta)$  = probabilidade de uma resposta, condicionada a um nível de traço latente

$f$  = é uma função probabilística

$\theta$  = nível de habilidade de um sujeito

$\gamma$  = parâmetros dos itens

# AF e TRI





AF e TRI

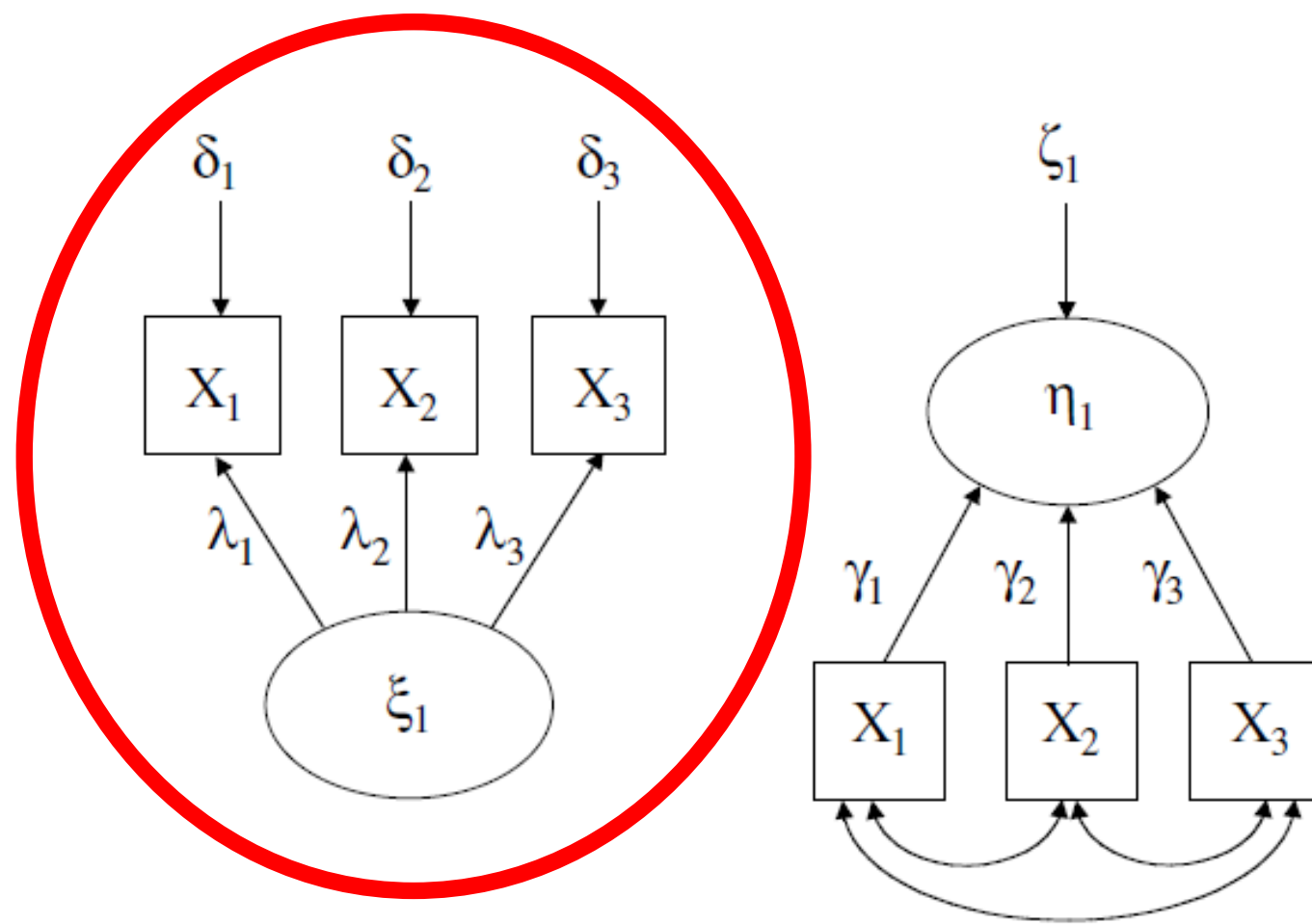
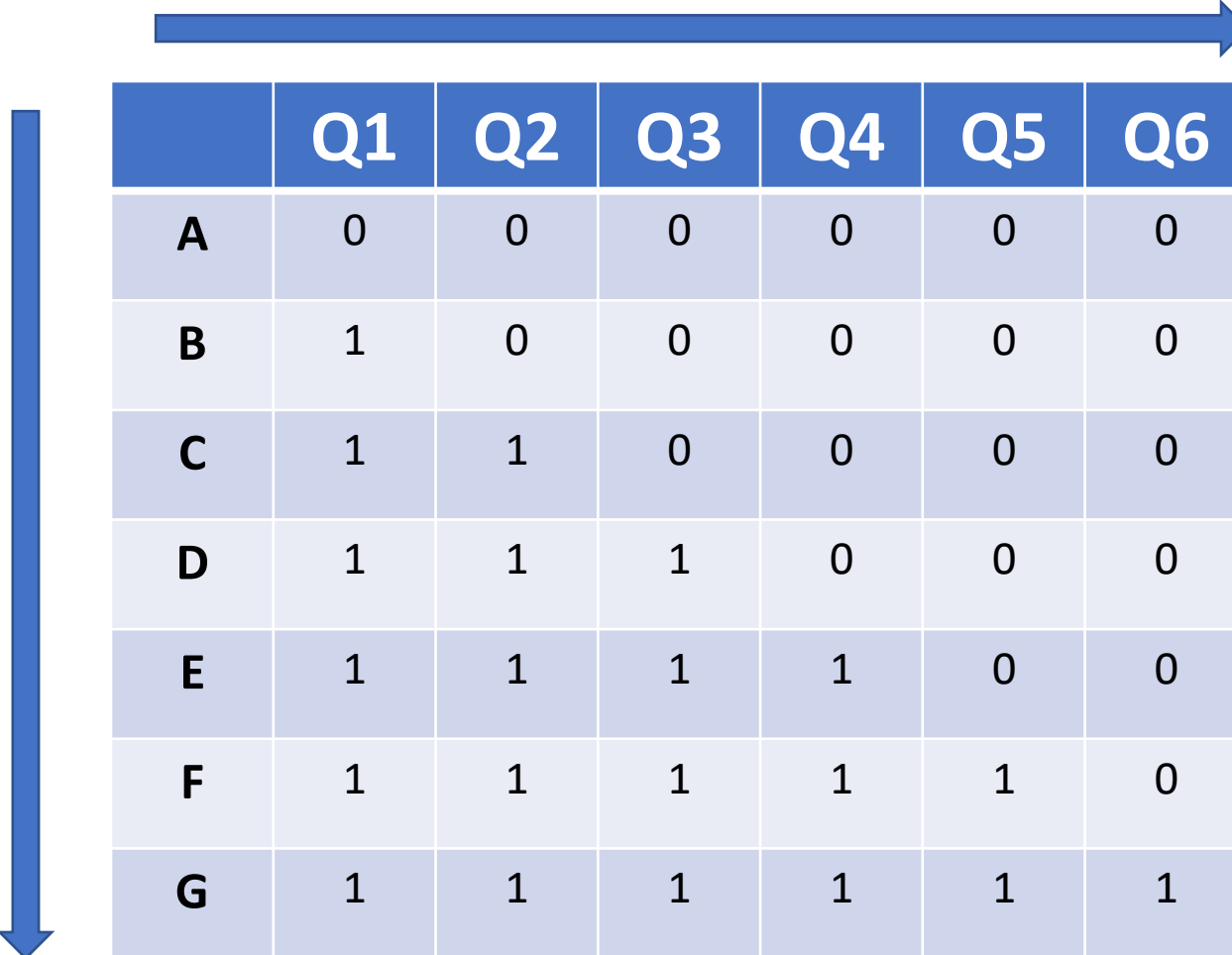


Figure 3.1. Two models for measurement. The figure in the left panel is the reflective measurement model. The  $X$ 's are observed variables,  $\xi$  is the latent variable,  $\lambda$ 's are factor loadings and the  $\delta$ 's are error terms. The right panel shows the formative model. The latent variable is denoted  $\eta$ , the  $\gamma$ 's are the weights of the indicators, and  $\zeta$  is a residual term.

# Escalograma de Guttman



|   | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 |
|---|----|----|----|----|----|----|
| A | 0  | 0  | 0  | 0  | 0  | 0  |
| B | 1  | 0  | 0  | 0  | 0  | 0  |
| C | 1  | 1  | 0  | 0  | 0  | 0  |
| D | 1  | 1  | 1  | 0  | 0  | 0  |
| E | 1  | 1  | 1  | 1  | 0  | 0  |
| F | 1  | 1  | 1  | 1  | 1  | 0  |
| G | 1  | 1  | 1  | 1  | 1  | 1  |

# **Fundamentos da Teoria da Resposta ao Item –TRI**

## **Basic Theory of Item Response Theory – IRT**

*Luiz Pasquali*

*Universidade de Brasília*

*Ricardo Primi*

*Universidade de São Francisco*

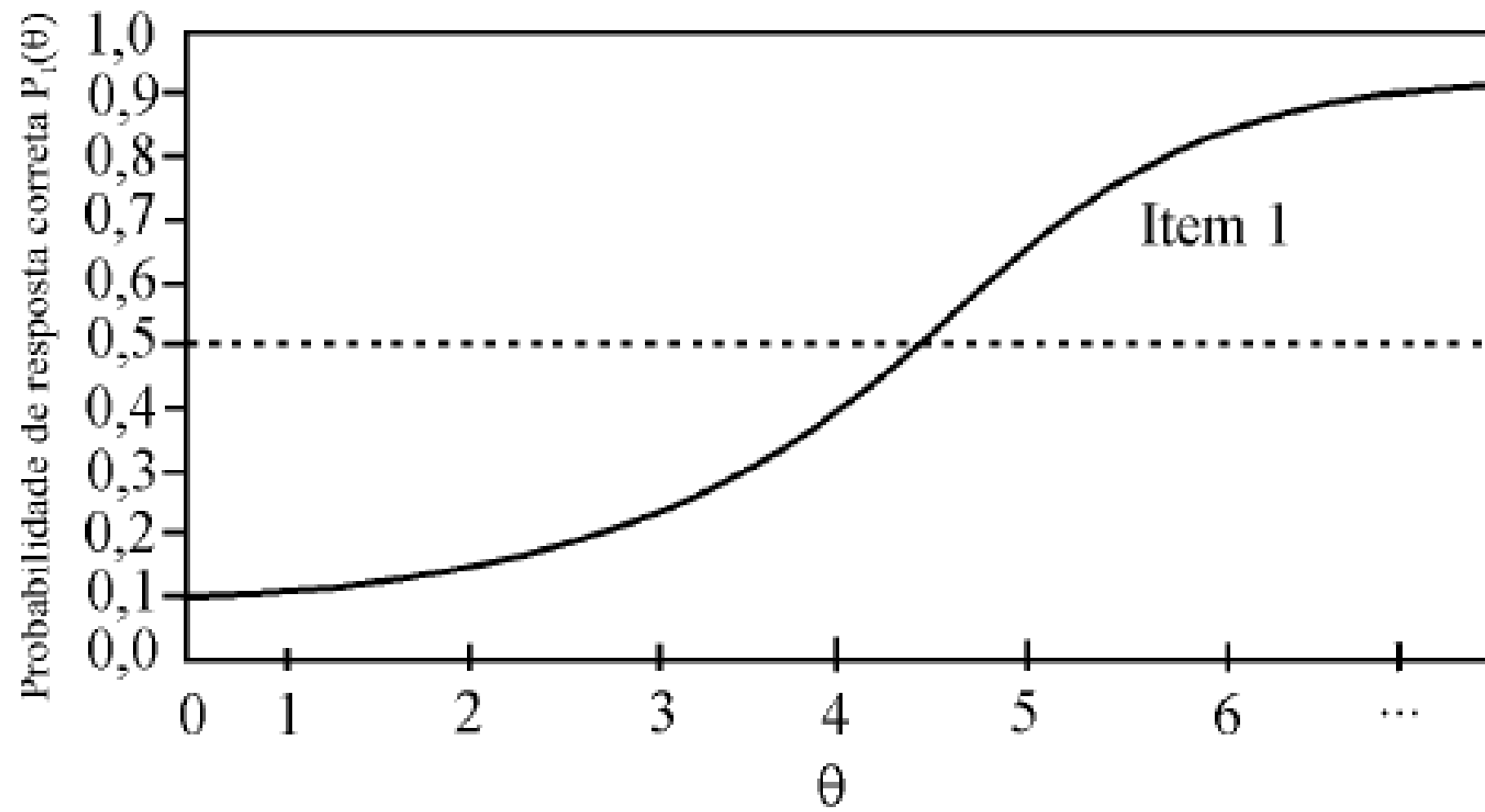
# Teoria da Resposta ao Item -TRI

- Surge nos anos 30, porém é sistematizada nos anos 60
- Entretanto, seu uso é recente, devido a dificuldades computacionais da época
- Teoria dos traços latentes
  - Resposta do sujeito a um estímulo, depende do nível do traço latente (teta -  $\theta$ )
- Foi inicialmente desenvolvidas para testes de aptidões (e.g. inteligência), ou seja, para respostas dicotômicas

# Proposições dos modelos de TRI e Rasch

- Mais habilidade → Maior probabilidade de acertar um determinado item, em relação a outro sujeito com menor habilidade
- Um item mais fácil tem maior probabilidade de ser acertado (aceitado, endossado) do que um item mais difícil, por um mesmo sujeito
- A resposta a um item se dá em função da habilidade do sujeito e da dificuldade do item. É uma função do traço latente.
- Conhecendo-se as respostas dadas a um determinado item, pode-se estimar o nível de traço latente em cada probabilidade de acerto

- Os ITENS são avaliados
- Cada item quando analisado, produz uma representação gráfica das probabilidades de acerto a um item ( $P_i$ ) em relação ao nível do traço latente ( $\theta$ ), expresso em “Logits”
- A escala de Teta ( $\theta$ ), vai de -4 a +4
- A análise do desempenho do item é comparada com uma curva chamada “ogiva normal”, parecida com a curva de Gauss, porém é uma curva logarítmica probabilística
- Essa curva é chamada Curva Característica do Item - ICC



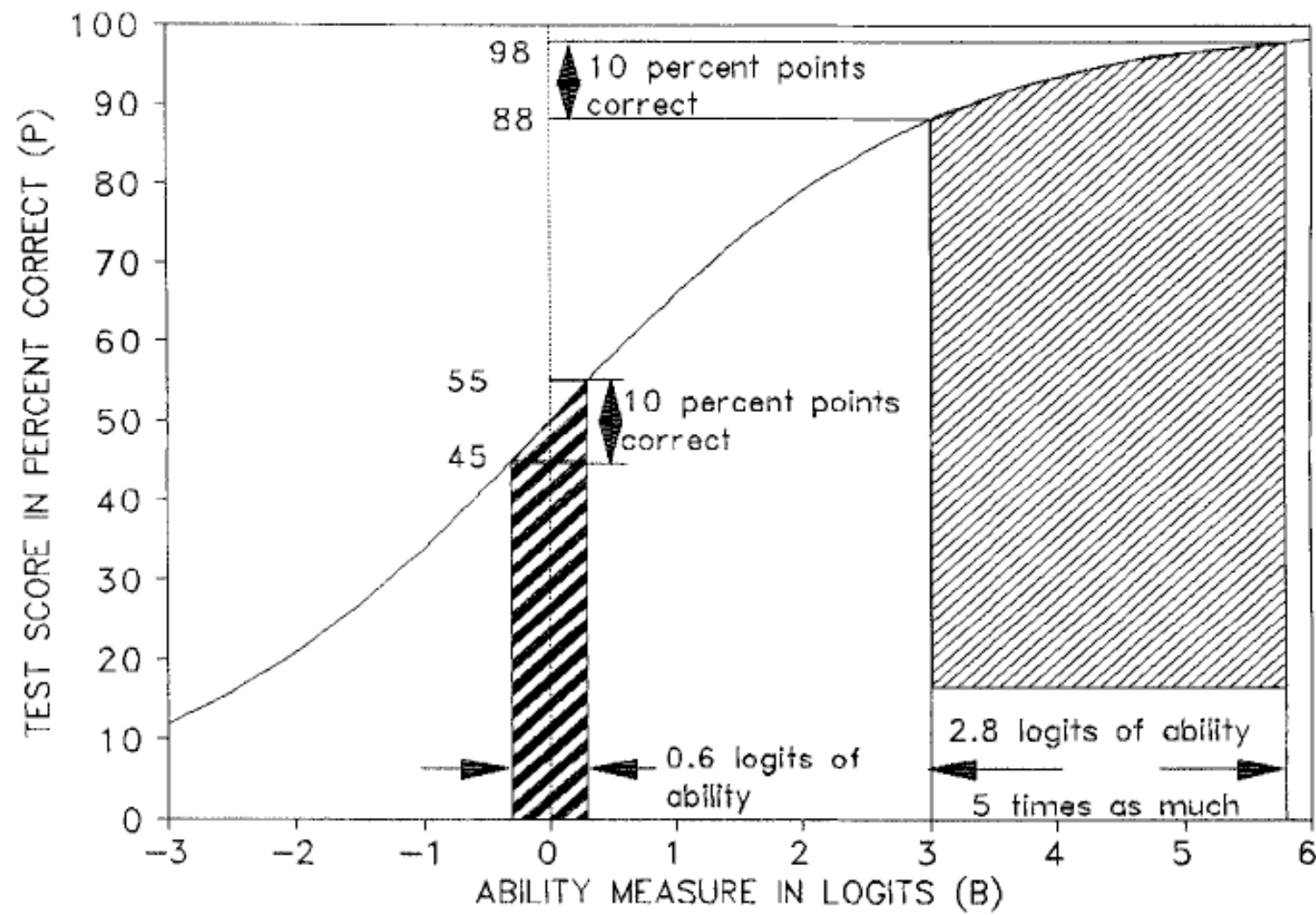


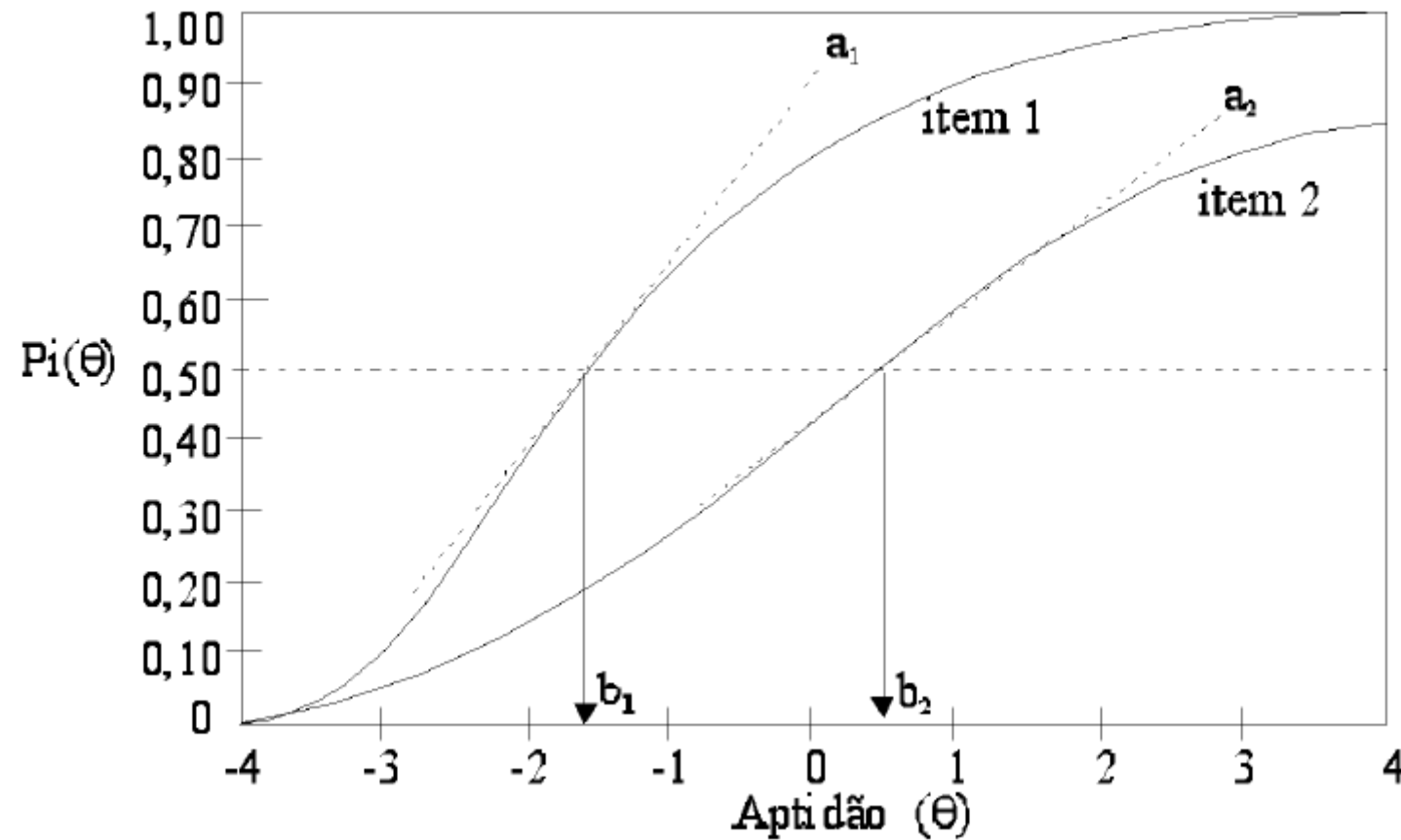
FIGURE 1. *Extreme raw scores are biased against measures*

(Wright, 1997)

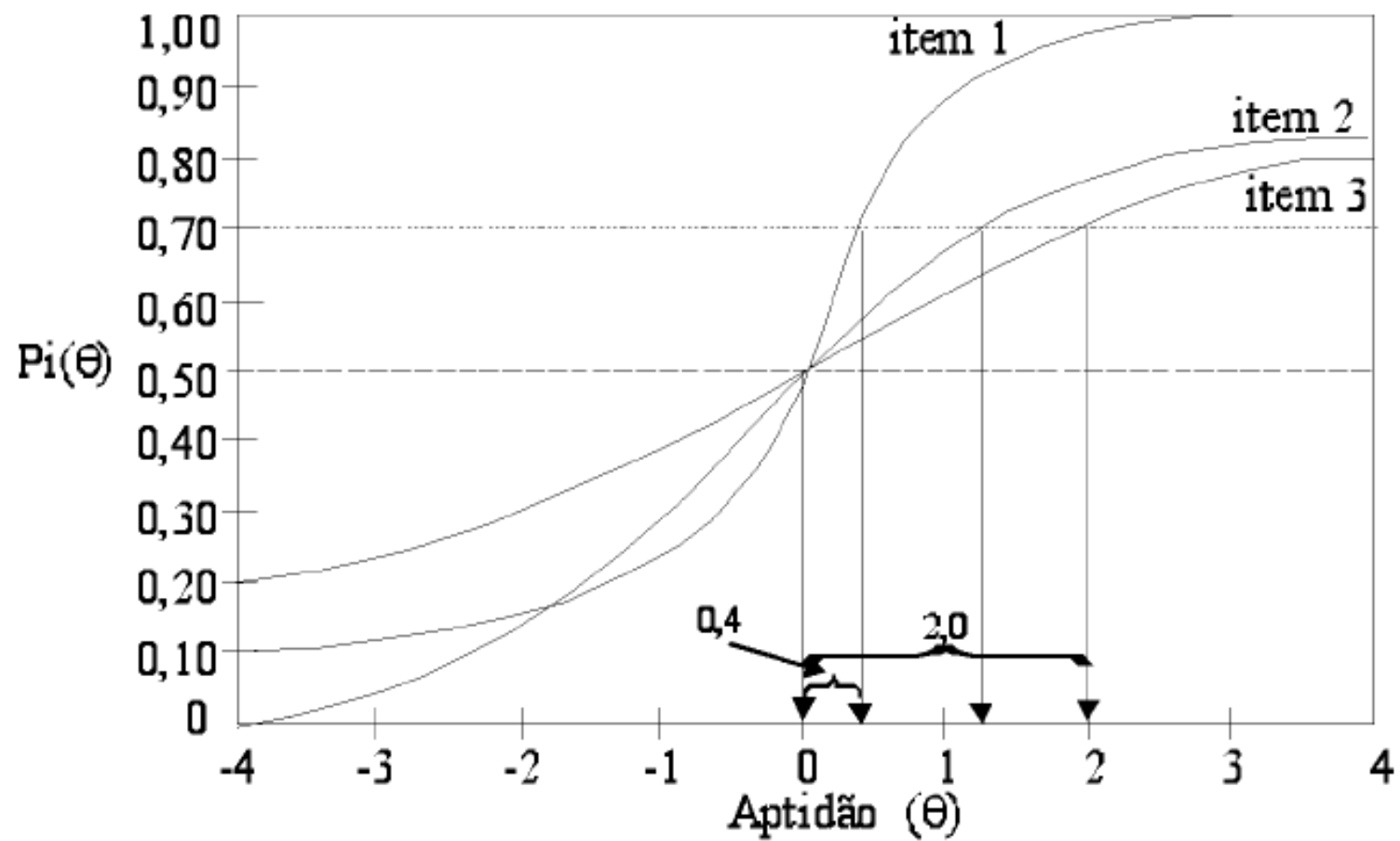


- São estimados os parâmetros dos itens no que diz respeito:
  - Dificuldade (b) (-3 a 3)
  - Poder de discriminação (a) (0 a 3)
  - Resposta ao acaso (chute; c)

- Discriminação (a) e dificuldade (b) de dois itens:



- Discriminação de 3 itens:



# EXEMPLO

## ESCALA DE SATISFAÇÃO DE VIDA

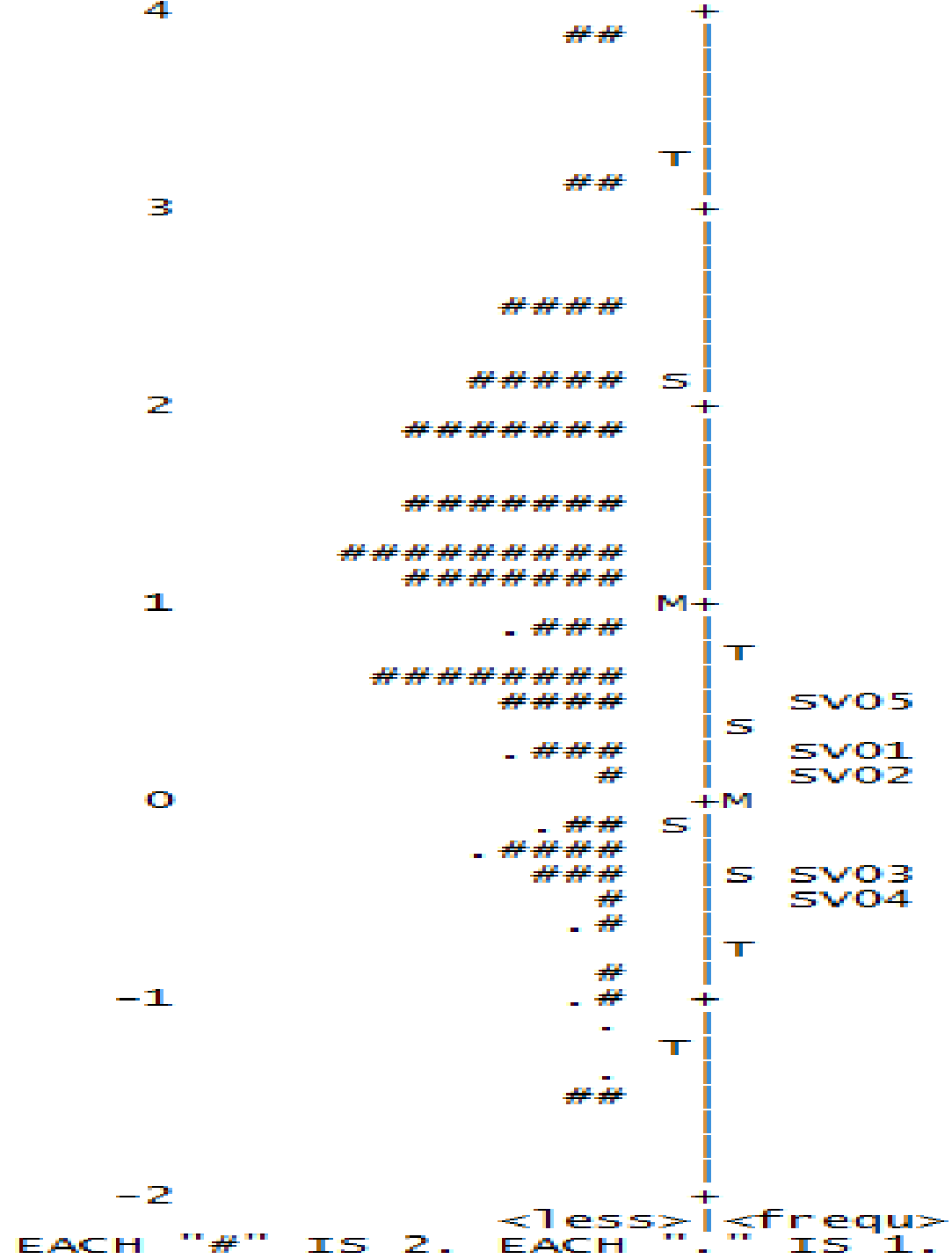
Logo abaixo, você encontrará 5 afirmativas. Assinale na escala ao lado de cada afirmativa o quanto ela descreve a sua situação pessoal. Não há respostas certas ou erradas, mas é importante você marcar com sinceridade como você se sente com relação a cada uma dessas afirmativas.

- 1- A minha vida está próxima do meu ideal.
- 2- Minhas condições de vida são excelentes.
- 3- Eu estou satisfeito com a minha vida.
- 4- Até agora eu tenho conseguido as coisas importantes que eu quero na vida.
- 5- Se eu pudesse viver a minha vida de novo eu não mudaria quase nada.

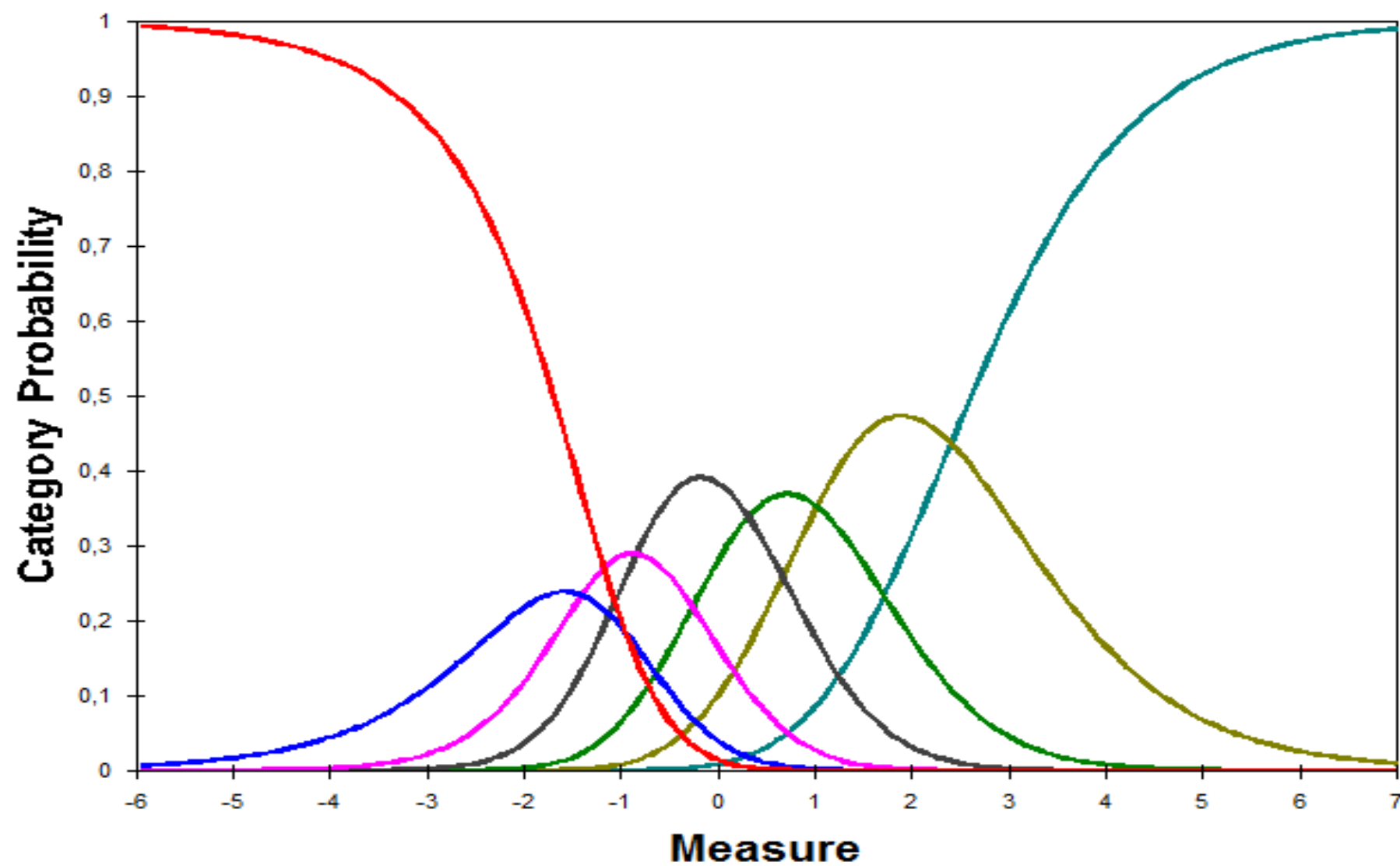
Discordo Plenamente \_|\_1\_|\_2\_|\_3\_|\_4\_|\_5\_|\_6\_|\_7\_|\_ Concordo Plenamente

## ITEM STATISTICS: MEASURE ORDER

| ENTRY  | TOTAL | TOTAL |         | MODEL | INFIT |      | OUTFIT |      | PT-MEASURE |      | EXACT MATCH |      |      |
|--------|-------|-------|---------|-------|-------|------|--------|------|------------|------|-------------|------|------|
| NUMBER | SCORE | COUNT | MEASURE | S.E.  | MNSQ  | ZSTD | MNSQ   | ZSTD | CORR.      | EXP. | OBS%        | EXP% | ITEM |
| 5      | 757   | 162   | .49     | .08   | 1.34  | 2.8  | 1.35   | 2.9  | .76        | .74  | 30.9        | 37.2 | SV05 |
| 1      | 796   | 162   | .25     | .08   | .84   | -1.5 | .88    | -1.1 | .69        | .73  | 40.1        | 39.0 | SV01 |
| 2      | 820   | 162   | .09     | .08   | 1.06  | .5   | 1.05   | .5   | .66        | .72  | 46.3        | 40.5 | SV02 |
| 3      | 879   | 162   | -.32    | .09   | .73   | -2.5 | .70    | -2.8 | .77        | .70  | 58.6        | 43.3 | SV03 |
| 4      | 902   | 162   | -.50    | .09   | .92   | -.7  | .83    | -1.5 | .74        | .69  | 48.1        | 43.5 | SV04 |
| MEAN   | 830.8 | 162.0 | .00     | .08   | .98   | -.3  | .96    | -.4  |            |      | 44.8        | 40.7 |      |
| S.D.   | 53.2  | .0    | .36     | .00   | .21   | 1.8  | .23    | 2.0  |            |      | 9.2         | 2.5  |      |

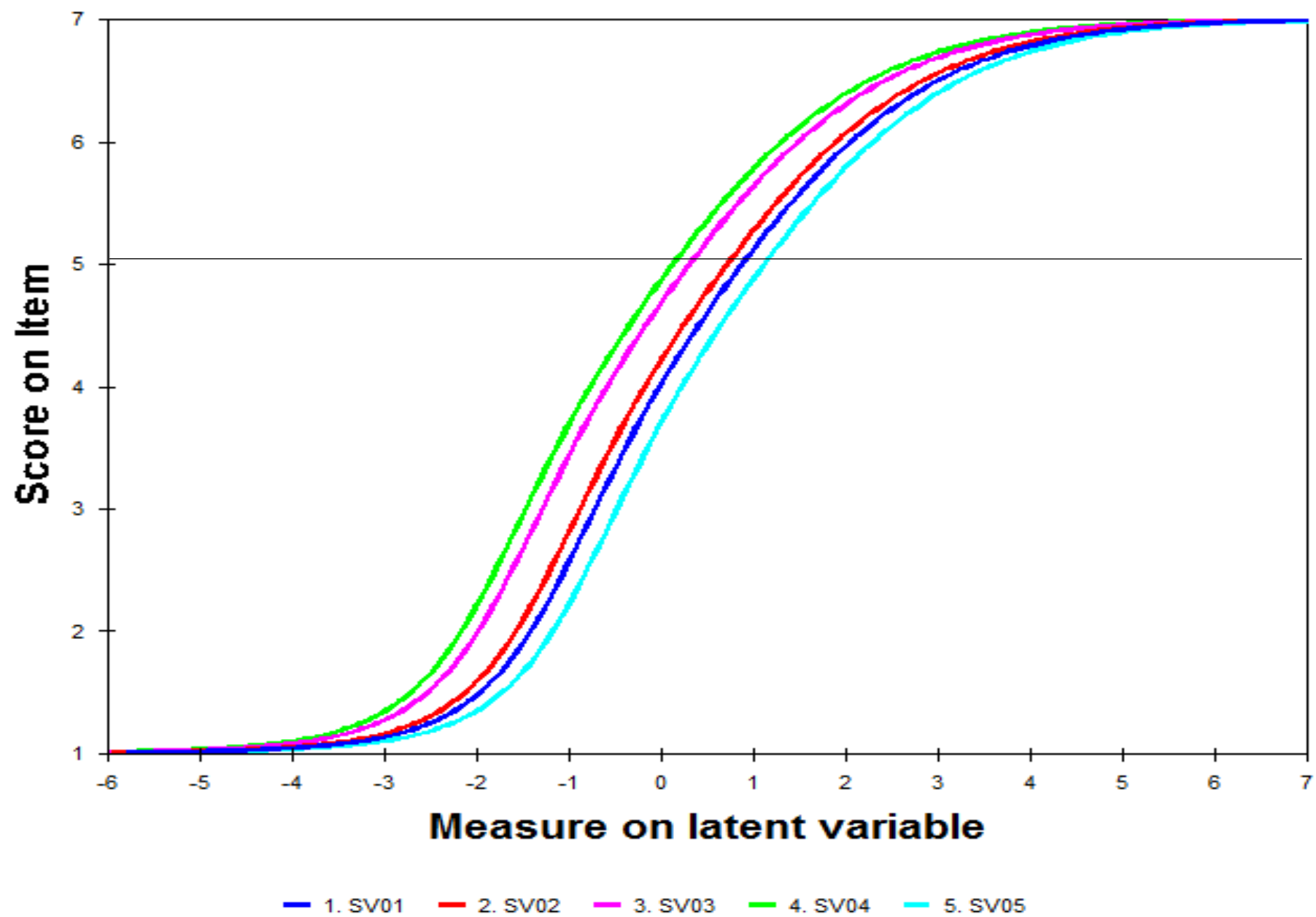


## 2. SV02



Category probability: 1    Category probability: 3    Category probability: 5    Category probability: 7  
Category probability: 2    Category probability: 4    Category probability: 6

Item Characteristic Curves





- A TRI não substitui a TCT, porém avança no sentido da estimação da fidedignidade
- Vantagens:
  - Avalia o desempenho dos itens de um teste
    - Dificuldade, discriminação e resposta ao acaso
  - Avalia a adequação do tipo de escala de resposta
  - O nível de aptidão do sujeito independe do teste
  - O cálculo dos parâmetros dos itens independem da amostra utilizada

# Modelo Rasch (Hauck, 2012)

- O “*modelo de Rasch*”, *atualmente*, designa uma *família de modelos* para a *construção de medidas* a partir de *dados de natureza ordinal*
- Esse modelo tem sido apresentado na literatura como uma das técnicas da Teoria de Resposta ao Item (TRI), pois compartilha muitas das características que definem os ditos modelos analíticos da TRI
- Algumas dessas características fazem as técnicas da TRI superiores às técnicas da Teoria Clássica dos Testes (análise fatorial, análise de componentes principais etc.) para a construção e análise de instrumentos psicométricos

- Algumas das vantagens incluem: estimativas mais realistas dos erros de medida e da fidedignidade do instrumento, avaliação dos itens de forma independente da amostra, estimativa do traço latente independente dos itens utilizados e do formato das escalas de resposta e melhor apreciação do significado dos escores de um indivíduo em um teste
- O modelo de Rasch, no entanto, diferencia-se de outros modelos da TRI por ser baseado em uma teoria consistente da medida, prescrevendo condições que os dados devem respeitar para que seja possível a mensuração
- A técnica é utilizada tanto para a construção e análise de instrumentos quanto para a formação de uma medida, a partir de escores brutos, para ser imputada em outras técnicas estatísticas que requerem dados com natureza intervalar

- A análise de Rasch, no entanto, não substitui a clareza teórica que o pesquisador deve possuir para construir uma medida com significado psicológico
- O modelo é orientado para a medida fundamental, a suficiência estatística e a divisibilidade
- O modelo é uma versão não-determinística do Escalograma de Guttman
- Dois parâmetros são utilizados para explicar a resposta de um indivíduo a um item qualquer: a dificuldade desse item ( $D_i$  ou  $\delta$ ) e a habilidade do indivíduo ( $B_n$  ou  $\theta$ ), ou seja,  $P = B_n - D_i$

# Tópicos importantes

## II. AJUSTE DOS ITENS E DAS PESSOAS

- Índices *infit* e *outfit*: são quadrados médios (*mean square*), ou seja, medidas de resíduos
- Como se comportaram os indivíduos com diferentes níveis de habilidade ( $B_n$  ou  $\theta$ ) ao responderem ao item?
- *Infit*: medida de ajuste sensível às respostas dos indivíduos com nível de habilidade compatível com o item que representa uma média dos resíduos padronizados ponderados pela variância individual de cada resíduo
- *Outfit*: medida de ajuste sensível às respostas dos indivíduos com nível de habilidade distante do item que representa uma média não-ponderada dos resíduos

# Tópicos importantes

## II. AJUSTE DOS ITENS E DAS PESSOAS

- O valor ideal é sempre 1,00 para ambos os índices de ajuste, o que representa que os dados se comportaram exatamente como o esperado pelo modelo. No entanto, abre-se um intervalo de tolerância para valores próximos, sendo a recomendação mais liberal de 0,50 até 1,50
- 
- *Infit* abaixo de 0,50: o item é redundante e não acrescenta informação à medida
- *Infit* acima de 1,50: o item apresenta um padrão distinto do Escalograma de Guttman, sendo potencialmente prejudicial ao modelo
- *Outfit* abaixo de 0,50: comum acontecer para respostas faltantes substituídas no banco de dados
- *Outfit* acima de 1,50: comum acontecer para respostas dadas ao acaso ou sem muito interesse

# Tópicos importantes

## III. UNIDIMENSIONALIDADE

- O modelo de Rasch **NÃO ASSUME** que os dados são unidimensionais
- Dados em ciências sociais e humanas são tipicamente multidimensionais
- O modelo, entretanto, prescreve sobre como os dados devem se comportar para que seja construída uma medida a partir das observações
- Toda medida é unidimensional
- A questão, na análise de Rasch, é descobrir se outras dimensões são salientes a ponto de interferir na capacidade do instrumento de medir o traço latente

# Tópicos importantes

## III. UNIDIMENSIONALIDADE

- Variância explicada: “receita tradicional” (60%)
- Análise de componentes principais dos resíduos do modelo
- Contrastes devem possuir autovalor  $< 2.0$
- Interpretação dos gráficos de contrastes principais
- Correlações residuais entre itens  $< .30$



# A History of Social Science Measurement

Benjamin D. Wright

*University of Chicago, MESA Psychometric Laboratory*

---

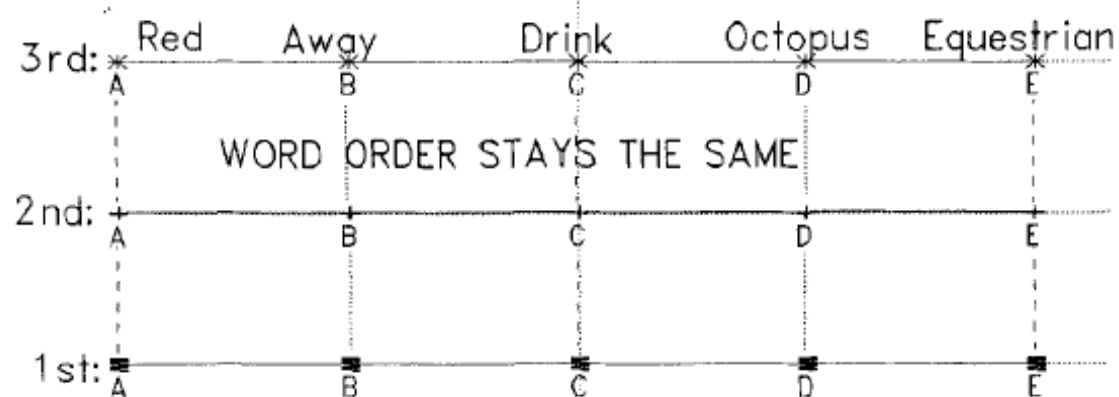
*What are the historic origins of measurement? What are the basic requirements for fundamental measurement? What are some common pitfalls to avoid? Do item response models meet these needs?*

set upon them by evil tax collectors. The more righteous law is justice and good conduct . . . I order you to take in taxes only the weight of seven. (Damascus, 723 AD)

The Magna Carta of King John of England requires that:

## Three Perceptions of One! Variable

Five Rasch Items



WORD ORDER STAYS THE SAME

LOGIT SCALE CALIBRATES RULER

-2

-1

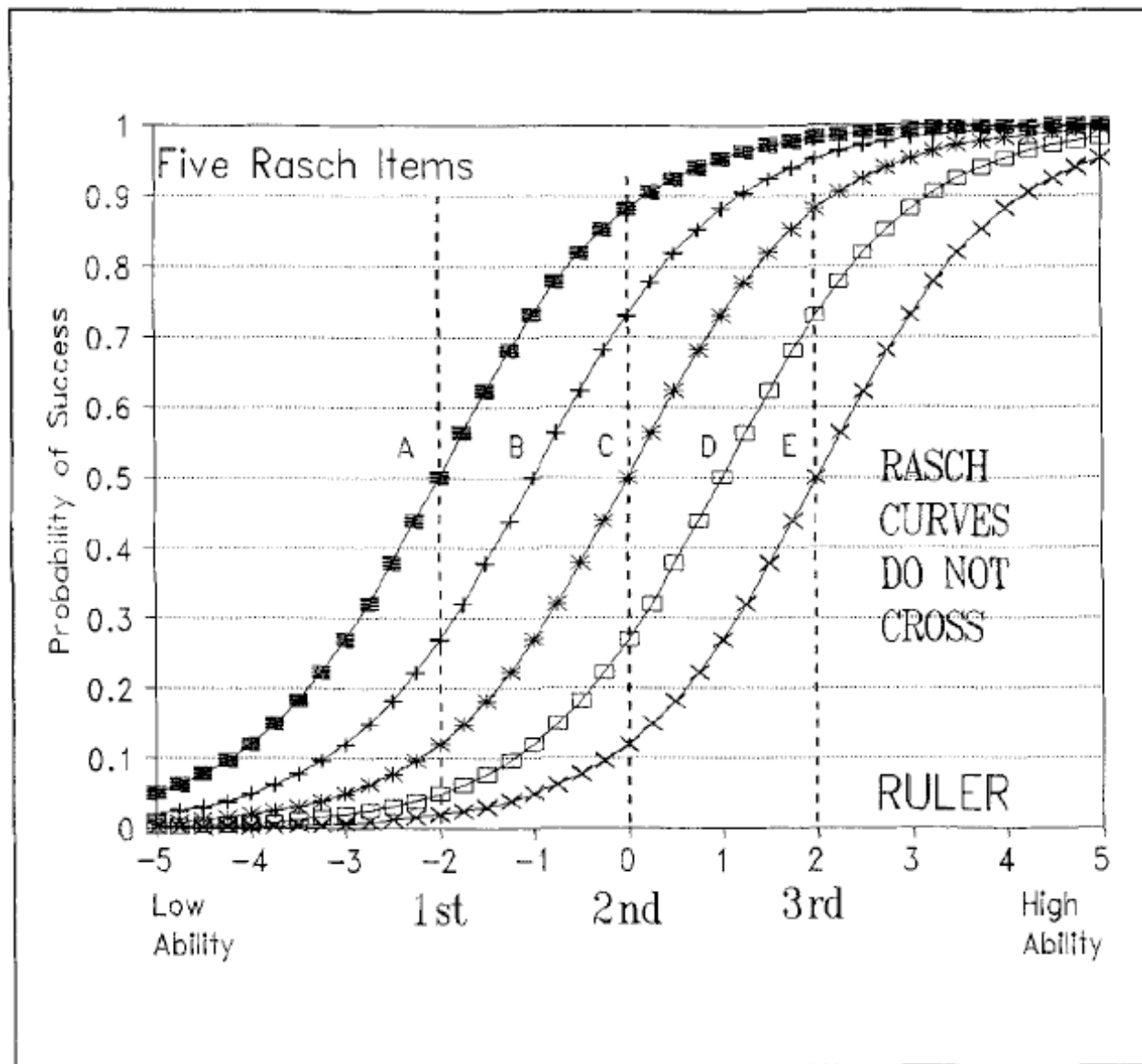
0

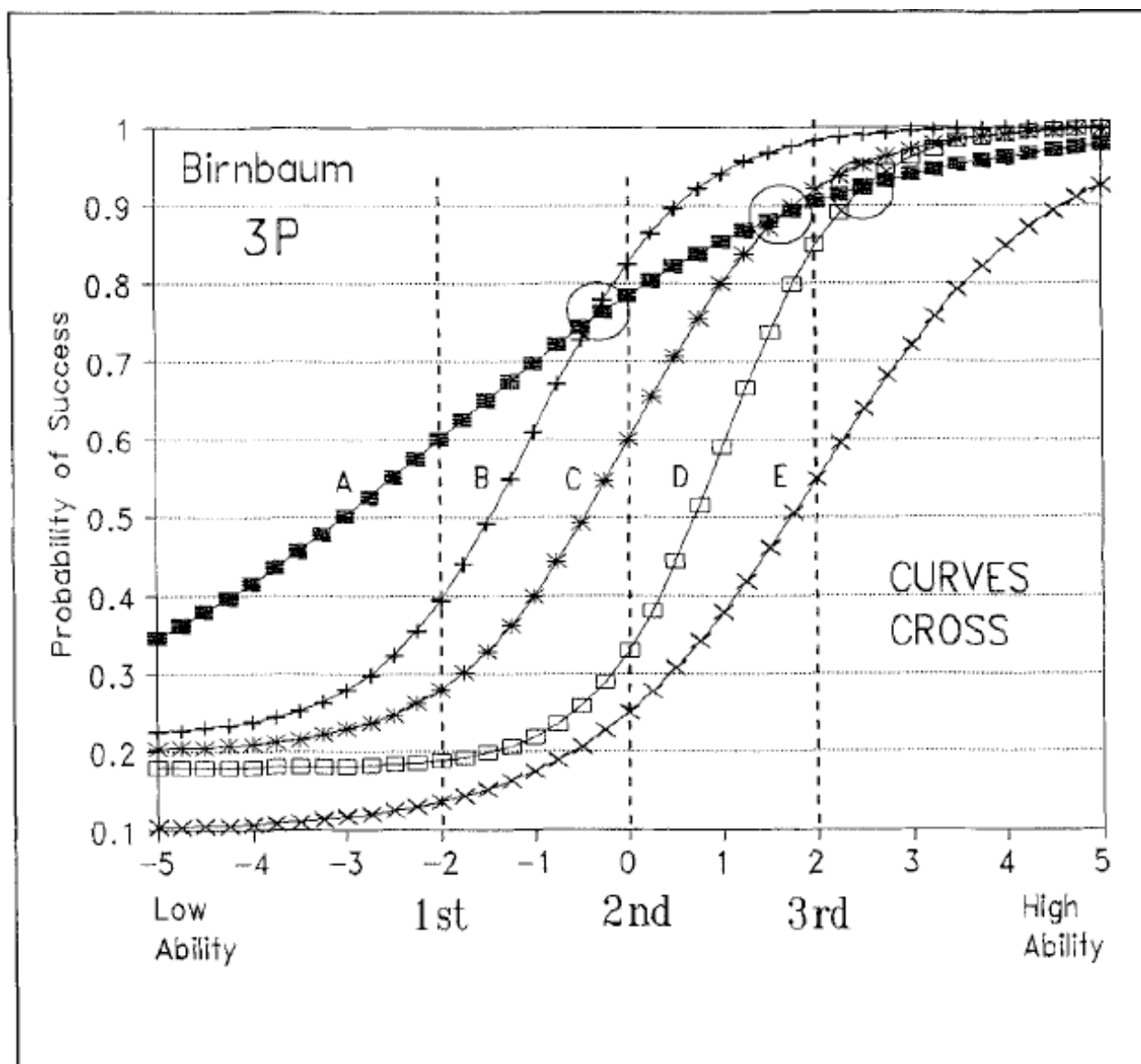
1

2

Item Difficulty (relative to item C)

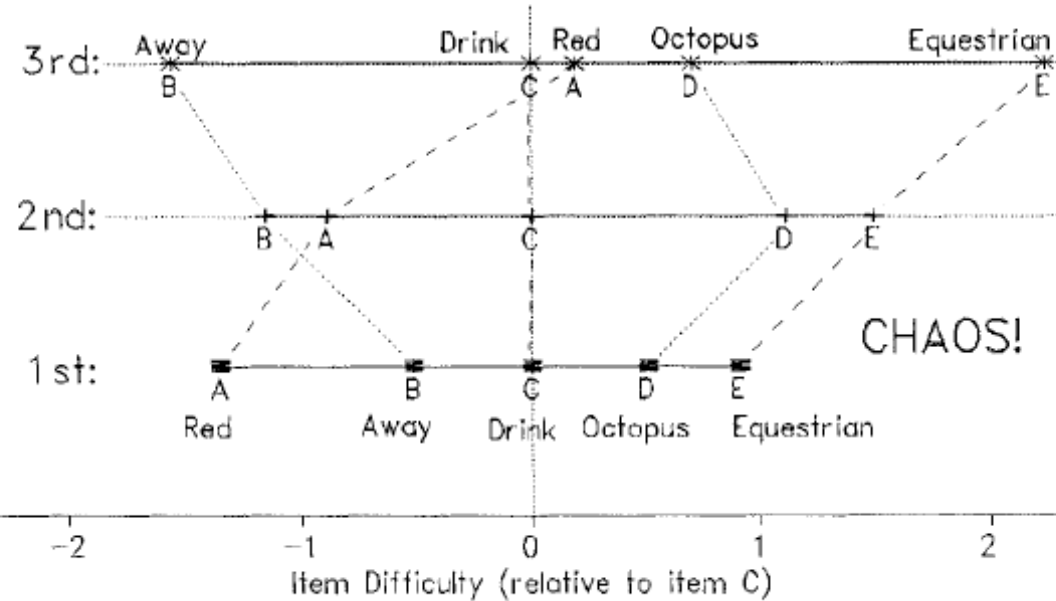
One Variable Defined!





## Three Different Variables

### Five Birnbaum Items



What is the Item Definition of this Variable?

# Exemplos:

- <https://github.com/wagnerLM/psicometria/blob/master/script%20TRI%20PL%20Rasch.R>

## De outras aulas:

- <https://github.com/wagnerLM/quantia/blob/master/script%20TRI%20continuacao.R>
- <https://github.com/wagnerLM/tutoriais/blob/master/script%20tri%20Igor.R>