



GOVERNO DE  
PORTUGAL

MINISTÉRIO DA EDUCAÇÃO  
E CIÊNCIA

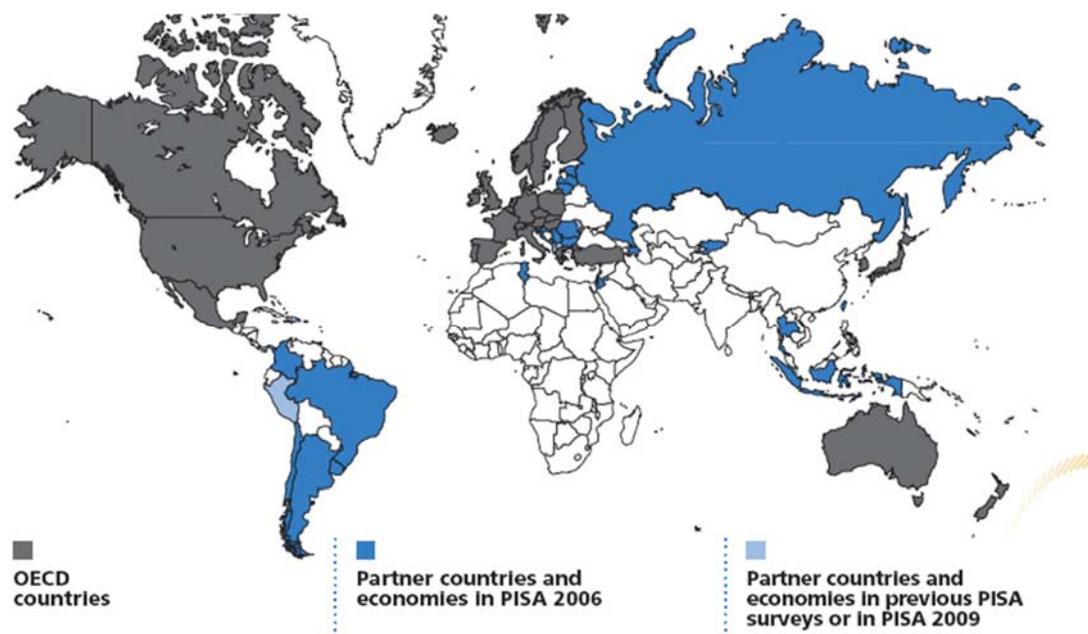
**IAVE** INSTITUTO  
DE AVALIAÇÃO  
EDUCATIVA, I.P.

# PISA 2015

... da “*Avaliação papel-e-lápis*” à  
“*Avaliação em Computador*” ...

João Marôco, Ph. D.

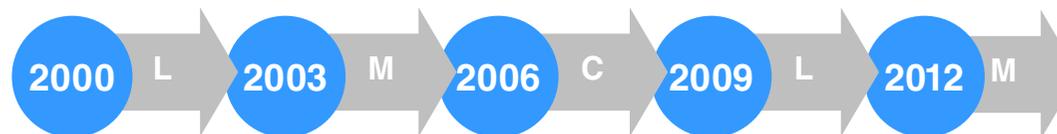
## Introdução



Os testes do PISA, iniciados em 2000, aplicados desde então com periodicidade trienal, avaliam os sistemas educativos de vários países de todo o mundo...

... por estimativa da capacidade que alunos de 15 anos têm para mobilizar conhecimentos e competências em Leitura, Ciências e Matemática para resolver problemas do dia-a-dia, fundamentais à plena cidadania nas sociedades contemporâneas.

## Introdução



**2000 - 2012:** formato tradicional “papel-e-lápis” ...



**2015 - :** formato “Computer Based Assesment” ...

## Introdução



**2015** - : formato “Computer Based Assesment”...

A avaliação em computador tem várias vantagens:

- Logística
- Codificação automática
- Fiabilidade de resultados
- Disponibilidade de Resultados
- Itens interativos e simulação de dados
- ...
- Custos financeiros

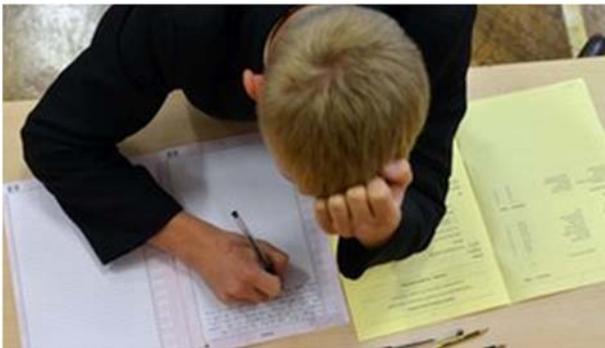
**OCDE:** Evolução do PISA de PBA ⇒ **CBA** ⇒ **CAT**



## Problemas com o CBA?

Invariância de modo?

Como “ligar” os resultados dos ciclos PBA vs. CBA?



- Testes CBA têm validade e fiabilidade?
- Os itens/testes em PBA vs. CBA:
  - Têm o mesmo *índice de dificuldade*?
  - Têm o mesmo *índice de discriminação*?



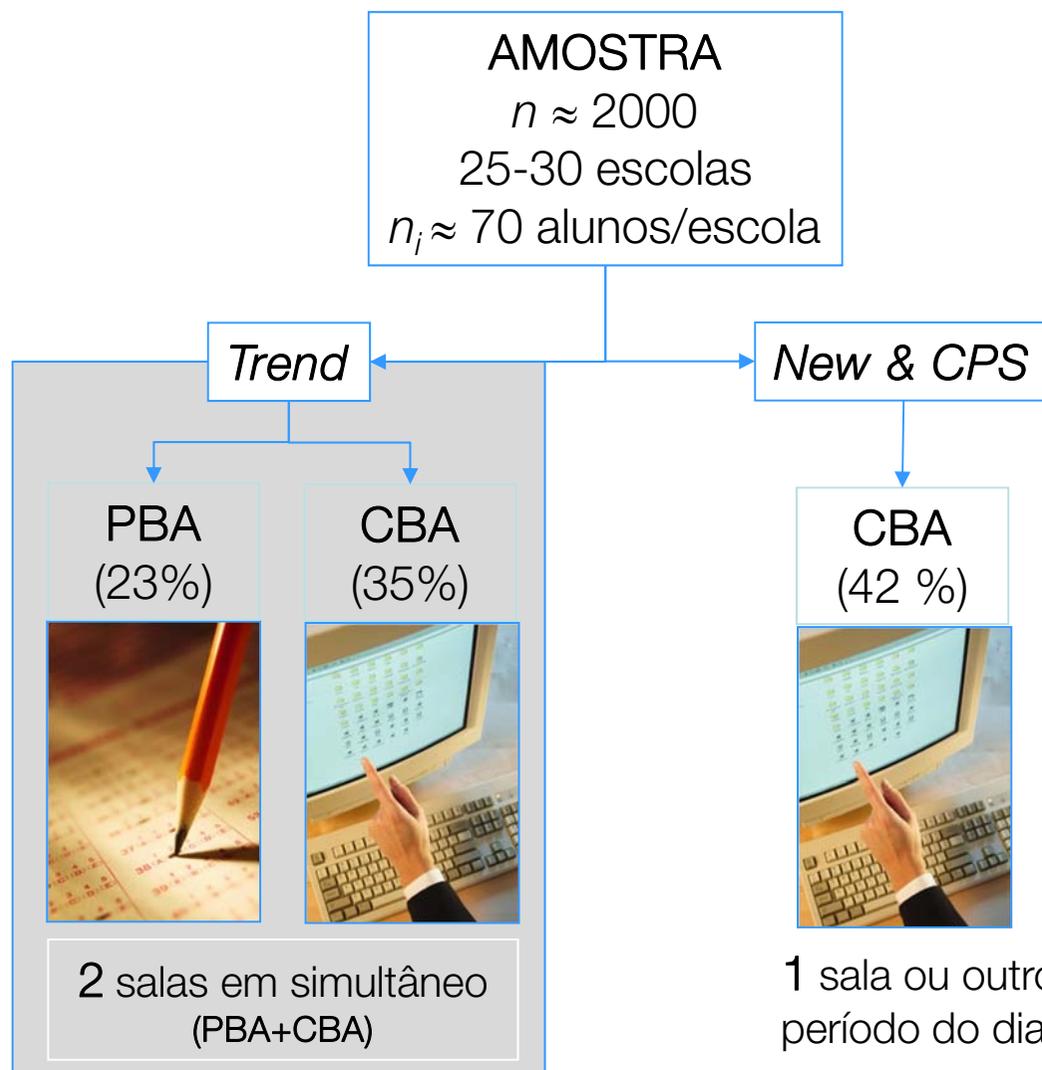
- Existe efeito de “modo” nos resultados e comparação de países PBA vs. CBA?
- É possível “ligar” os resultados (itens *trend*) dos ciclos PBA com os novos ciclos CBA?

# Piloto PISA 2015

## Design

(Preferencial)

Domínio principal: Ciências  
72 Combinações L, M & C





## Piloto PISA 2015

### Resultados Internacionais

- *Base de Dados da OECD*: não é do domínio público
- *Teste Piloto PISA 2015 dos 45 Países/territórios que realizaram CBA & PBA*
- 318 itens de Leitura (*trend*), Matemática (*trend*) e Ciências (*trend & new*)
- $n = 120\ 303$  alunos com 15 anos (53 Países/territórios no Piloto)
- ETS (USA): *Outorgante* responsável pelo “scaling”



[von Davier & Sakamoto (ETS), Pers. Com., Novembro 2014]

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

## Piloto PISA 2015

### Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

*Taxa de Não Respostas* (318 itens, aplicados em 45 países com CBA & PBA):

2015	Média Internacional			2015		
Formato	CBA	PBA	CBA - PBA	Formato	CBA	PBA
% não-respostas	4,7	5,2	-0,5	Nº itens (média/país) com <i>Ef. Om.</i> elevado	32,5	44,5
<i>Efeito de Omissão</i>	2,0	2,6	-0,6			

*Efeito de Omissão:*

$P^+$  (calculados com 'não respostas' como errados) –  $P^+$  (não considerando as 'não respostas')

Efeito de Omissão forte: > 5%

( $P^+$ : proporção de respostas corretas; ou média relativa  $[M/\text{maxscore}]$  dos scores nos itens politómicos)

**CBA: reduz ligeiramente a % não respostas e efeito de omissão.**

## Piloto PISA 2015

### Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

#### Teoria Clássica de Testes:

**$P^+$** : proporção de respostas corretas; ou média relativa dos scores dos itens politómicos

Domínio	Médias Internacionais		
	CBA	PBA	CBA – PBA
Matemática	0,45	0,47	-0,02
Leitura	0,60	0,59	-0,01
Ciências- <i>Trend</i>	0,49	0,50	-0,01
Ciências- <i>New (FT Test)</i>	0,41	-	-

*... CBA ligeiramente mais difícil; diferenças de  $P^+$  irrelevantes entre os dois modos!*

## Piloto PISA 2015

### Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

#### Teoria Clássica de Testes:

**$P^+$**  Proporção de respostas corretas por posição dos blocos no teste (médias internacionais)

CBA (2015)	<i>Posição1</i>	<i>Posição2</i>	<i>Posição3</i>	<i>Posição4</i>	<i>Pos.4 – Pos.1</i>
Matemática	0,439	0,454	0,432	0,427	-0,012
Leitura	0,575	0,580	0,566	0,544	-0,031
Ciências	0,409	0,418	0,401	0,385	-0,024
PBA (2009)	<i>Posição1</i>	<i>Posição2</i>	<i>Posição3</i>	<i>Posição4</i>	<i>Pos.4 – Pos.1</i>
Matemática	0,411	0,401	0,384	0,371	-0,040
Leitura	0,581	0,557	0,532	0,499	-0,083
Ciências	0,490	0,478	0,457	0,435	-0,055

*CBA reduz efeito da localização dos itens no teste ...*

# Piloto PISA 2015

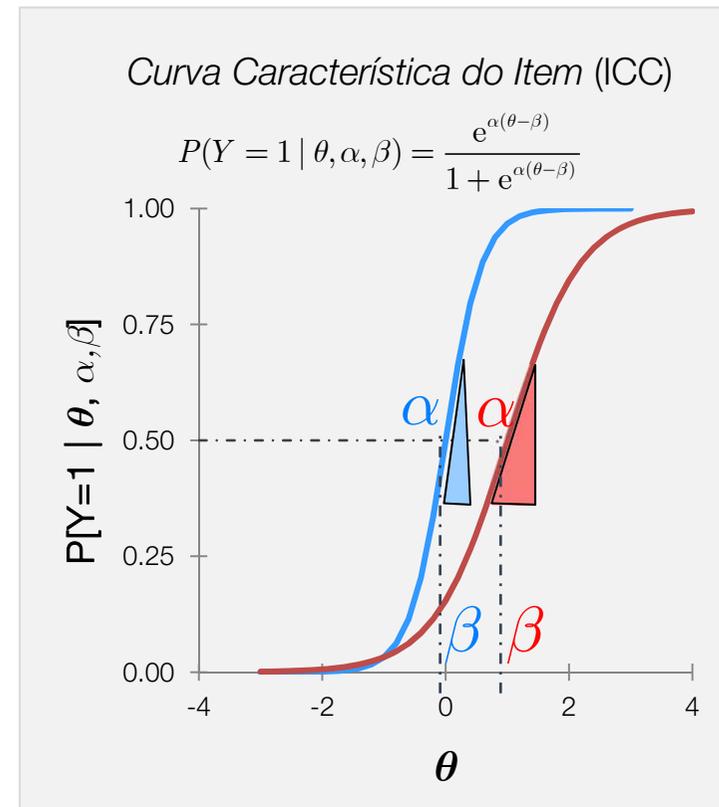
## Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

### Teoria de Resposta ao Item

No estudo Piloto do PISA2015 é seguida uma nova estratégia de “*scaling*”:

1. Avaliação da tendência histórica dos dados do PISA de 2000 a 2012
2. Teste a vários modelos de TRI:
  - 2.1. Modelos *Mixed Coefficients Multinomial Logit Model* (MCMLM): Modelação generalizada do **Modelo de Rasch** (itens dicotómicos) e **Crédito Parcial** (PCM) (itens politómicos) em múltiplos grupos (em vigor 2000-2012)
  - 2.2. Modelo de 2 parâmetros (2PL) e **Crédito Parcial Generalizado** (GPCM) *multigrupo* com parâmetro de interação entre grupos (países)



# Piloto PISA 2015

## Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

### Teoria de Resposta ao Item

#### Comparação entre os itens *trend* CBA 2015 e PBA 2000-2012

*mixed IRT models* (Rasch/PCM, 2PL/GPCM)

Domínio	RMSD > 0,2	Comparabilidade ICC	Fiabilidade Marginal
Matemática	2,2%	97,8%	0,83
Leitura	2,7%	97,3%	0,81
Ciências	4,9%	95,1%	0,80

$$RMSD = \sqrt{\int [P_o(\theta) - P_e(\theta)]^2 f(\theta) d\theta}$$

- Diferença quadrática entre  $P(Y=1|\theta)_{Obs} - P(Y=1|\theta)_{Exp}$  ponderada pela distribuição das aptidões;
- Reflete impacto dos índices de discriminação ( $a$ ) e dificuldade do item ( $b$ ) do item;
- RMSD > 0,2- Item com “*mau ajustamento*” relevante...

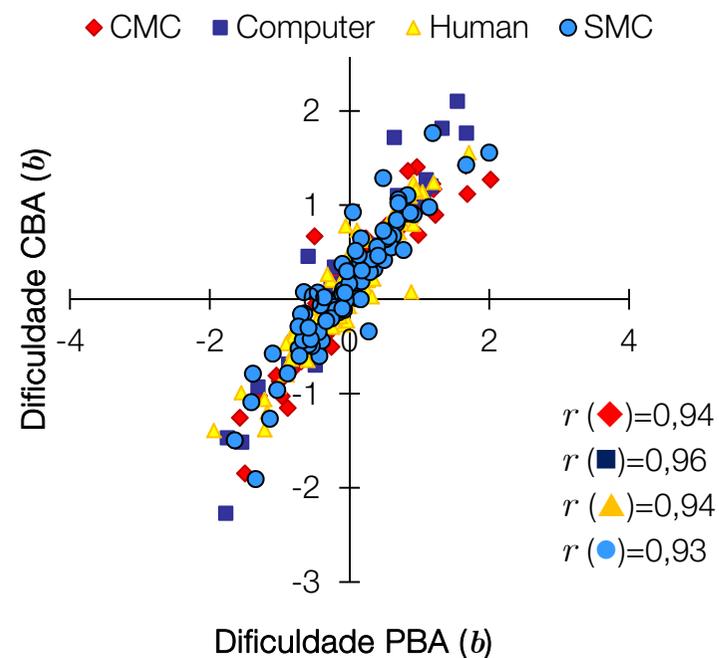
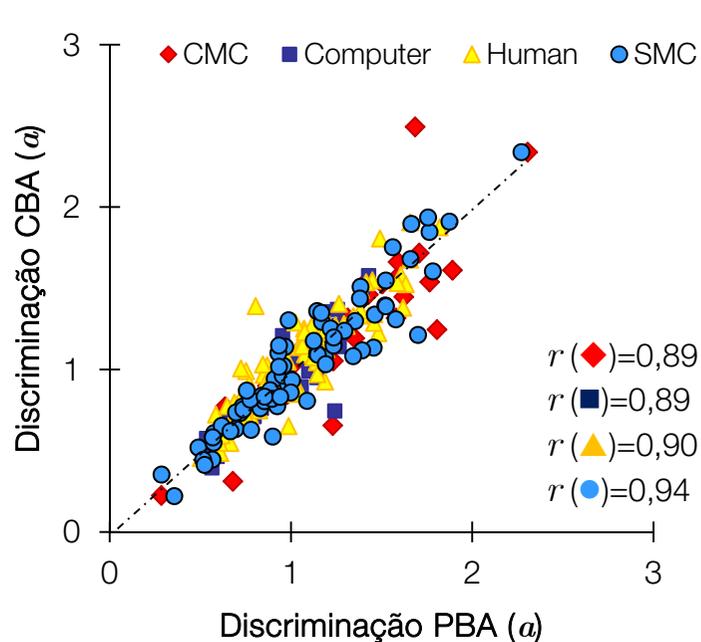
# Piloto PISA 2015

## Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

### Teoria de Resposta ao Item

### Comparação entre os itens PBA vs. CBA *mixed IRT models* (Rasch/PCM, 2PL/GPCM)



*Correlações fortes entre parâmetros de discriminação ( $a$ ) e dificuldade ( $b$ ) das ICC por tipologia de codificação do item...*

## Piloto PISA 2015

### Resultados Internacionais

Nota: A dimensão das amostras nacionais ( $n=1000-2000$ ) é demasiado pequena para inferências válidas por país.

#### Teoria de Resposta ao Item

#### Comparação entre os itens PBA vs. CBA

*mixed IRT models (Rasch/PCM, 2PL/GPCM)*

Domínio	Parâmetro de Discriminação $r(a_{PBA}, a_{CBA})$	Parâmetro de Dificuldade $r(b_{PBA}, b_{CBA})$
Matemática	0,91	0,95
Leitura	0,89	0,94
Ciências	0,94	0,93
Todos os domínios	0,91	0,94

*Correlações fortes entre parâmetros de discriminação ( $a$ ) e dificuldade ( $b$ ) das ICC por domínio...*



## Piloto PISA 2015

### Considerações finais...

- CBA reduz a taxa de não respostas
- Itens CBA ligeiramente mais difíceis que PBA (diferenças irrelevantes entre  $P^+$ )
- CBA reduz *Efeito de Omissão*
- CBA reduz *Efeito de Posição* dos itens/blocos no teste ...
- Comparabilidade dos itens *trend* entre 2000-2012 (PBA) e 2015 (CBA *trend*)
- Correlações entre os parâmetros de dificuldade e discriminação das ICC para PBA e CBA elevadas no geral, por tipologia de item e por domínio ( $r = 0,89-0,95$ )
- *Mixed multigroup IRT (RM/PCM; 2PL/GPCM)* permitem manter a continuidade da série PISA (2000-2015-...)
- Não há *efeito de modo* de aplicação (PBA vs. CBA) relevante quer de acordo com a *TCT* quer com a *TRI* (invariância de modo)
- Entre PBA e CBA são esperados 75-85% parâmetros comuns por domínio; *i. e.* ~ 4-5 *clusters* em 6 *clusters trend* comuns: uma ligação mais forte do que a observada em ciclos anteriores!



GOVERNO DE  
PORTUGAL

MINISTÉRIO DA EDUCAÇÃO  
E CIÊNCIA

**IAVE** INSTITUTO  
DE AVALIAÇÃO  
EDUCATIVA, I.P.

**Obrigado pela atenção!...**

joao.maroco@iave.pt