

UNIVERSIDADE
SÃO FRANCISCO

Teoria de Resposta ao Item

Dr. Ricardo Primi

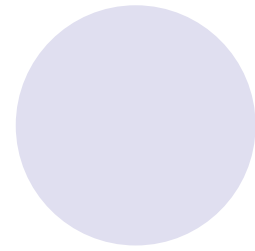
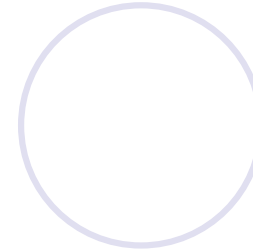
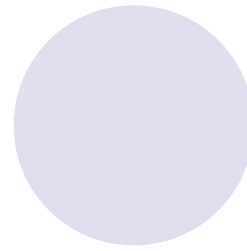
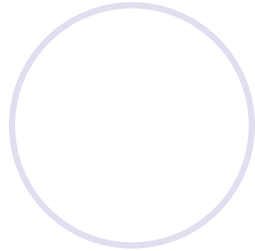
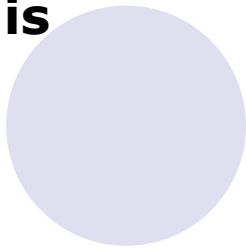
Programa de Mestrado e Doutorado em Avaliação Psicológica

Universidade São Francisco

Programa

- 
- Revisão dos conceitos da Psicometria Clássica: Índice de dificuldade, índice de discriminação e consistência interna. Limitações da Psicometria Clássica. Introdução à análise gráfica dos itens.
 - A Curva Característica dos Itens, Modelos da TRI (1, 2 e 3 parâmetros), Revisão de conceitos matemáticos (Capítulos 1 e 2, Baker, 2001)
 - Prática 1. Preparação dos bancos de dados para análise. Uso dos programas ITEMAN, XCALIBRE e WINSTEPS para estimação dos parâmetros dos itens e dos sujeitos.
 - Estimação dos Parâmetros dos Itens e dos Sujeitos. Métodos de Estimação. Invariância dos Parâmetros. (Capítulos 3 e 5, Baker, 2001)
 - Índices de ajuste ao modelo. Índices para os Itens e para os Sujeitos. Pressuposto da unidimensionalidade Essencial.
 - Prática 2. Verificação de ajuste nos programas XCALIBRE e WINSTEPS. Análise fatorial dos resíduos no WINSTEPS.
 - Equalização de Notas. Questões sobre a Definição da Métrica da Escala. Procedimentos de Equalização.
 - Prática 3. Procedimento de equalização no XCALIBRE e WINSTEPS.
 - O conceito de precisão na TRI: Curva de Informação do Teste. Fatores que afetam a informação.
 - Funcionamento diferencial do Item – DIF
 - Prática 4 Procedimentos de análise do DIF no WINSTEPS
 - Introdução aos modelos Créditos parciais e Escalas de Pontuações
 - Prática 5. Créditos parciais no WINSTEPS.

Materiais



- Programas de análise: BIRT, ITEMAN, XCALIBRE, WINSTEPS e SPSS
- Programa de edição de texto: TextPad (<http://www.textpad.com/>)
- Sites:
 - <http://edres.org/irt/>
 - <http://www.winsteps.com/> e <http://www.rasch.org/>
 - www.assess.com
 - <http://work.psych.uiuc.edu/irt/>
- Texto básico
 - Baker, F. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. (disponível on-line em <http://edres.org/irt/>)
 - Exercícios!!!!
- Outros materiais: transparências, planilhas e arquivos de dados

O que é TRI



- *Item Response Theory* (IRT)
- Estado atual das teorias (métodos) psicométricas sobre medida
- Teoria no sentido estatístico e não substancial (Goldstein, 2001)
- Conjunto de modelos matemáticos para representar elementos básicos da medida (parâmetros).
- Psicometria clássica era centrada no teste. A TRI é centrada no item
- Supera limitações do modelo clássico e o engloba
- Traz mais informação (sobre a consistência do instrumento e da medida das pessoas, sobre o construto) e permite algumas aplicações interessantes com relativa facilidade (análise de viés – DIF, equalização de notas, testagem adaptativa, banco de itens dentre outras).
- Pioneiros:
 - Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, New Jersey, Lawrence Erlbaum Associates:
 - Lawley, D. N. (1943). The application of the maximum likelihood method to factor analysis. British Journal of Psychology 33: 172-175.
 - Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

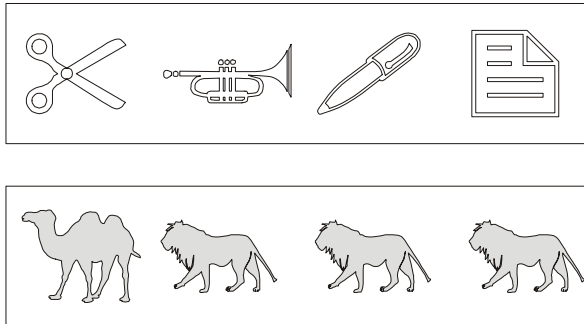
Teoria de Resposta ao Item

(from wikipedia ... http://en.wikipedia.org/wiki/Item_Response_Theory)

- Item response theory (IRT) is a body of related psychometric theory that provides a foundation for scaling persons and items based on responses to assessment items. The central feature of IRT models is that they relate item responses to characteristics of individual persons and assessment items. Expressed in somewhat more technical terms, IRT models are functions relating person and item parameters to the probability of a discrete outcome, such as a correct response to an item. Among other things, as a body of theory, IRT provides a basis for estimating parameters, ascertaining how well data fits a model, and investigating the psychometric properties of assessments. Psychometricians apply IRT in order to achieve tasks such as developing and refining exams, maintaining banks of items for exams, and equating for the difficulties of successive versions of exams (for example, to allow comparisons between results over time).
- IRT is often referred to as latent trait theory, strong true score theory, or modern mental test theory. The term latent is used to emphasise that discrete item responses are taken to be observable manifestations of the trait or attribute, the existence of which is hypothesized and must be inferred from the manifest responses. The other major body of psychometric theory of relevance to IRT is classical test theory. For tasks that can be accomplished using classical test theory, IRT generally brings greater flexibility and provides more sophisticated information. Some applications, such as computerized adaptive testing are enabled by IRT and cannot reasonably be performed using only classical test theory.

O Que é medida: Exemplo de um procedimento clássico

- Escala de Maturidade Mental Columbia (CMMS, 1992)



- Aplica-se, seguindo procedimentos padronizados, um conjunto de problemas que avaliam consistentemente uma determinada habilidade (CMMS, Gf).
- Corrige-se os resultados obtendo-se o escore bruto.
- Converte-se o resultado em uma escala padronizada.
- *Regra de conversão*: comparação com um grupo de referência para a indicação da posição relativa da criança em relação ao grupo.
- *Limitação* : Em função do mecanismo de mensuração, os escores informam quão boa ou ruim foi uma pessoa quando comparada com um grupo de referência, isto é, posição competitiva do sujeito em relação aos pares. Entretanto, dizem pouco sobre que processos mentais ocorrem e que culminavam nas diferenças observadas no desempenho.

Modelos de medidas: TCT

- Representar com números (medidas) as pessoas e os itens a partir das respostas
 - O que permite predizer as respostas a partir dos números

- O modelo clássico:

- $E = V + e$

V: valor verdadeiro (valor autêntico se medíssemos sem erro)

e: erro de medida

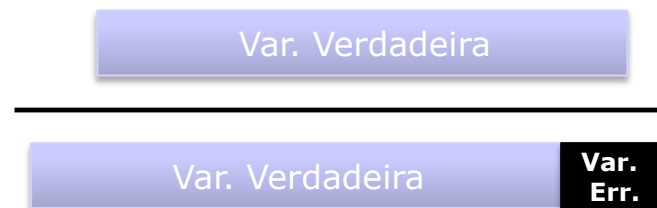
E: o escore observado que é a soma de pontos no teste. É uma estimativa do escore verdadeiro

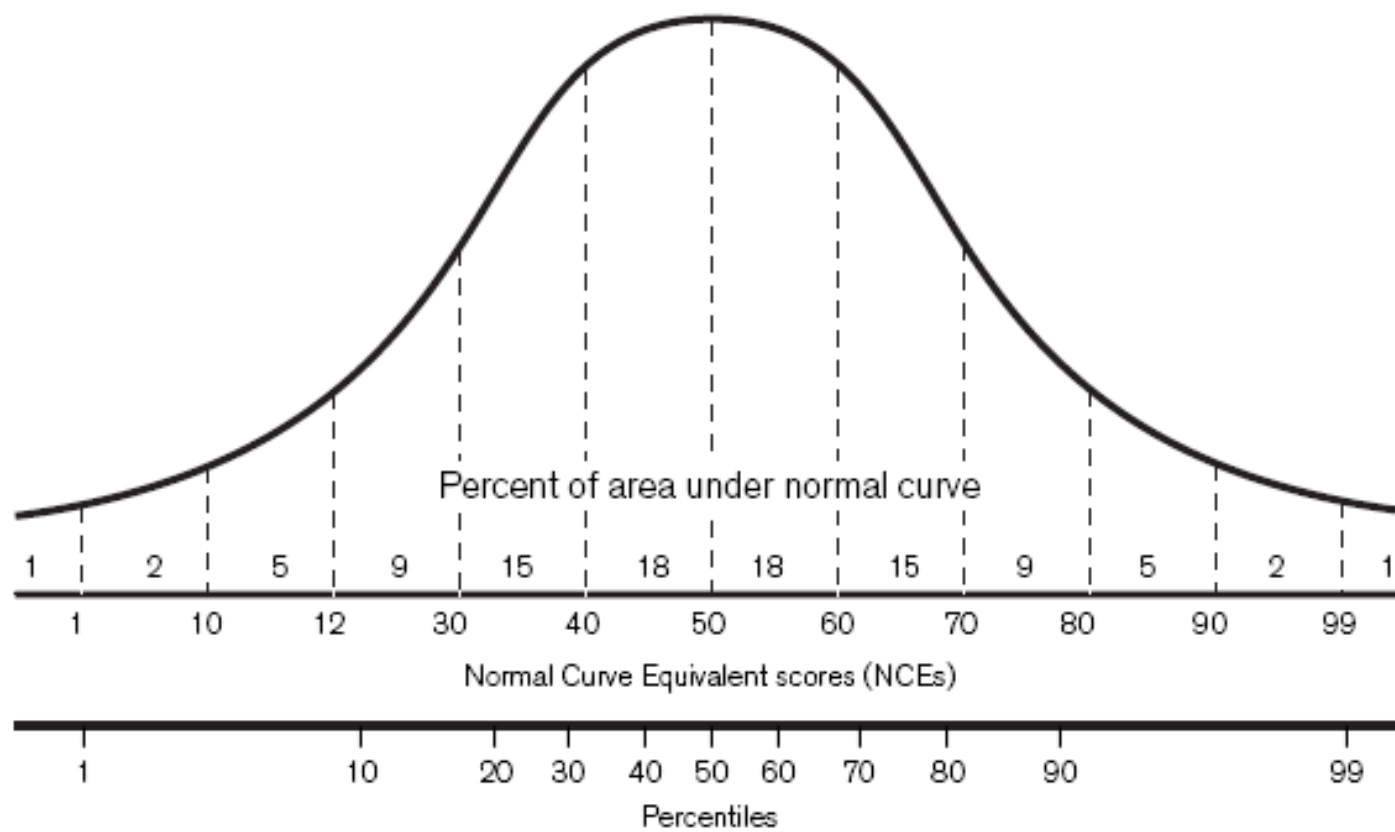
Obs: não importa qual item a pessoa respondeu ao teste, eles sempre representarão a mesma magnitude

Interpretação é feita por normatização

Controle de qualidade (precisão e validade)

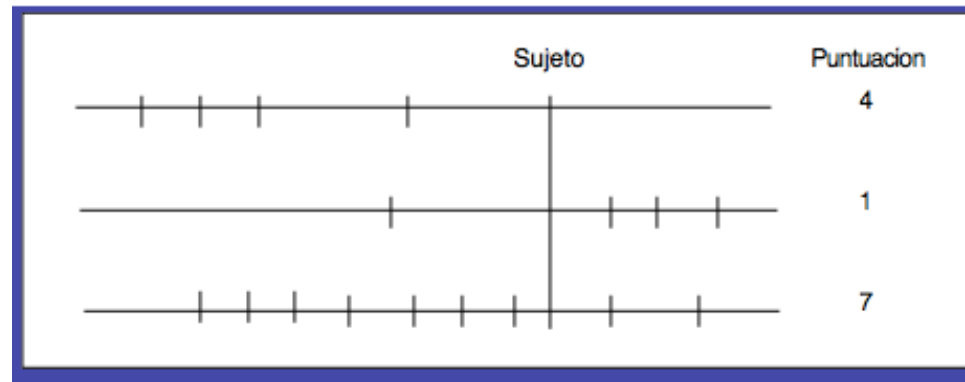
Precisão é feita pelas estimativas de grupo da variância do escore observado e dela a parte que corresponde à variância do escore verdadeiro.





Algumas limitações da TCT

- Pontuação das pessoas dependem do teste



- A capacidade das pessoas e dificuldades dos itens estão em escalas diferentes
- Não é possível fazer interpretações referenciadas nos itens
- A precisão do teste é um índice global quando na verdade varia em função do nível da escala que se pretende medir
- A partir dessa e de outras limitações é que se desenvolveu a TRI

Exemplos de escalas via TRI (Rasch)

EMBRETSON

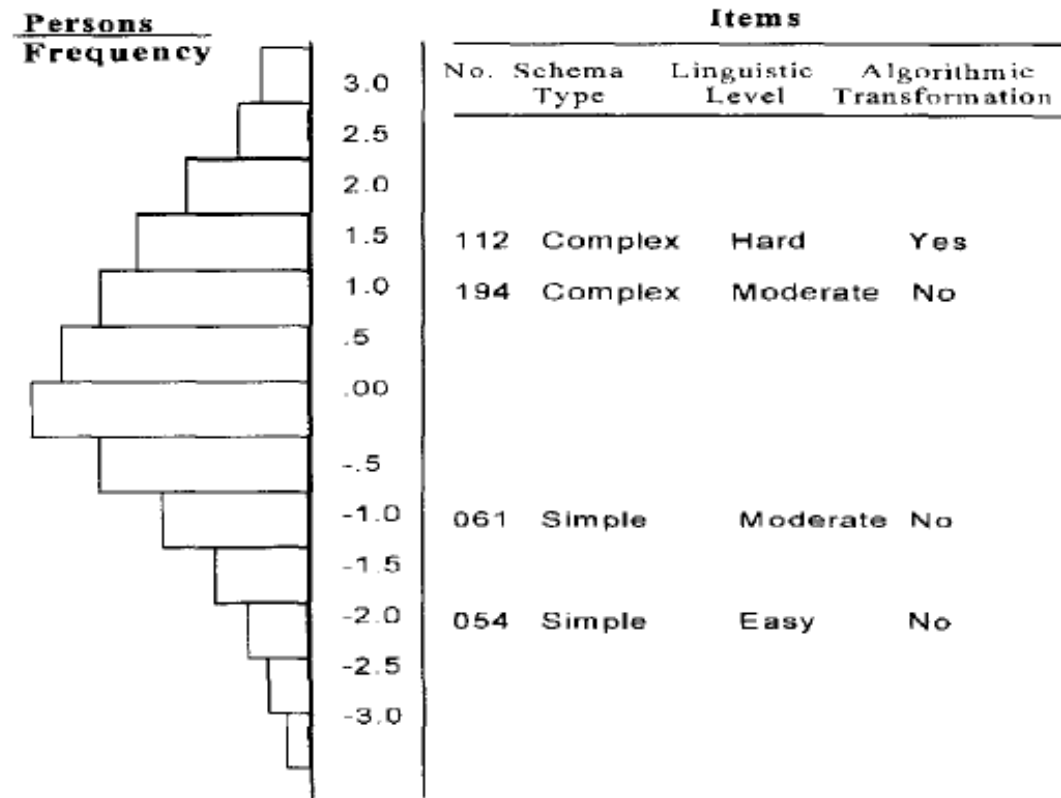
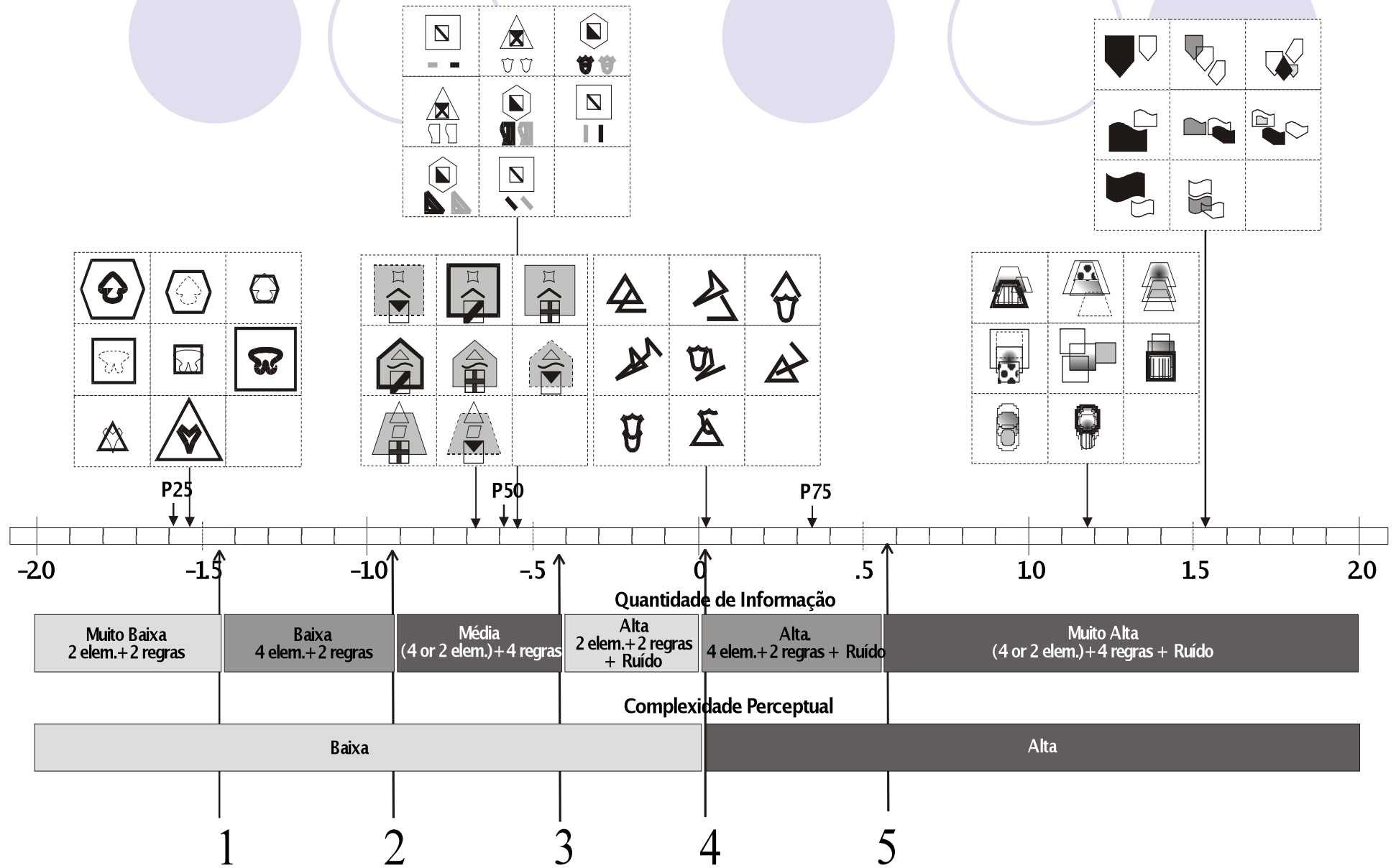


Figure 5. Common scale measurement of item difficulty and trait scores.

Primi (2002) Citado por Urbina (2004) pag. 118



	-6	-4	-2	0	2	4	6	ITEM
0				0	1	2		2 16 Auditory hallucinations
0				0	1	2		2 62 Holding ideas not yours
0				0	1	2		2 82 Fear passing out w others
0				0	1	2		2 47 Fear public transport
0				0	1	2		2 07 Feeling others direct yr thinking
0				0	1	2		2 25 Fear leave home by self
0				0	1	2		2 63 Impulse to hurt another
0				0	1	2		2 73 Ill-at-ease food/drink w others
0				0	1	2		2 13 Fear exposed places
0				0	1	2		2 35 Others know what you think
0				0	1	2		2 65 Do again: feel/number/clean
0				0	1	2		2 15 Thinking about causing own death
0				0	1	2		2 53 Trouble swallowing
0				0	1	2		2 48 Sense: not getting enough air
0				0	1	2		2 74 Recurrent quarrels
0				0	1	2		2 81 Yelling/tossing objects
0				0	1	2		2 75 Anxious when by oneself
0				0	1	2		2 17 Shakiness
0				0	1	2		2 23 Inexplicably frightened
0				0	1	2		2 12 Sharp/tight pangs in upper body
0				0	1	2		2 72 Attacks of intense anxiety
0				0	1	2		2 19 Lack of hunger
0				0	1	2		2 04 Woozy/light-headed
0				0	1	2		2 67 Impulse to destroy objects
0				0	1	2		2 84 Very disturbed by lewd ideas
0				0	1	2		2 24 Unmanageable anger episodes
0				0	1	2		2 86 Scary ideas or mental pictures
0				0	1	2		2 58 Extremities seem weighed down
0				0	1	2		2 50 Evading scary objects/places/events
0				0	1	2		2 85 Deserving discipline for doing evil
0				0	1	2		2 70 Discomfort in large groups of Ps
0				0	1	2		2 39 Overly rapid/strong heartbeat
0				0	1	2		2 49 Episodes: feel too warm/too cool
0				0	1	2		2 52 Loss of feeling/prickling
0				0	1	2		2 87 Thoughts of severe physical problems
0				0	1	2		2 08 Attributing problems to other Ps

Severe
Agora phobia

Serious
Psychosis

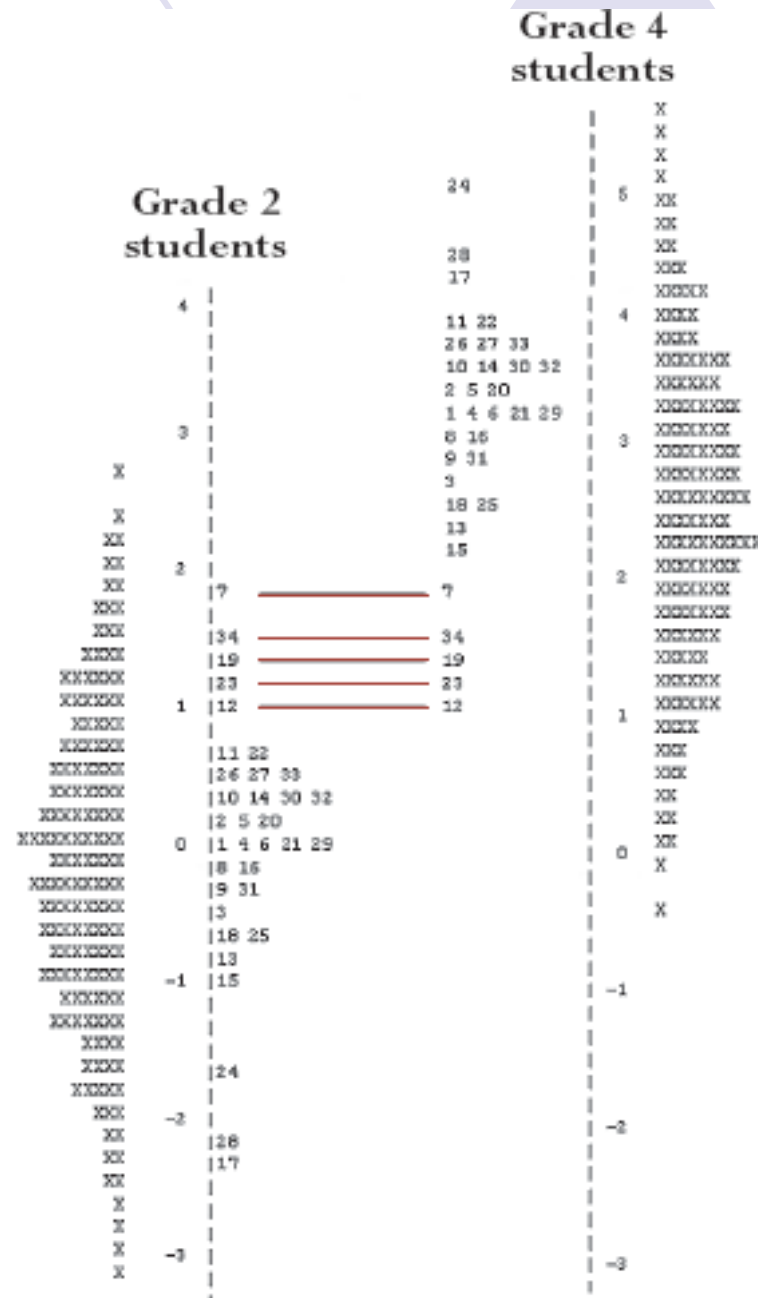
Aggression

Panic/
strong
anxiety

Serious
medical
concerns

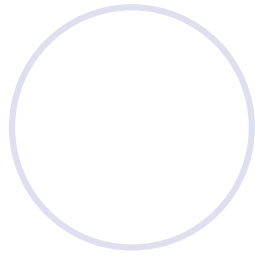
0	0	:	1	:	2	2	58	Extremities seem weighed down
0	0	:	1	:	2	2	50	Evading scary objects/places/events
0	0	:	1	:	2	2	85	Deserving discipline for doing evil
0	0	:	1	:	2	2	70	Discomfort in large groups of Ps
0	0	:	1	:	2	2	39	Overly rapid/strong heartbeat
0	0	:	1	:	2	2	49	Episodes: feel too warm/too cool
0	0	:	1	:	2	2	52	Loss of feeling/prickling
0	0	:	1	:	2	2	87	Thoughts of severe physical problems
0	0	:	1	:	2	2	08	Attributing problems to other Ps
0	0	:	1	:	2	2	20	Weeping readily
0	0	:	1	:	2	2	59	Ideas of passing away
0	0	:	1	:	2	2	40	Queasy or distressed belly
0	0	:	1	:	2	2	43	Ideas of reference
0	0	:	1	:	2	2	56	Sense: debility in physical sites
0	0	:	1	:	2	2	80	Sense: impending doom
0	0	:	1	:	2	2	90	Sense: mental dysfunction
0	0	:	1	:	2	2	78	Inability to stop moving
0	0	:	1	:	2	2	21	Uncomfortable/timid w other gender
0	0	:	1	:	2	2	22	Sense: Ensnared/captured
0	0	:	1	:	2	2	68	Holding unusual views/assumptions
0	0	:	1	:	2	2	37	View others as cold/rejecting
0	0	:	1	:	2	2	76	Accomplishments underappreciated
0	0	:	1	:	2	2	64	Waking up too soon
0	0	:	1	:	2	2	45	Inspect/reinspect everything
0	0	:	1	:	2	2	38	Overly methodical to avoid error
0	0	:	1	:	2	2	88	Always isolated from others
0	0	:	1	:	2	2	05	Lack: erotic attraction/enjoyment
0	0	:	1	:	2	2	61	Discomfort: being observed/discussed
0	0	:	1	:	2	2	18	Thinking: others unreliable
0	0	:	1	:	2	2	51	Mental lapses
0	0	:	1	:	2	2	33	Afraid
0	0	:	1	:	2	2	44	Insomnia
0	0	:	1	:	2	2	03	Persistent disagreeable ideas
0	0	:	1	:	2	2	69	Highly ill-at-ease around Ps

Precisão e métricas comuns

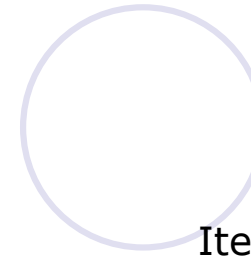


O que é medida ?

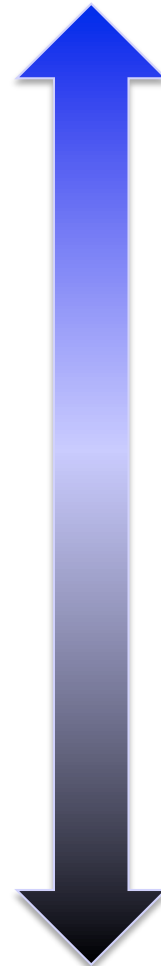
Pessoas



Escala/variável/construto



Itens



Modelo de medida (Wright & Stone, 1979)

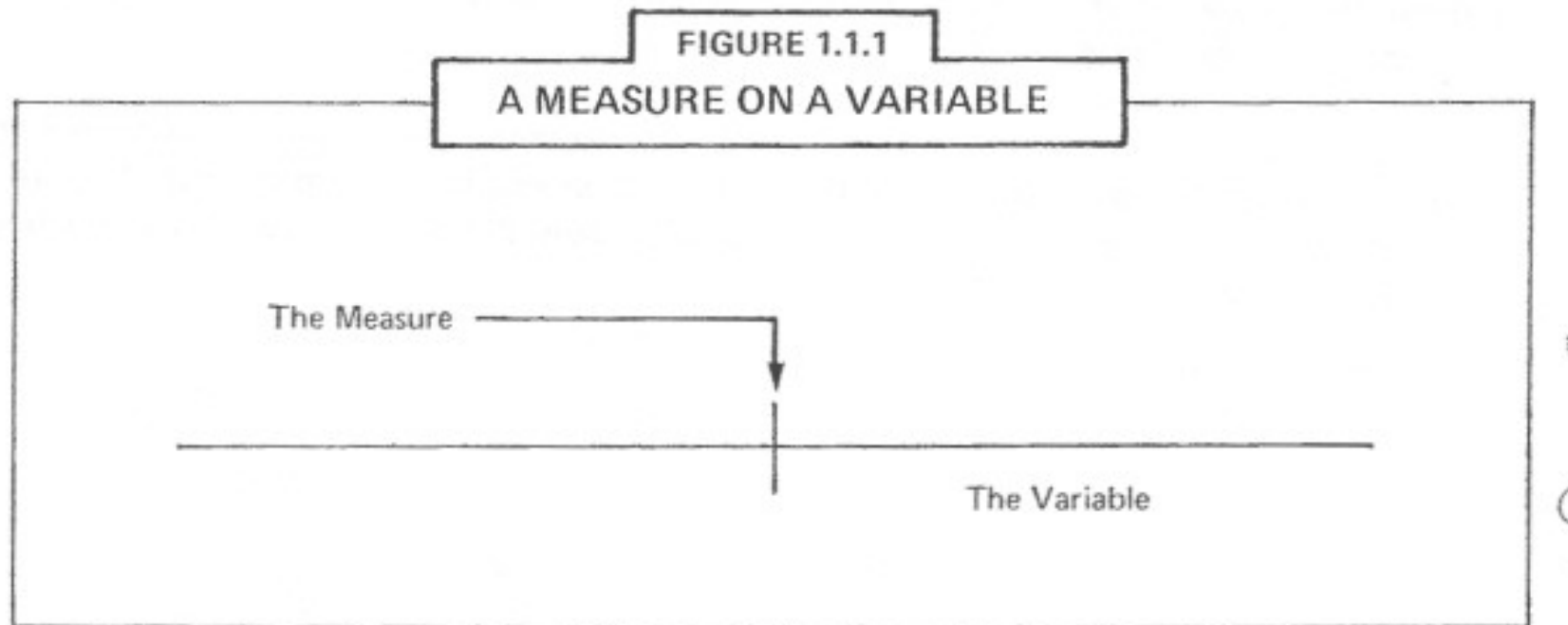


FIGURE 1.1.2
DEFINING A VARIABLE

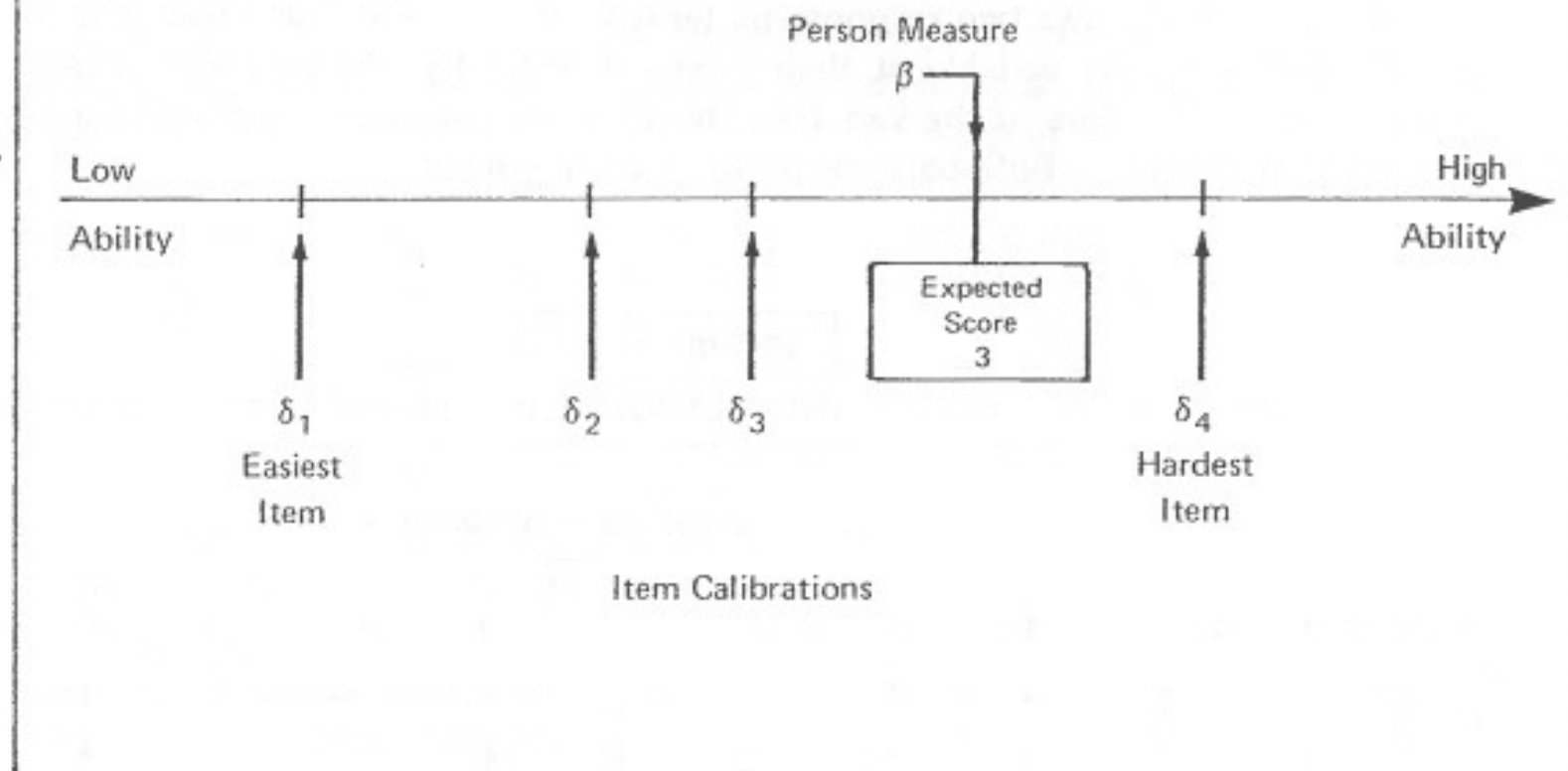
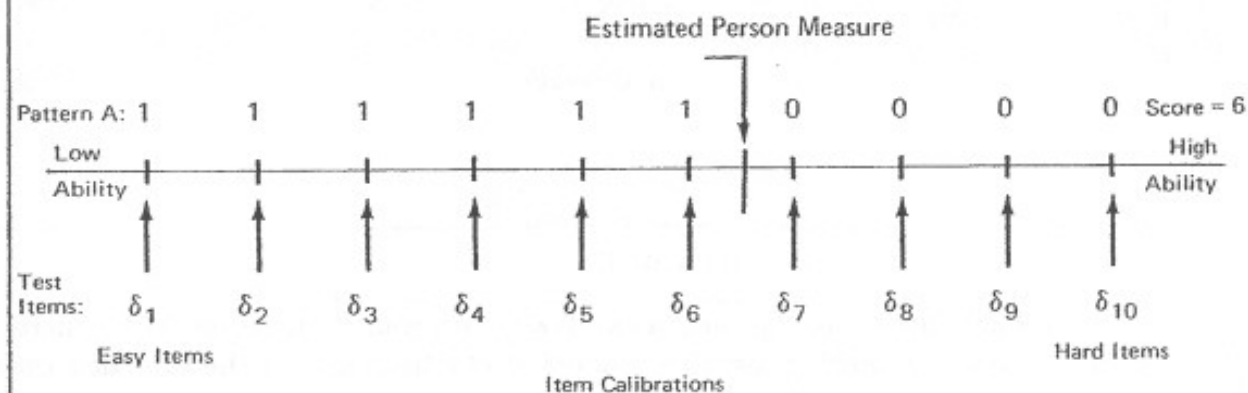
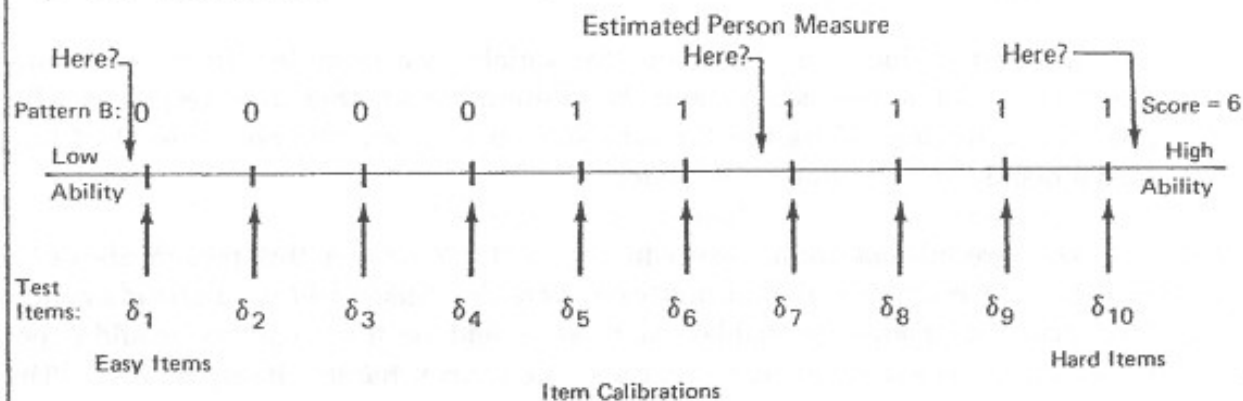


FIGURE 1.1.3
VALIDATING RESPONSE PATTERNS

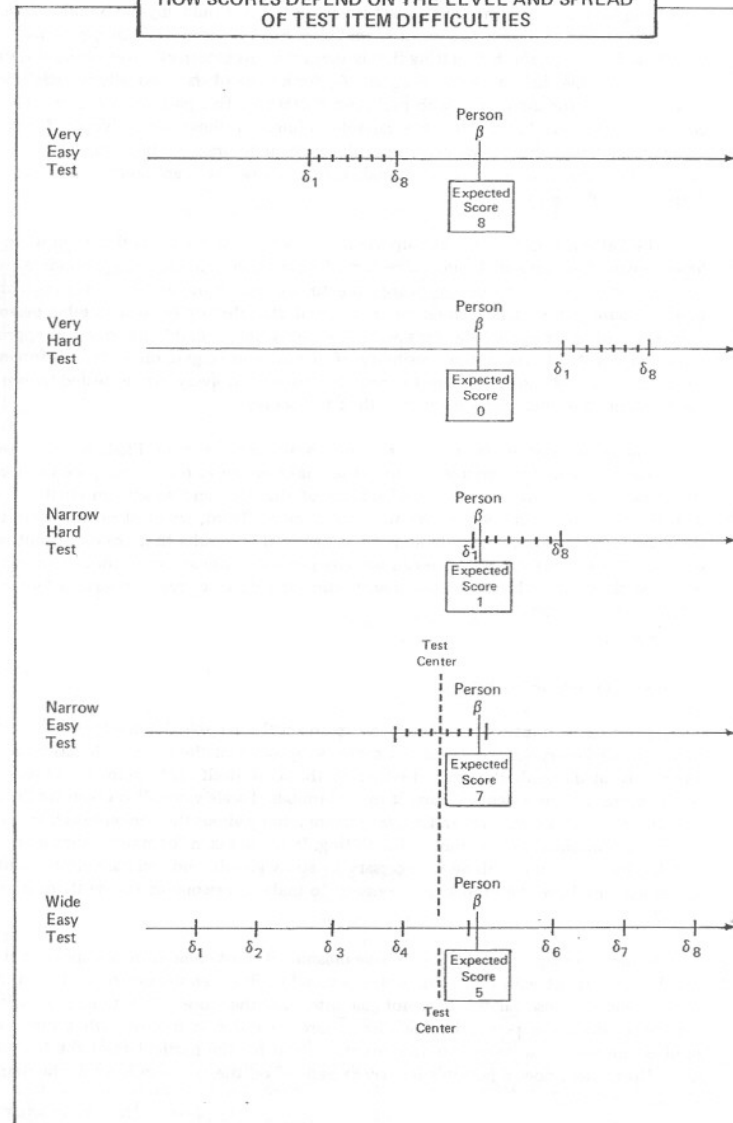


* No Reasonable measure can be derived



1 = Correct Response
0 = Incorrect Response
Score = Sum of Correct Responses

FIGURE 1.2.1
HOW SCORES DEPEND ON THE LEVEL AND SPREAD
OF TEST ITEM DIFFICULTIES



Person's abilities remain constant
↓
Score depends on its difficulty

encl

FIGURE 1.2.2
THE NONLINEARITY OF SCORES

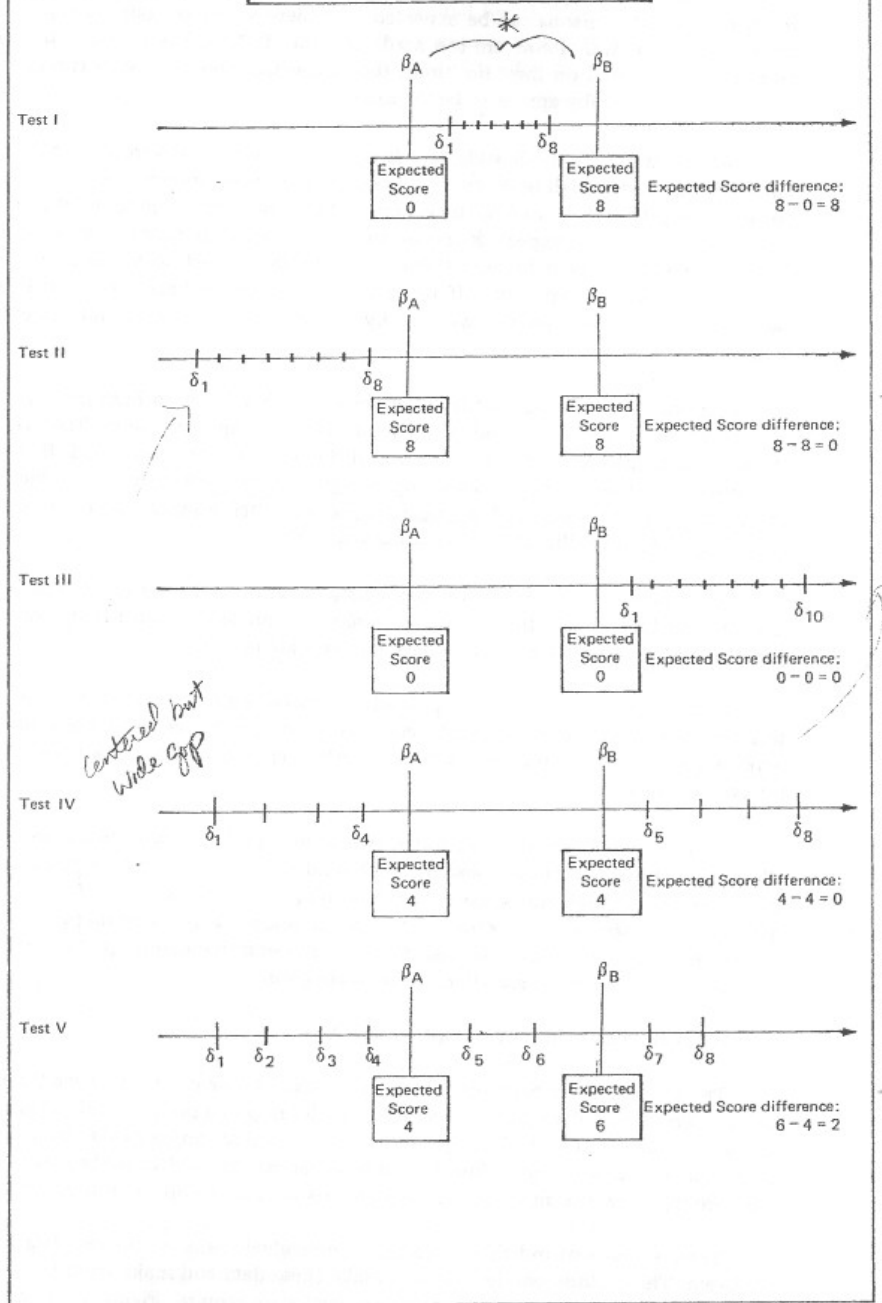
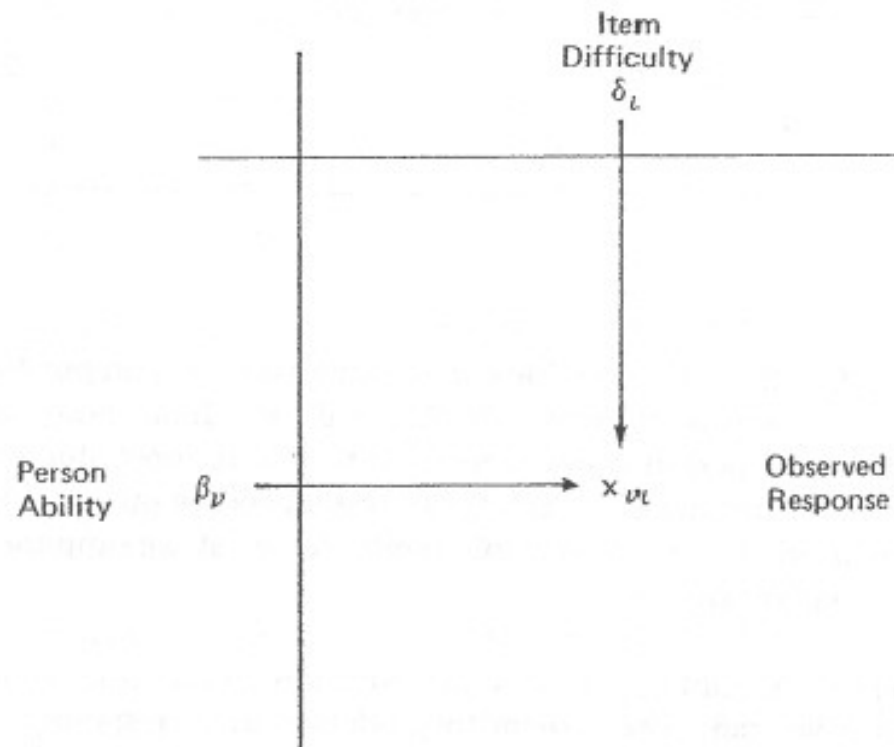


FIGURE 1.3.1
A DATA MATRIX
OF OBSERVED RESPONSES

Person Name	1	2	Item Name 3	4	5	Person Score
a	1	0	0	0	0	1
b	0	1	0	0	0	1
c	1	1	0	0	0	2
d	1	0	1	0	0	2
e	1	1	1	0	0	3
f	1	1	0	1	0	3
g	1	1	1	1	0	4
h	1	1	1	0	1	4
	7	6	4	2	1	20
			Item Score			

FIGURE 1.3.2

THE ESSENTIAL CONDITIONS CAUSING A RESPONSE



When the response is "correct",
 $x = 1$

When the response is "incorrect",
 $x = 0$

FIGURE 1.3.3

HOW DIFFERENCES BETWEEN
PERSON ABILITY AND ITEM DIFFICULTY OUGHT TO
AFFECT THE PROBABILITY OF A CORRECT RESPONSE

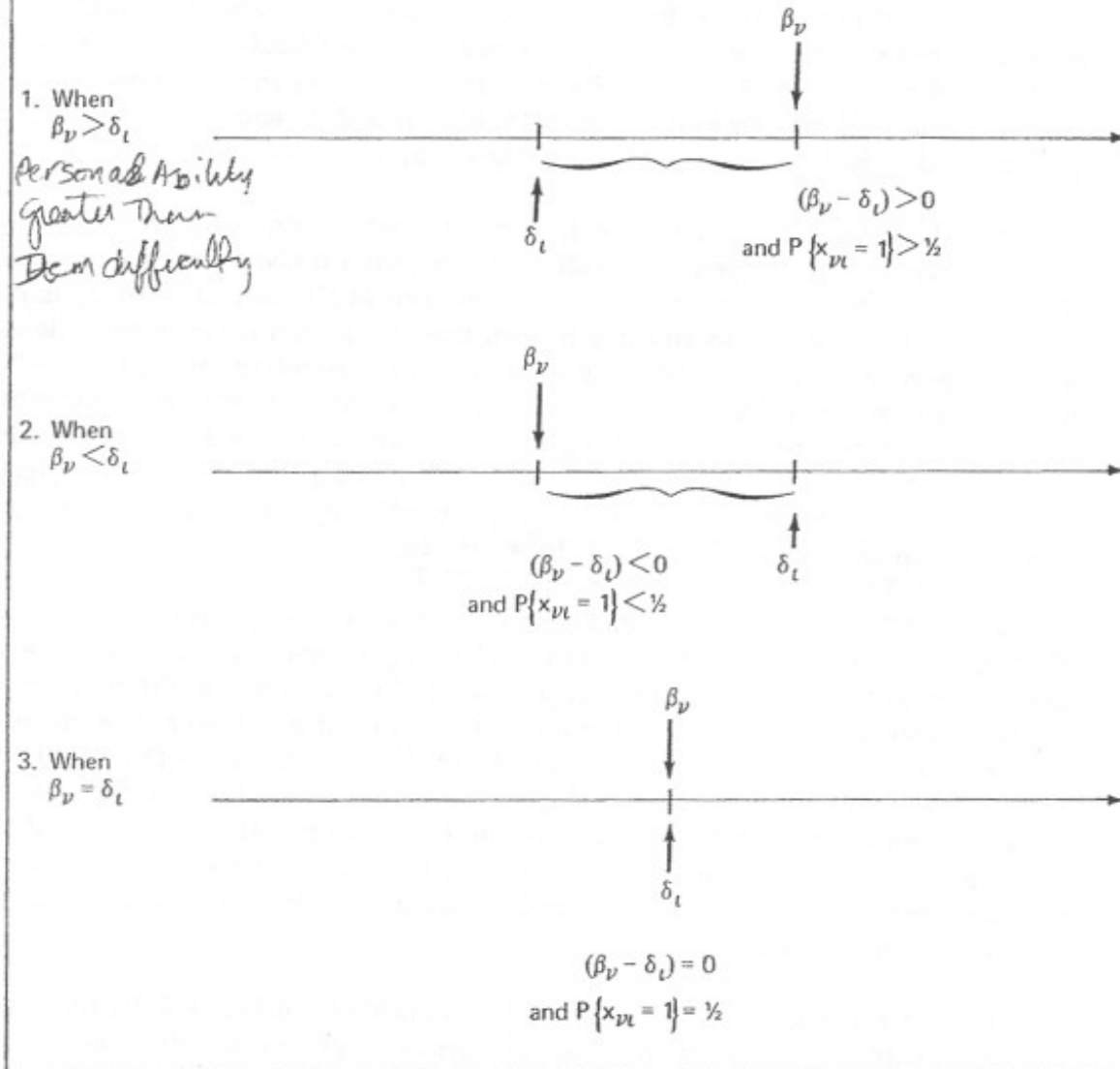
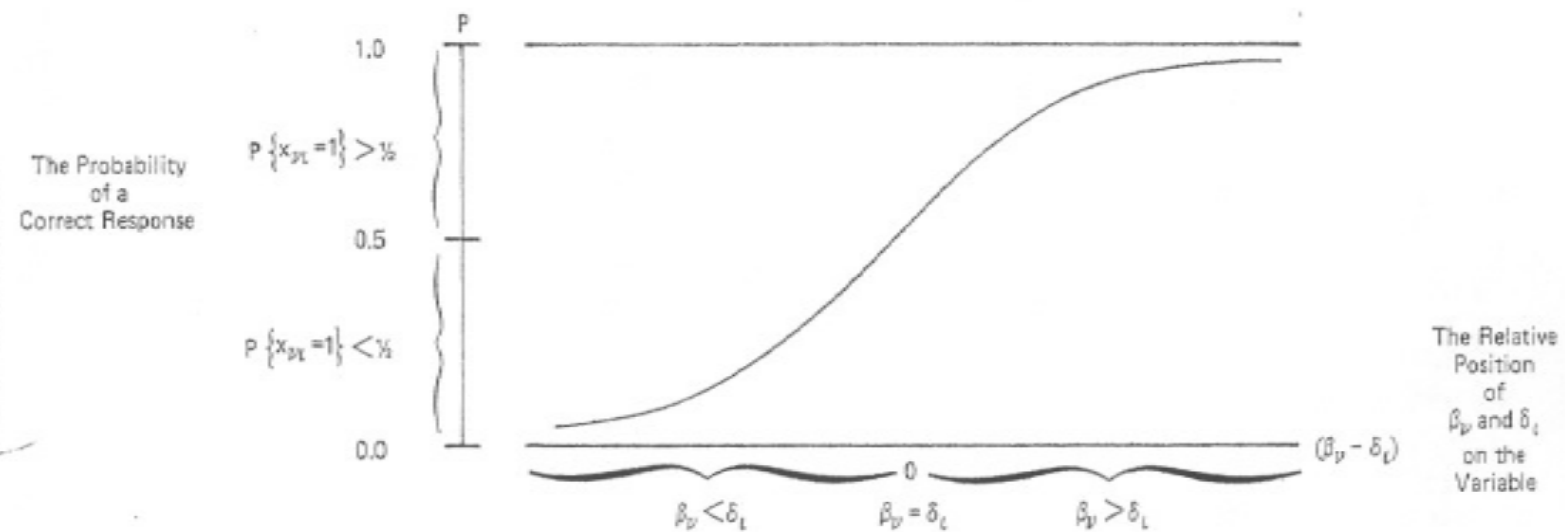


FIGURE 1.3.4
THE RESPONSE CURVE



$$P \{x_{ui} = 1 | \beta_u, \delta_i\} = f(\beta_u - \delta_i)$$

The Probability That The Observed Response is correct given The Ability of The Person and The difficulty of The item = (is a) Function of Ability minus Difficulty

TRI e Teoria Clássica dos Testes (TCT)

- Situação de testagem: pessoa respondendo a problemas ou perguntas (pessoa, item e respostas)

ord	Item 1	Item 2	Item 3	Item 4	Item 5	Total		
1 Suj 2	1	1	1	1	1	5	Soma	36
2 Suj 5	1	1	1	1	0	4	Média	2,57
3 Suj 1	1	1	1	1	1	5	Var	3,10
4 Suj 11	1	0	0	0	0	1	DP	1,76
5 Suj 3	1	1	1	1	0	4	Alfa	
6 Suj 13	0	0	0	0	0	0	N/N-1	1,08
7 Suj 10	1	0	0	0	0	1		
8 Suj 12	1	0	0	0	0	1	Alfa/KR	0,71
9 Suj 9	1	0	0	0	0	1		
10 Suj 7	1	1	1	0	0	3		
11 Suj 8	1	1	1	0	0	3		
12 Suj 4	1	1	1	1	0	4		
12 Suj 14	0	0	0	0	0	0		
13 Suj 6	1	1	1	1	0	4		
ID	0,86	0,57	0,57	0,43	0,14	2,57		
Corr It-Tot	0,60	0,94	0,94	0,87	0,56	1,00		
Desv. Padr	0,35	0,49	0,49	0,49	0,35	Soma:	2,18	
Var	0,13	0,26	0,26	0,26	0,13	Soma:	1,05	

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1,00				
Item 2	0,47	1,00			
Item 3	0,47	1,00	1,00		
Item 4	0,35	0,75	0,75	1,00	
Item 5	0,17	0,35	0,35	0,47	1,00

TRI e TCT

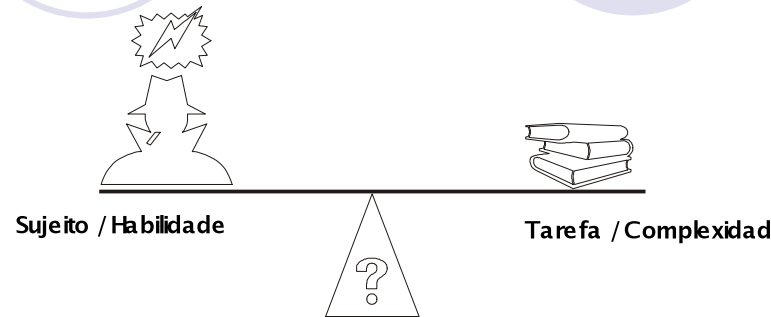
- Ordenando pelos escores e pelas pessoas

	Item 1	Item 2	Item 3	Item 4	Item 5	Total		
Suj 1	1	1	1	1	1	5	Soma	36
Suj 2	1	1	1	1	1	5	Média	2,57
Suj 3	1	1	1	1	0	4	Var	3,10
Suj 4	1	1	1	1	0	4	DP	1,76
Suj 5	1	1	1	1	0	4	Alfa	
Suj 6	1	1	1	1	0	4	N/N-1	1,08
Suj 7	1	1	1	0	0	3		
Suj 8	1	1	1	0	0	3	Alfa/KR	0,71
Suj 9	1	0	0	0	0	1		
Suj 10	1	0	0	0	0	1		
Suj 11	1	0	0	0	0	1		
Suj 12	1	0	0	0	0	1		
Suj 13	0	0	0	0	0	0		
Suj 14	0	0	0	0	0	0		
ID	0,86	0,57	0,57	0,43	0,14	2,57		
Corr It-Tot	0,60	0,94	0,94	0,87	0,56	1,00		
Desv. Padr	0,35	0,49	0,49	0,49	0,35	Soma:		2,18
Var	0,13	0,26	0,26	0,26	0,13	Soma:		1,05

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1,00				
Item 2	0,47	1,00			
Item 3	0,47	1,00	1,00		
Item 4	0,35	0,75	0,75	1,00	
Item 5	0,17	0,35	0,35	0,47	1,00

O que é TRI

- É um modelo matemático que representa os aspectos essenciais da situação de avaliação.

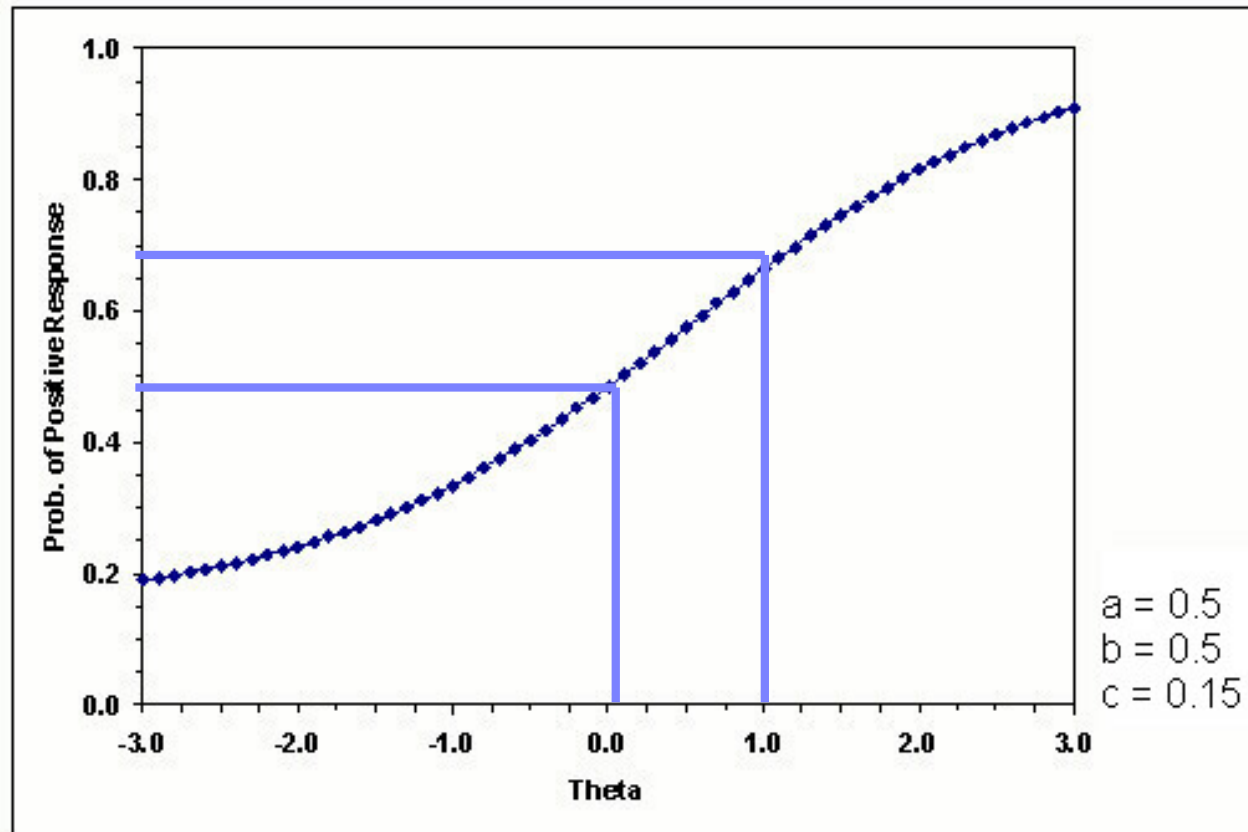


- Ele representa essencialmente três elementos dessa situação: (a) a habilidade da pessoa, (b) a dificuldade do problema e, a partir da interação entre elas, (c) a previsão do acerto ou erro.
- A implicação mais importante e interessante é que esse modelo mudou a *regra de comparação*.
- Os escores são atribuídos por um processo de comparação entre Sujeito/Habilidade – Tarefa/Dificuldade/Discriminação/Acertos ao acaso.
- *Conseqüência* Se existe um estudo sistemático dos fatores de complexidade caracterizando melhor a tarefa, atrela-se um significado psicológico aos diferentes níveis de complexidade do problema. Como a mensuração atribui números comparando o desempenho da pessoa com a escala de complexidade do problema a medida passa a ter um significado psicológico direto já que as escalas de complexidade e habilidade são as mesmas.

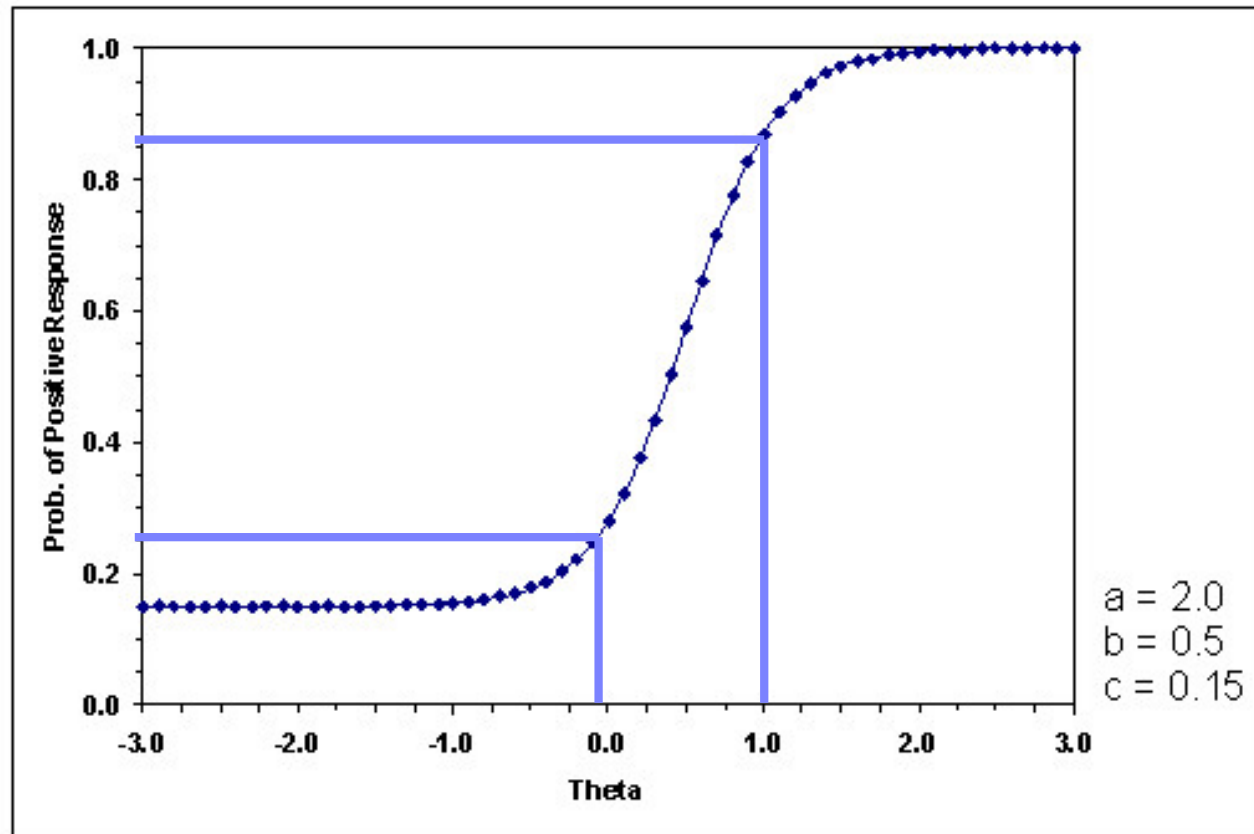
Primeira apresentação da TRI: Elementos (parâmetros) do Modelo e Representação Gráfica

- O que é um modelo
 - “Um modelo é uma representação ou interpretação simplificada da realidade, ou uma interpretação de um fragmento de um sistema segundo uma estrutura de conceitos” ... “Em Teoria de modelos um modelo é uma estrutura composta por um conjunto universo e por constantes, relações e funções definidas no conjunto universo” (<http://pt.wikipedia.org/wiki/>)
 - “A mathematical model is an abstract model that uses mathematical language to describe the behaviour of a system”
- Modelo TRI de três parâmetros:
 - Pessoas
 - Theta: dimensão mensurada (-4 a + 4)
 - Itens
 - a: discriminação (0,6 a 1,8)
 - b: dificuldade (-4 a +4)
 - c: acerto ao acaso (“chute”) (0 a 0,40)
 - Efeito/Resultado/“Outcome”
 - P: probabilidade de acerto (0 a 1) ... Estranho não ...
 - P (theta) dado a, b e c

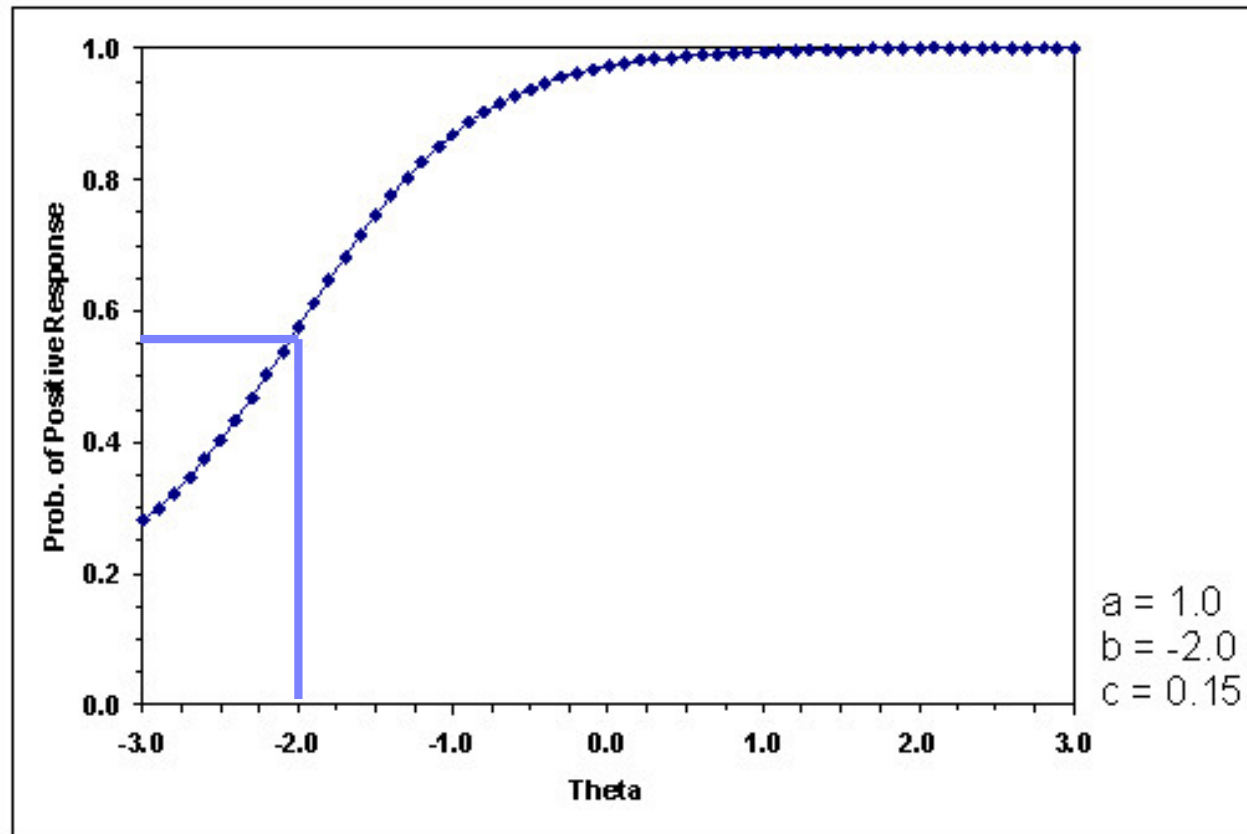
O “Modelo”: uma aproximação gráfica (baixo a)



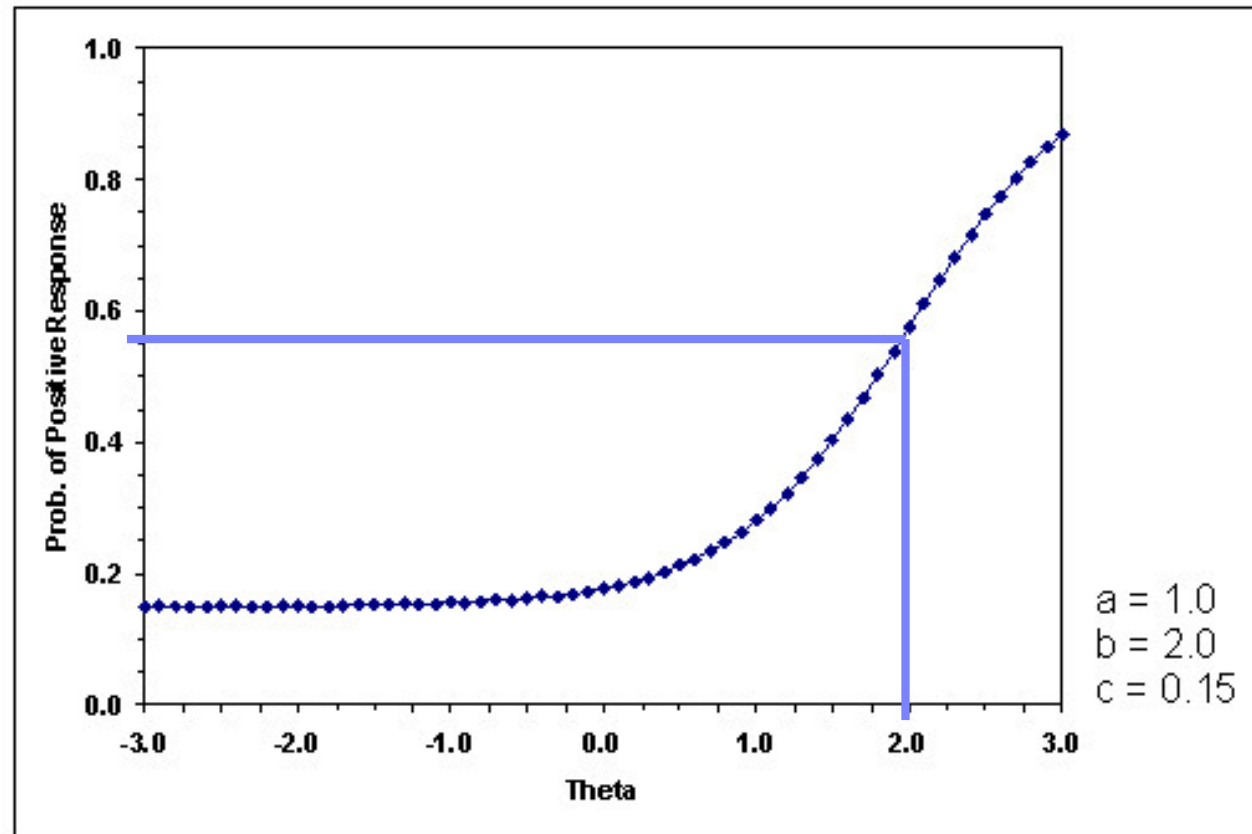
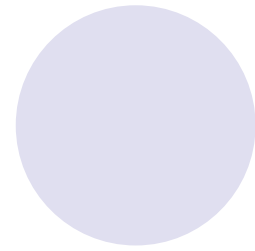
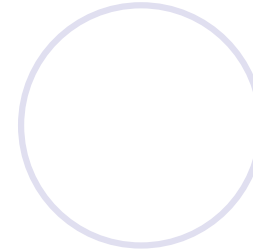
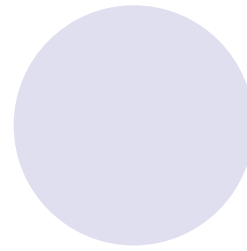
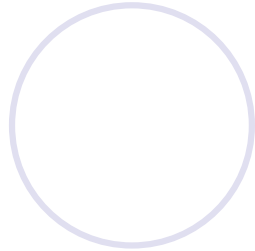
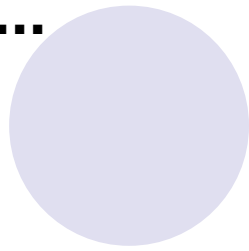
Alta discriminação ... a



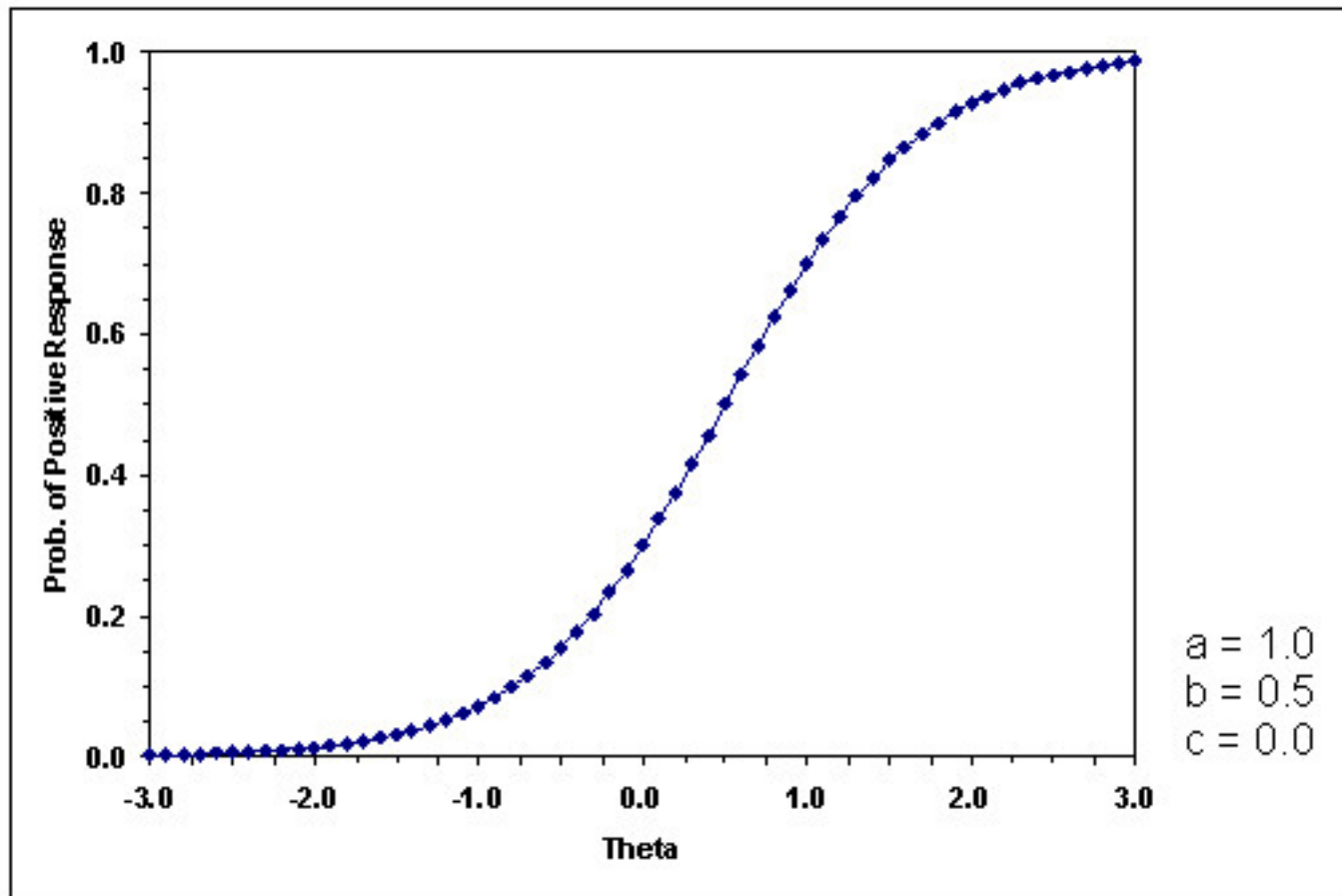
Parâmetro b ...



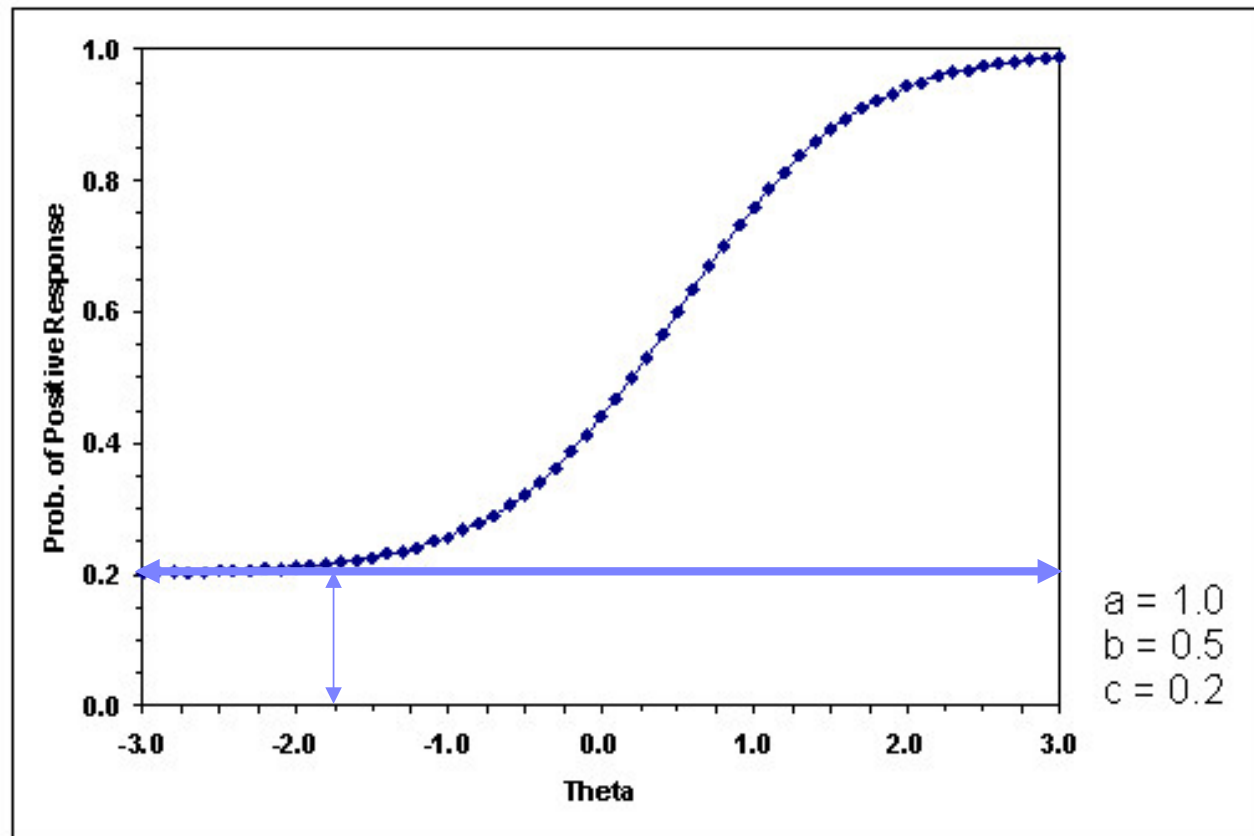
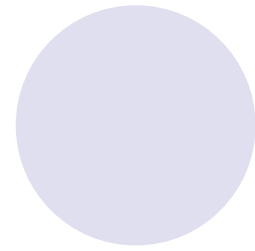
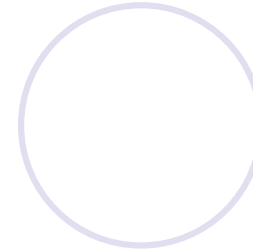
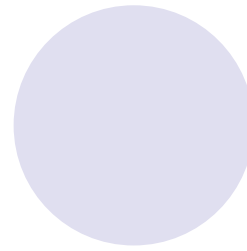
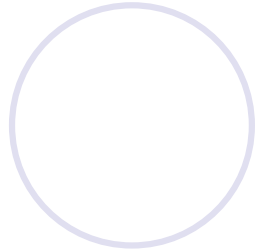
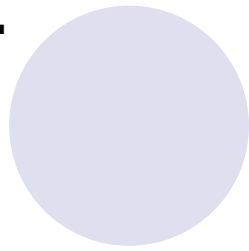
B alto



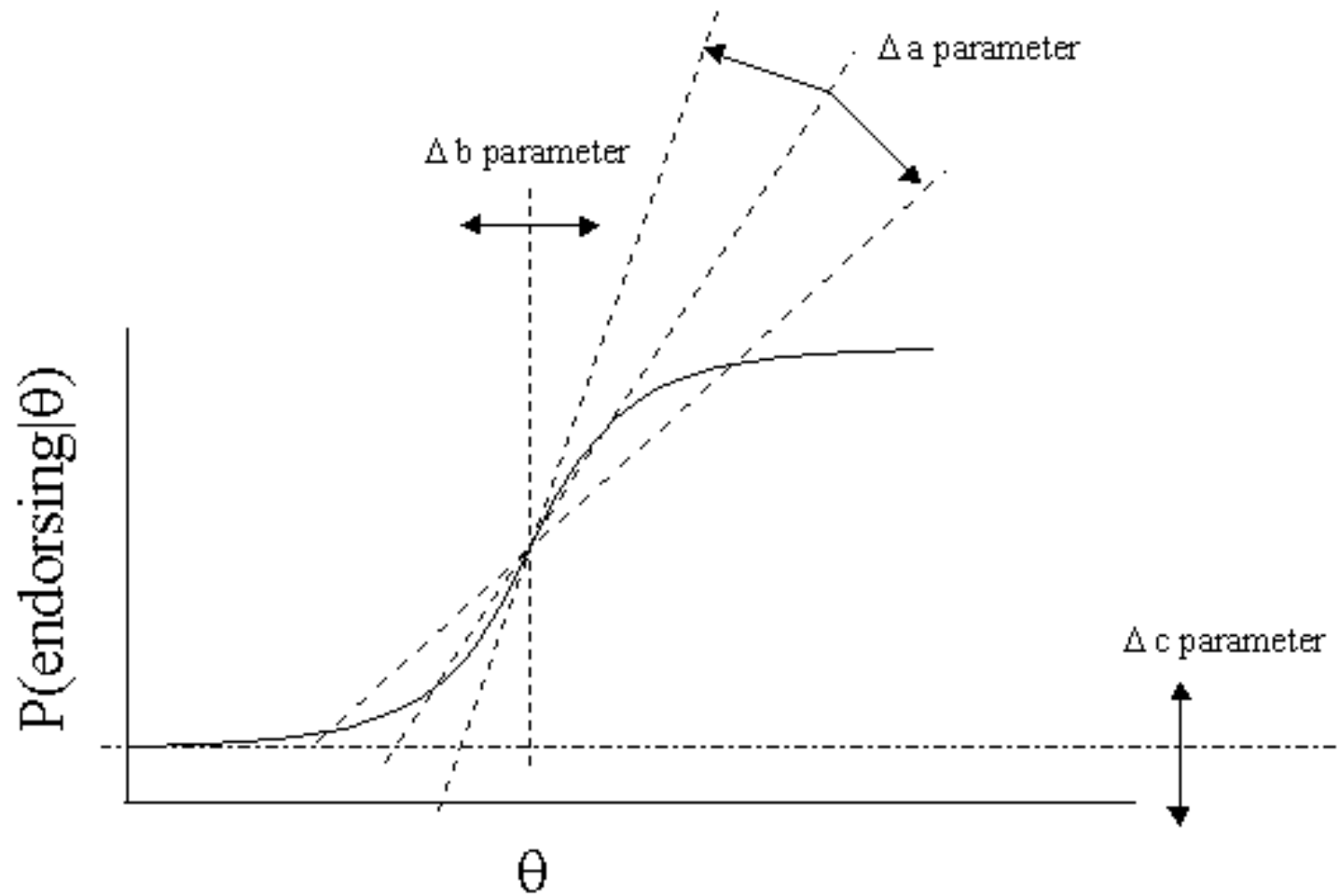
Parâmetro c ($= 0$)



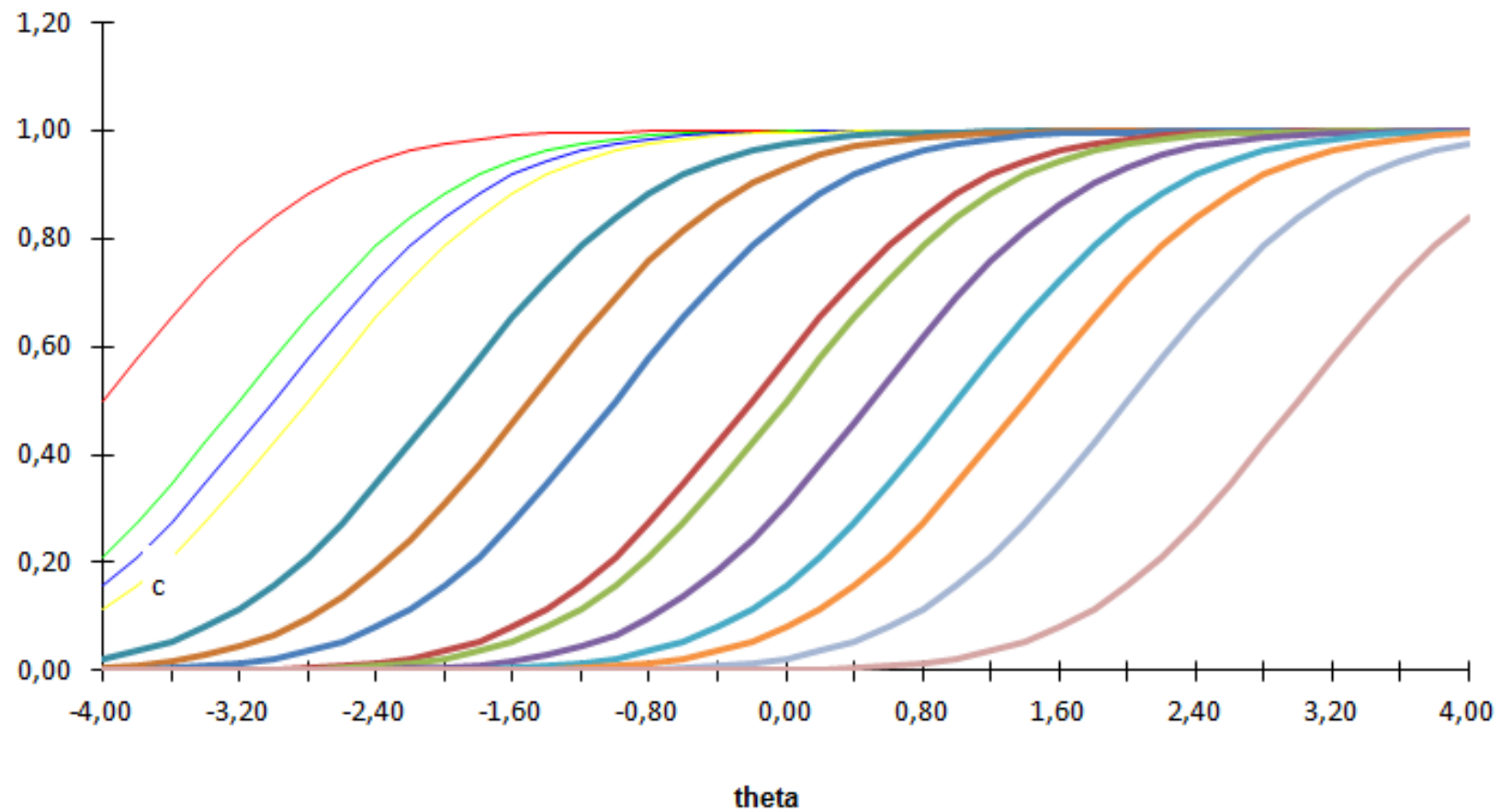
$c > 0 \dots$



Abstraindo



Como fica representado um teste ?



Item 1 Item 2 Item 3 Item 4 Item 5 Item 6 Item 7
Item 8 Item 9 Item 10 Item 11 Item 12 Item 13 Item 14

Matriz de dado clássica para representação via CCI



- TRI: Modelo matemático
 - Fórmulas e parâmetros
 - Dados (observação)
 - Calibração (resolução das equações, estimação dos parâmetros do modelo)
 - Controle de qualidade: cálculo dos resíduos (quanto a representação corresponde à observação)
- Como organizo os dados para calcular os parâmetros do modelo ?
 - Planilha do GfRI
 - Insight sobre a transição do modelo clássico para a TRI a partir dos dados observados
 - Análise do PAC
 - Análise gráfica do Item, insight entre o modelo e a observação

Exemplo das limitações da Psicometria Clássica

- Lembrando dos conceitos vistos na disciplina de Psicometria
- Modelo:

$$X_i = T_i + e_i$$

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

$$\rho_{tt} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_e^2} = \frac{\sigma_t^2}{\sigma_x^2}$$

- Então na análise dos itens procuramos:
 - Itens com variância
 - Itens com covariância com os outros
- Variância: índice de dificuldade (ID) ou média e desvio padrão em escalas likert
- Correlação: correlação ponto bisserial item-total
- Limitações
 - Parâmetros dos itens e sujeitos dependentes da amostra
 - Precisão da medida é "genérica"
- Planilha X Exemplos