Detecting Item Bias

Bias Conference, Nov. 2000,

Houston TX

Maria Orlando, Ph.D.

RAND Corp.

Overview

- Item bias – differential item functioning
  - » DIF detection
    - – Classical approaches
    - – IRT approaches
- IRT Basics
- DIF detection with IRT
- Application
- Discussion

Item Bias

- Bias and Differential item functioning
- Subgroups
  - » Age, gender, ethnicity, education
  - » Mode or language of administration
- Impact - Non-equivalent scores
  - » Group differences may be misleading
  - » Associations among variables may be inaccurate

DIF Detection

- Classical approaches
  - » Contingency tables (M-H and extensions)
  - » Logistic regression
- IRT approaches
- Comparison
  - » Advantages
  - » Limitations

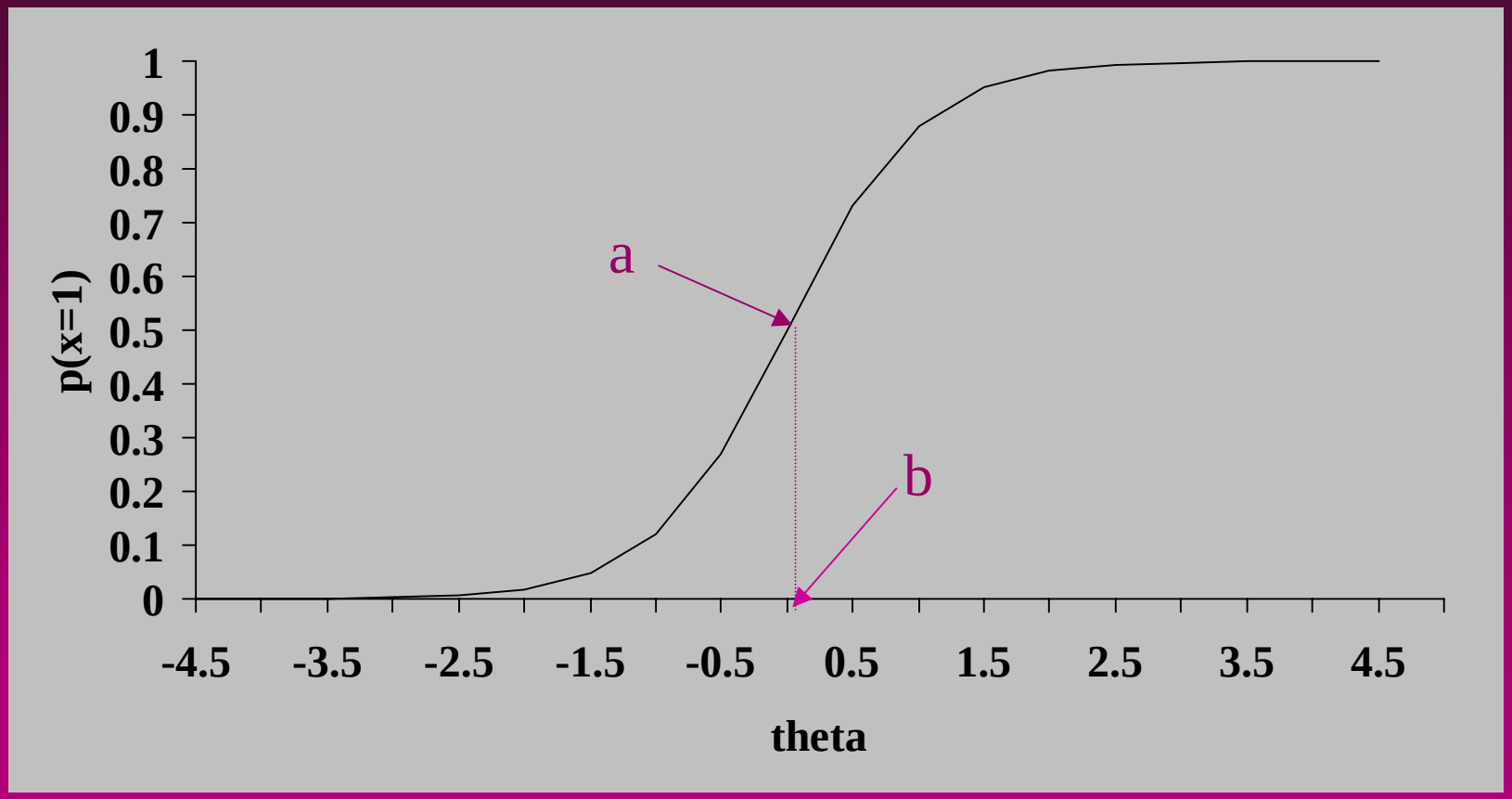
IRT Basics

- IRT describes the processes underlying a person's response to a question
- The Item Characteristic Curve, or Trace Line, most commonly defined by a logistic function, is at the core of IRT
- Item parameters reflect the properties of the item and determine the shape of the trace lines

IRT Models

- Dichotomous items (yes/no; true/false)
  - » 1-, 2-, and 3-PL models

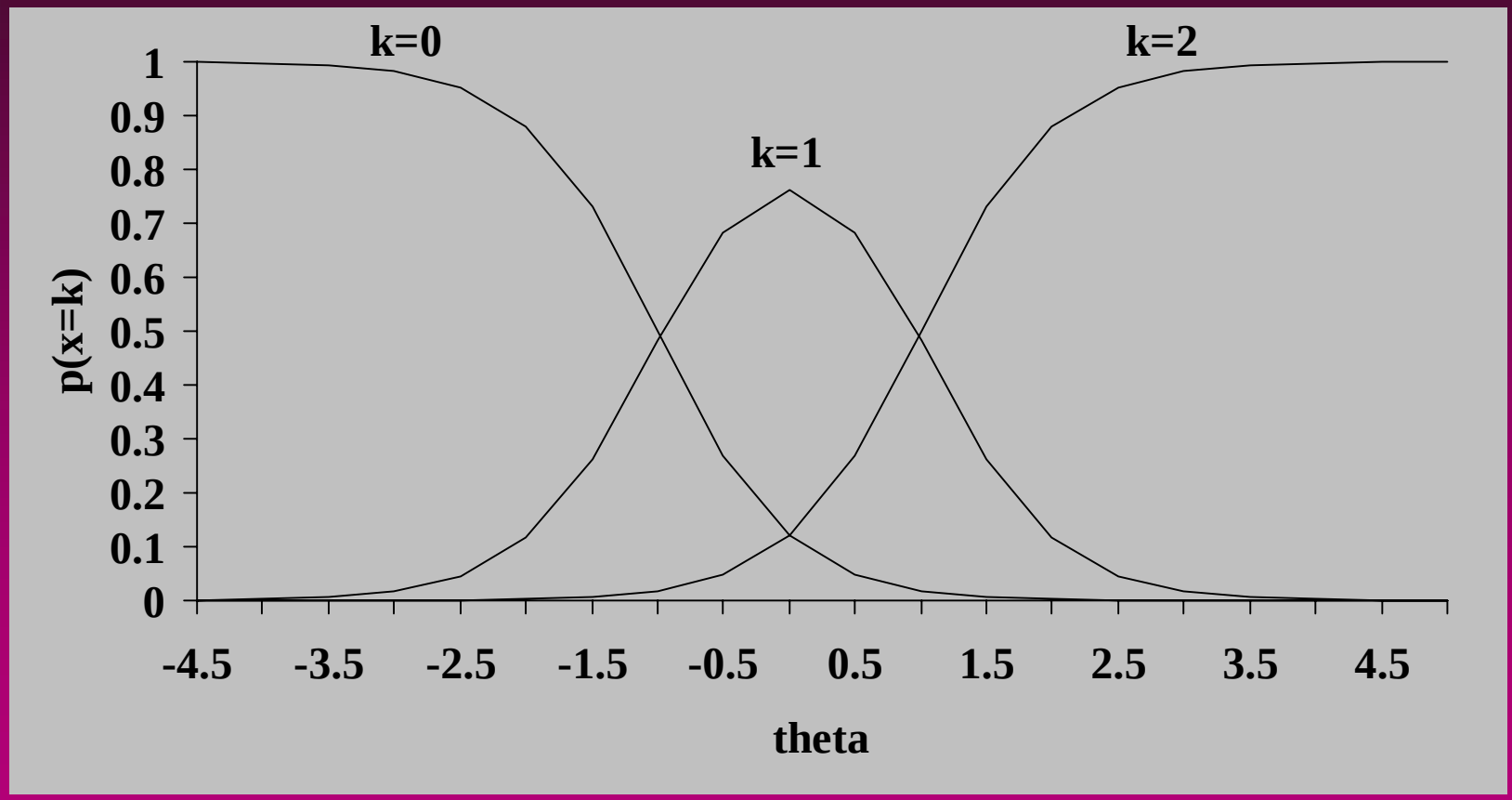
2-PL trace line



IRT Models (cont.)

- Polytomous items (Likert-type scales)
  - » Graded response model
  - » Partial credit model

Graded trace lines



IRT Models (cont.)

- Assumptions, choice of model, model fit

DIF Detection with IRT

- DIF indices
  - » Lord's chi-square
  - » Signed Area, Unsigned Area
  - » CDIF, NCDIF, DTF
- Nested model likelihood comparisons

DIF Detection with IRT (cont.)

- Linking the two groups
  - » Separate calibration and iterative equating
  - » Simultaneous calibration with anchor items

Application

- RAND/USC Injury Study
- Sample description
  - » 338 respondents, 26% (89) took Spanish translation of survey
- Survey items
  - » Demographic measures
  - » Psychological measures
- Can we combine the English and Spanish versions?

Method

- Identify dif for each scale of interest
  - » Identify anchor items and suspect items with discriminant function analysis
    - – Predict group membership from total score, item score, and interaction
    - – Anchor items have non-significant main and interaction effects

Method (cont.)

- Use Multilog to test suspect items for dif
  - Simultaneous calibration with anchor items to link groups
  - Compare fit of dif and non-dif models for each suspect item
  - Final model accounting for significant dif produces item parameter estimates and IRT-scores

Method (cont.)

- Investigate potential impact
  - » Examine item parameters, trace lines and boundary response functions of dif items
  - » Compare results of interest from dif-adjusted and non-dif-adjusted scores

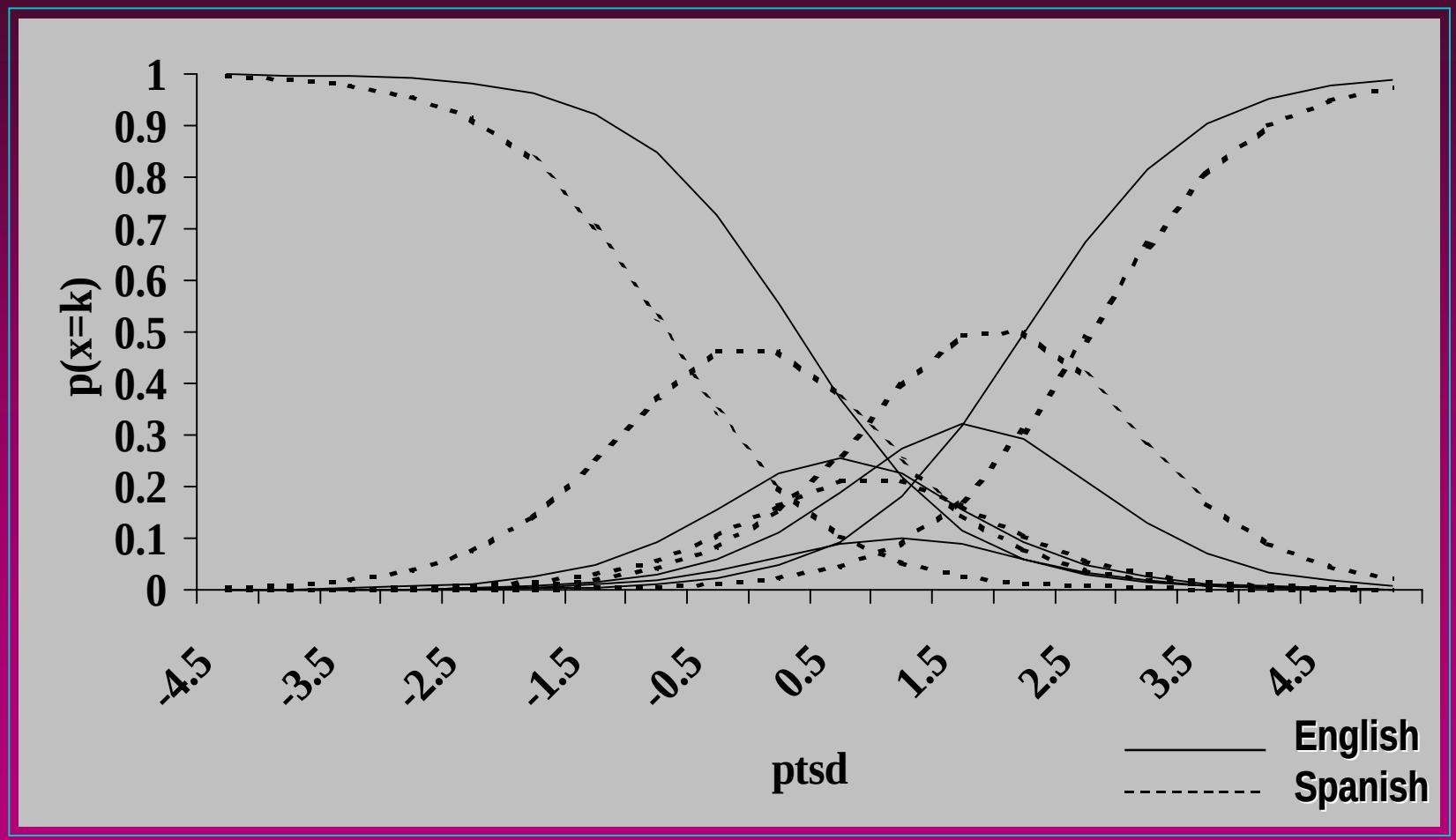
Results

- PTSD 17-items
  - » Discriminant analysis identified 7 anchor items
  - » 6 of 10 tested items had significant dif in location parameters

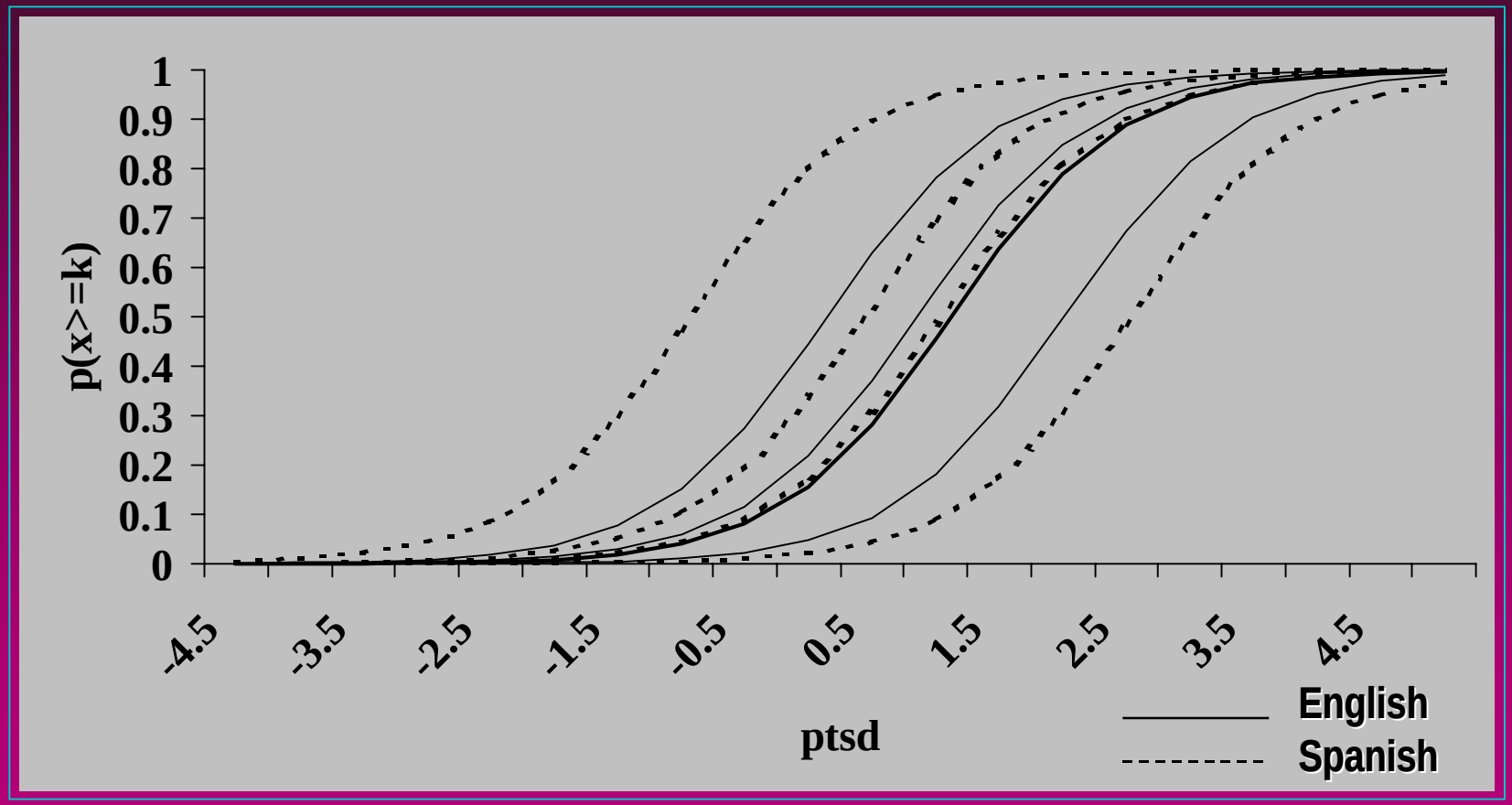
Model results

Item	-2LL Free	-2LL Equal	Difference (5 df)	Diff. a only (1 df)
2*	3590.8	3627.3	36.5	1.7
4*	3706.1	3742.4	36.3	0
6	3660.3	3663.9	3.6	
7*	3611.7	3651.4	39.7	.7
8*	3615.4	3635.7	20.3	2.8
9	3665.1	3675.7	10.6	
10	3672.2	3682.2	10	
11*	3427.5	3445.7	18.2	3.2
12	2630.9	3640.3	9.4	
14*	3659.7	3683.1	23.4	.3

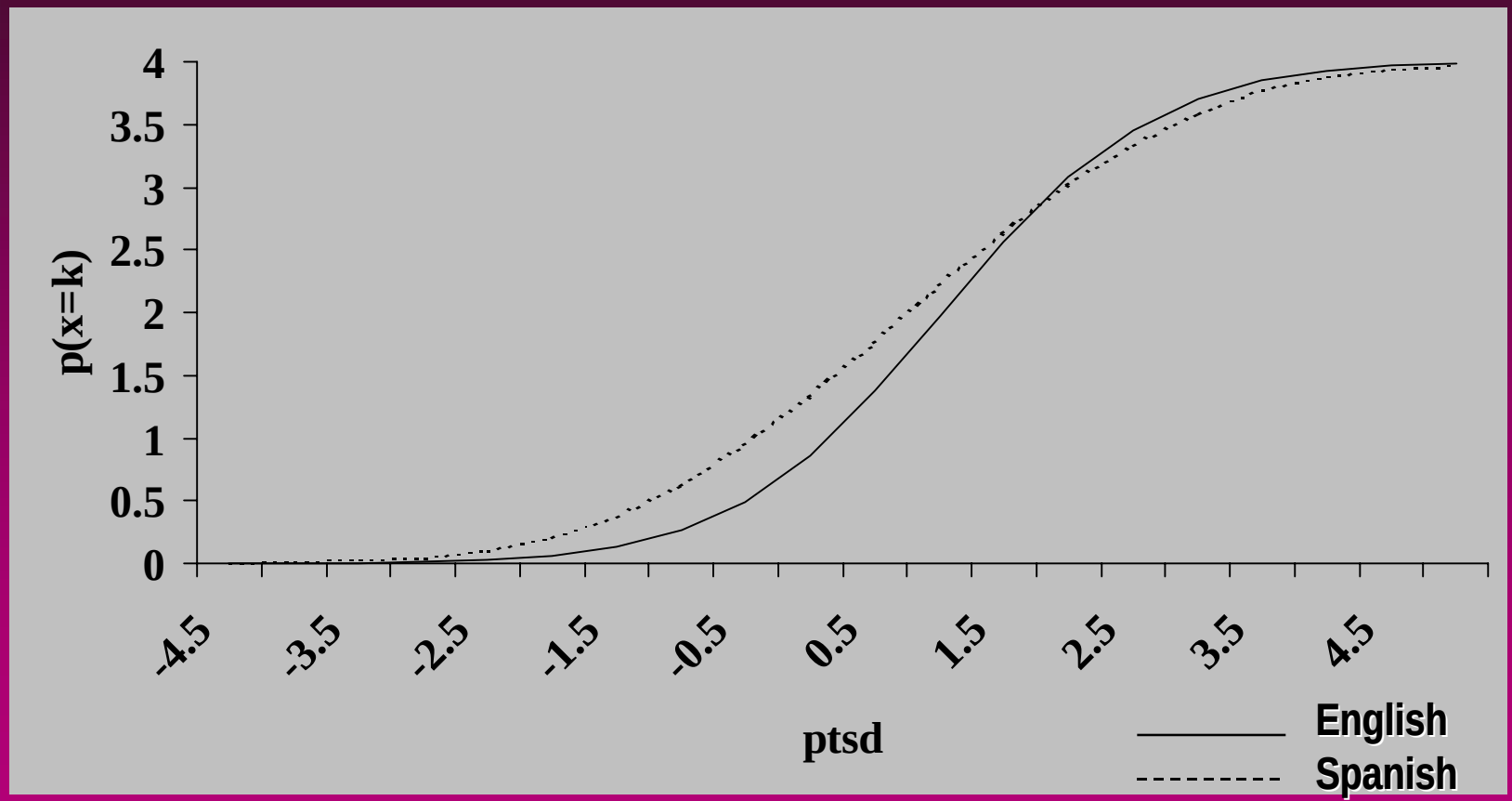
Trace lines - “Disturbing dreams”



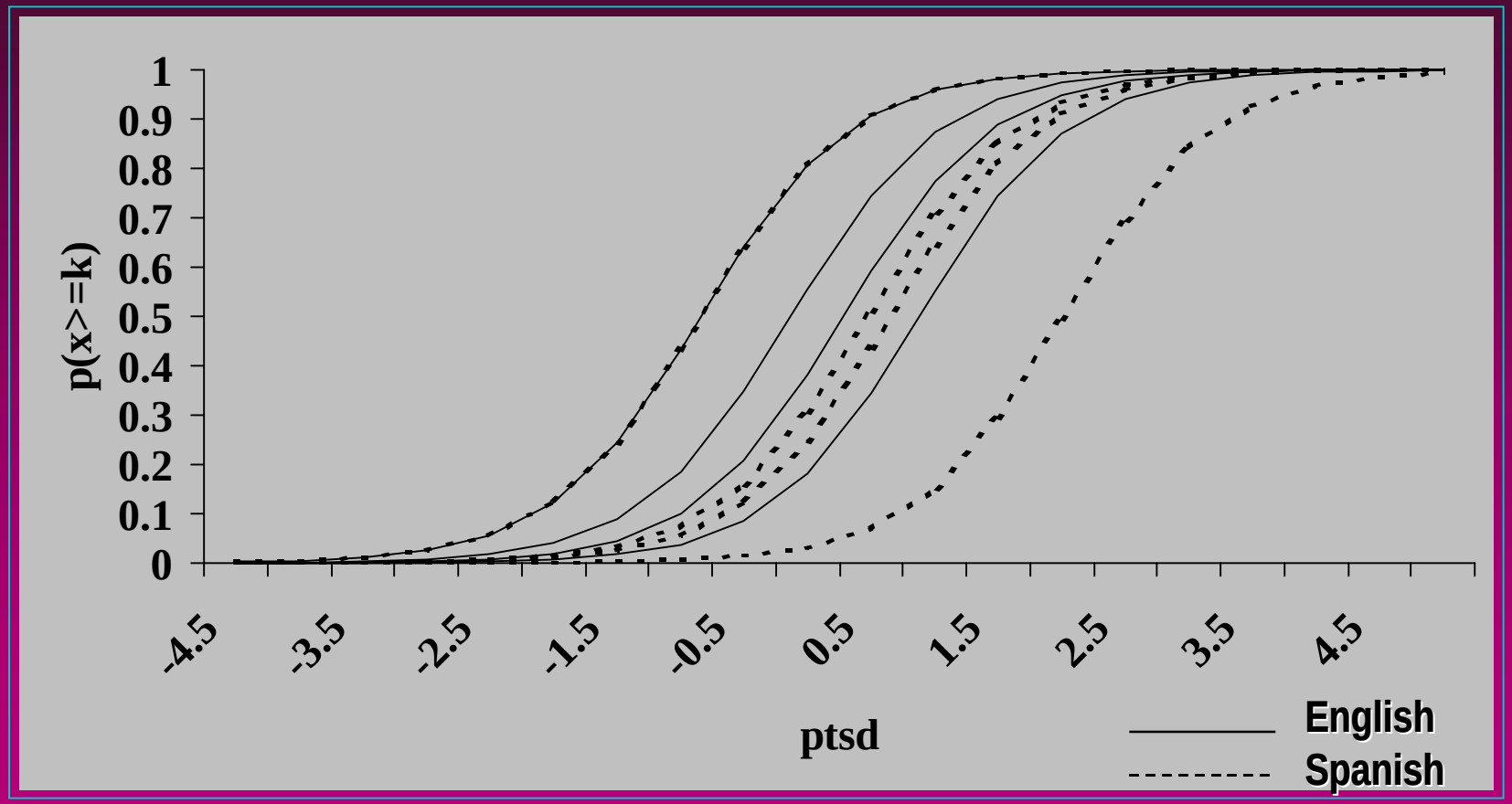
Boundary response functions - "Disturbing dreams"



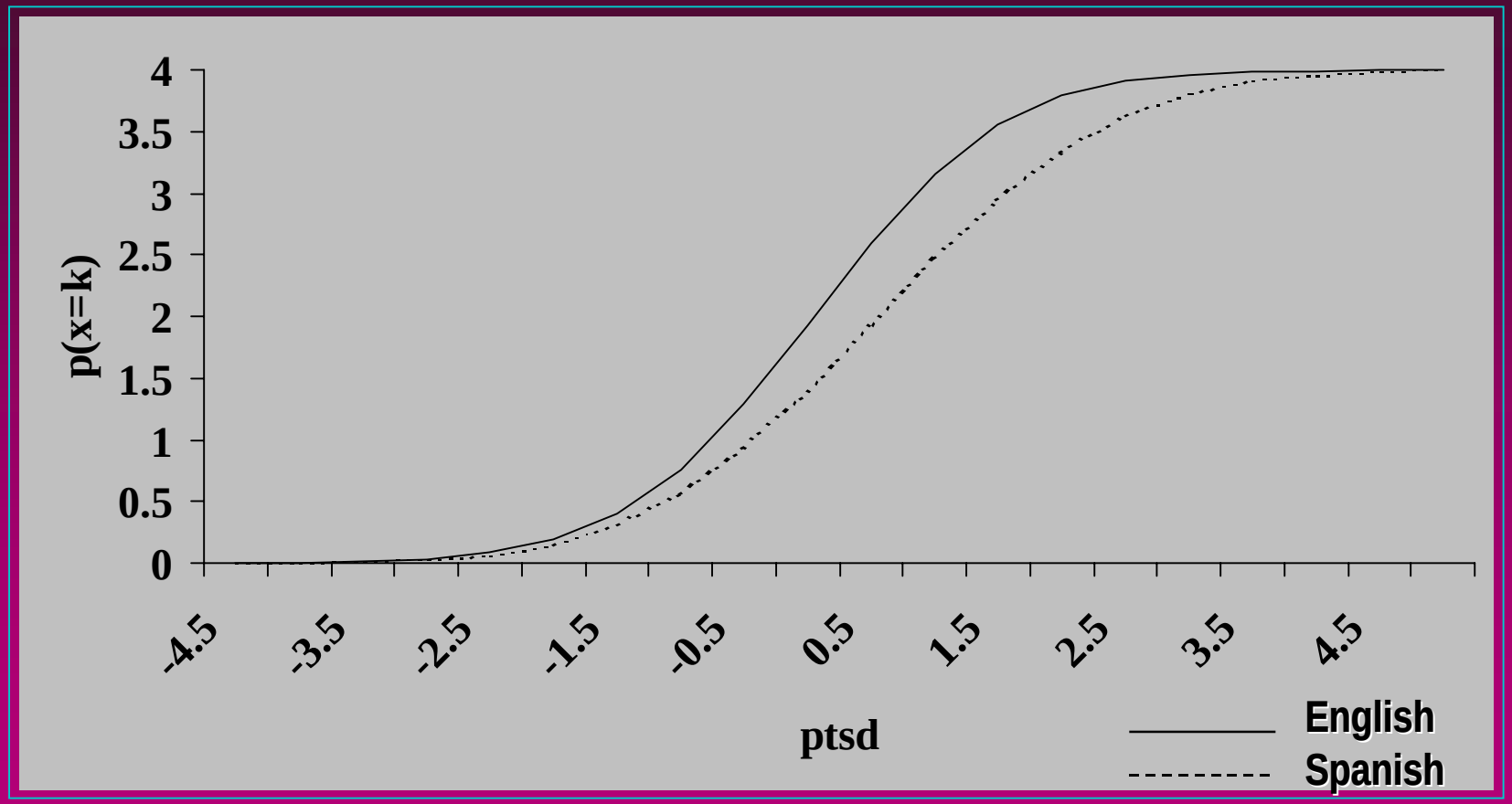
Item response functions - "Disturbing dreams"



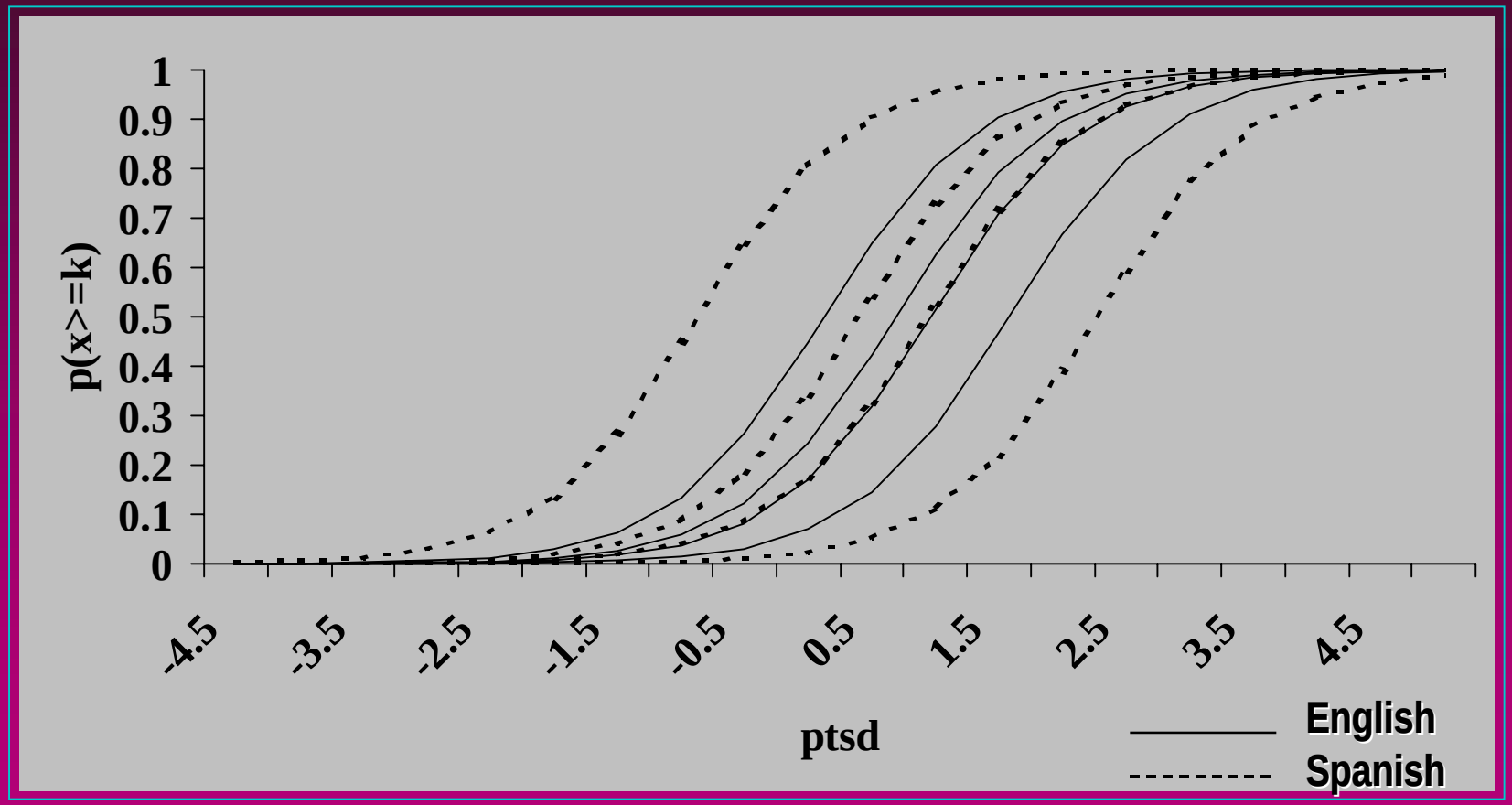
BRF - "Upset when reminded"



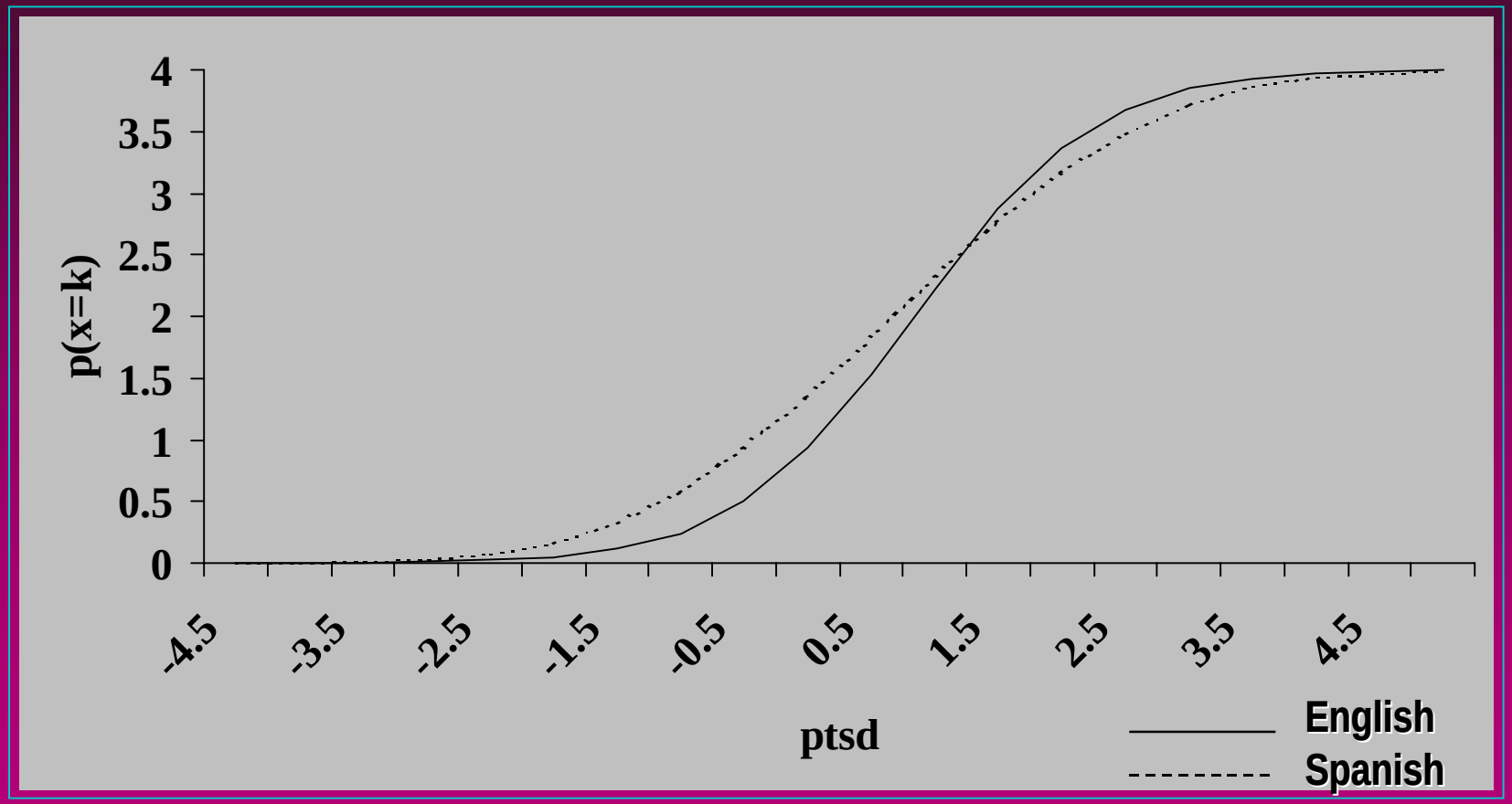
IRF - "Upset when reminded"



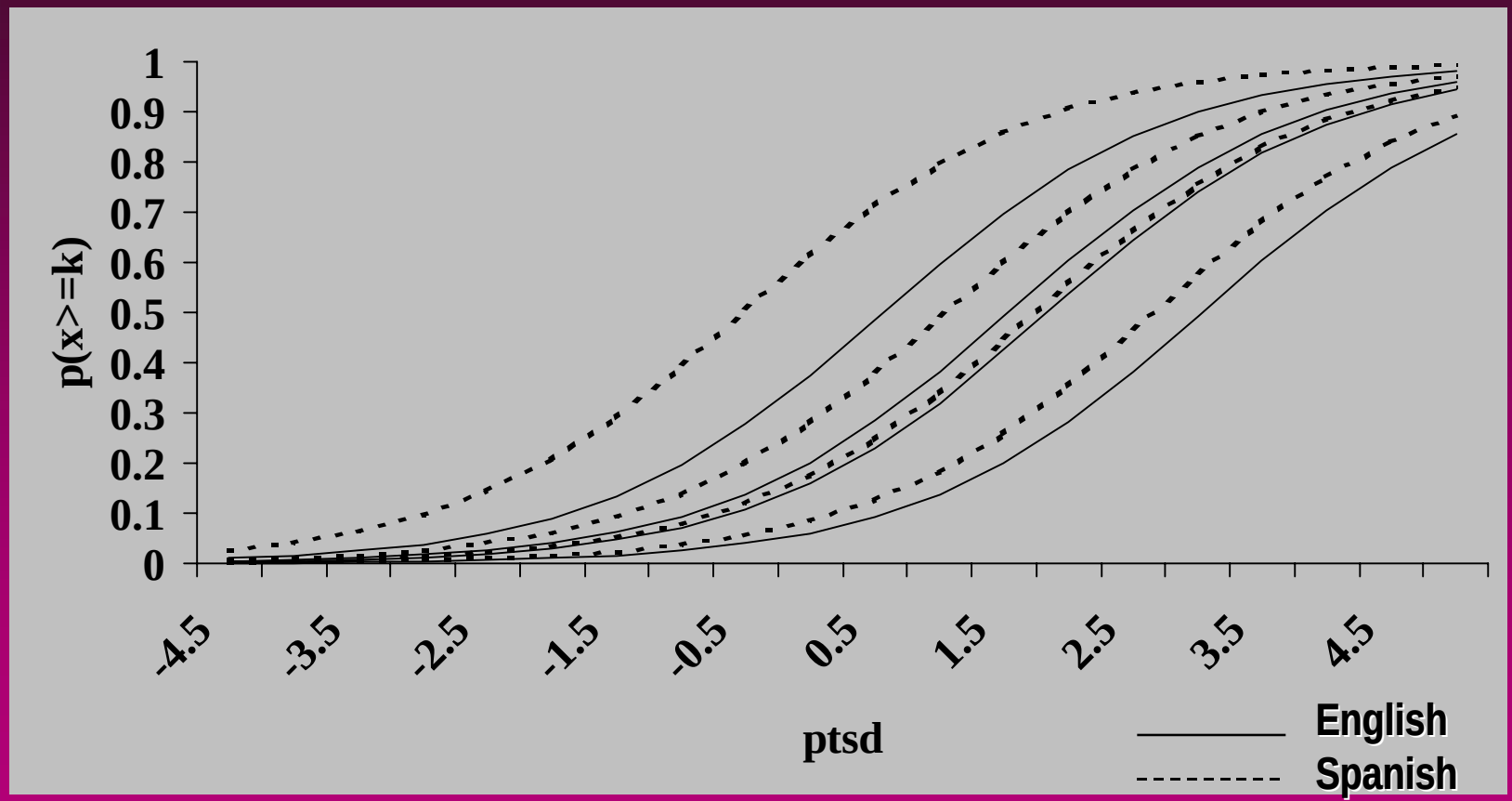
BRF - "Avoiding activities"



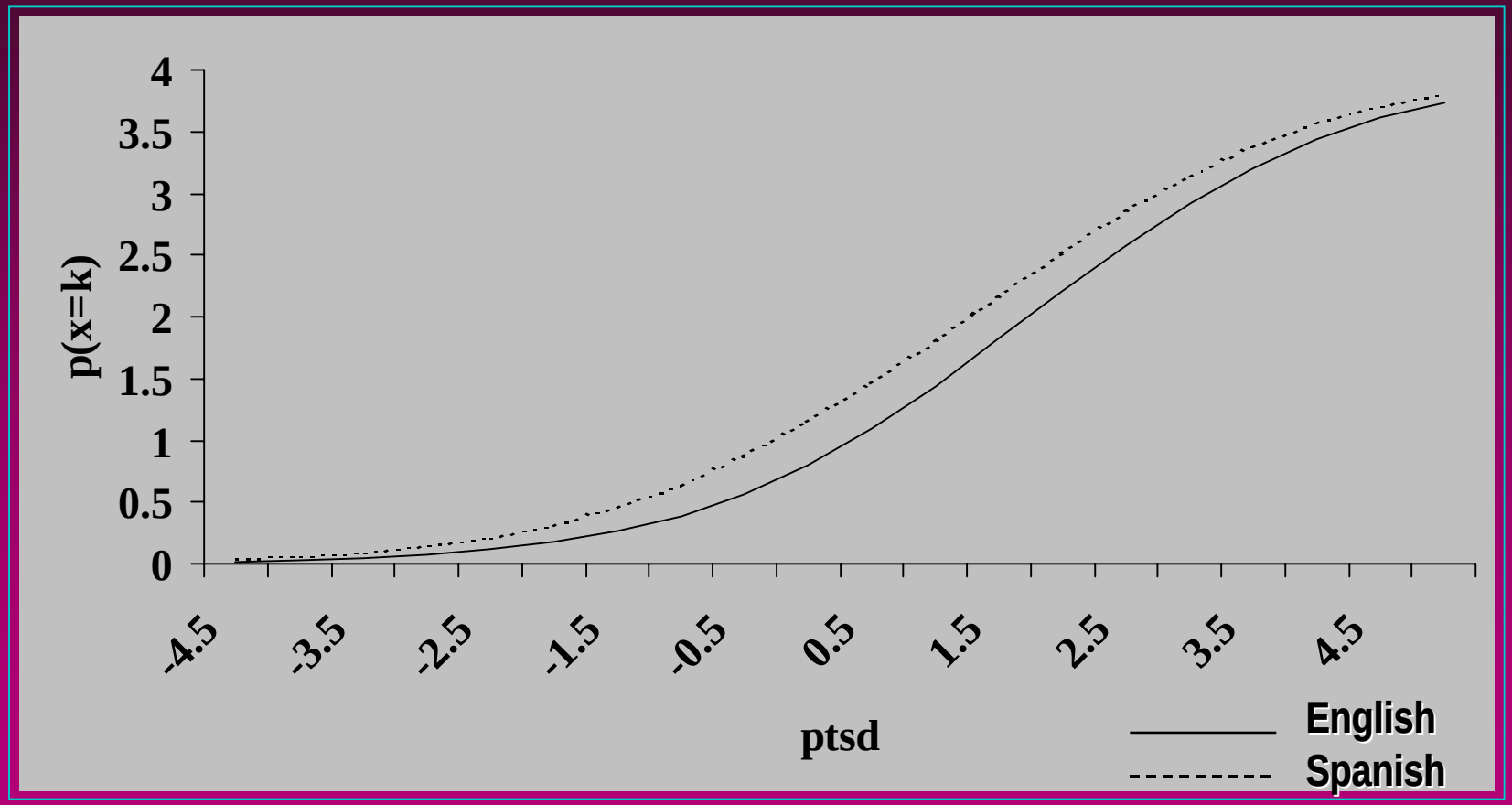
IRF - "Avoiding activities"



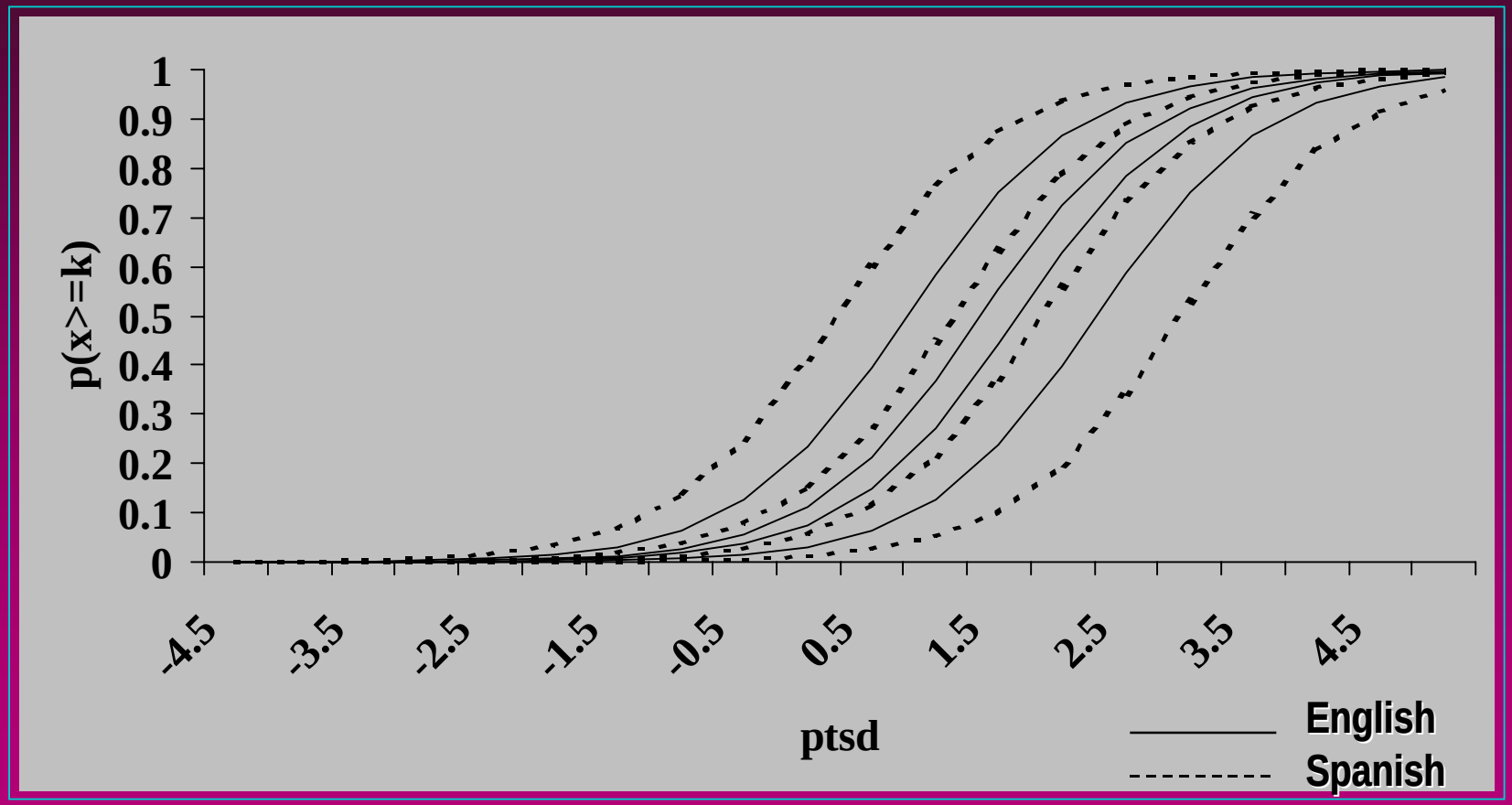
BRF - “Trouble remembering”



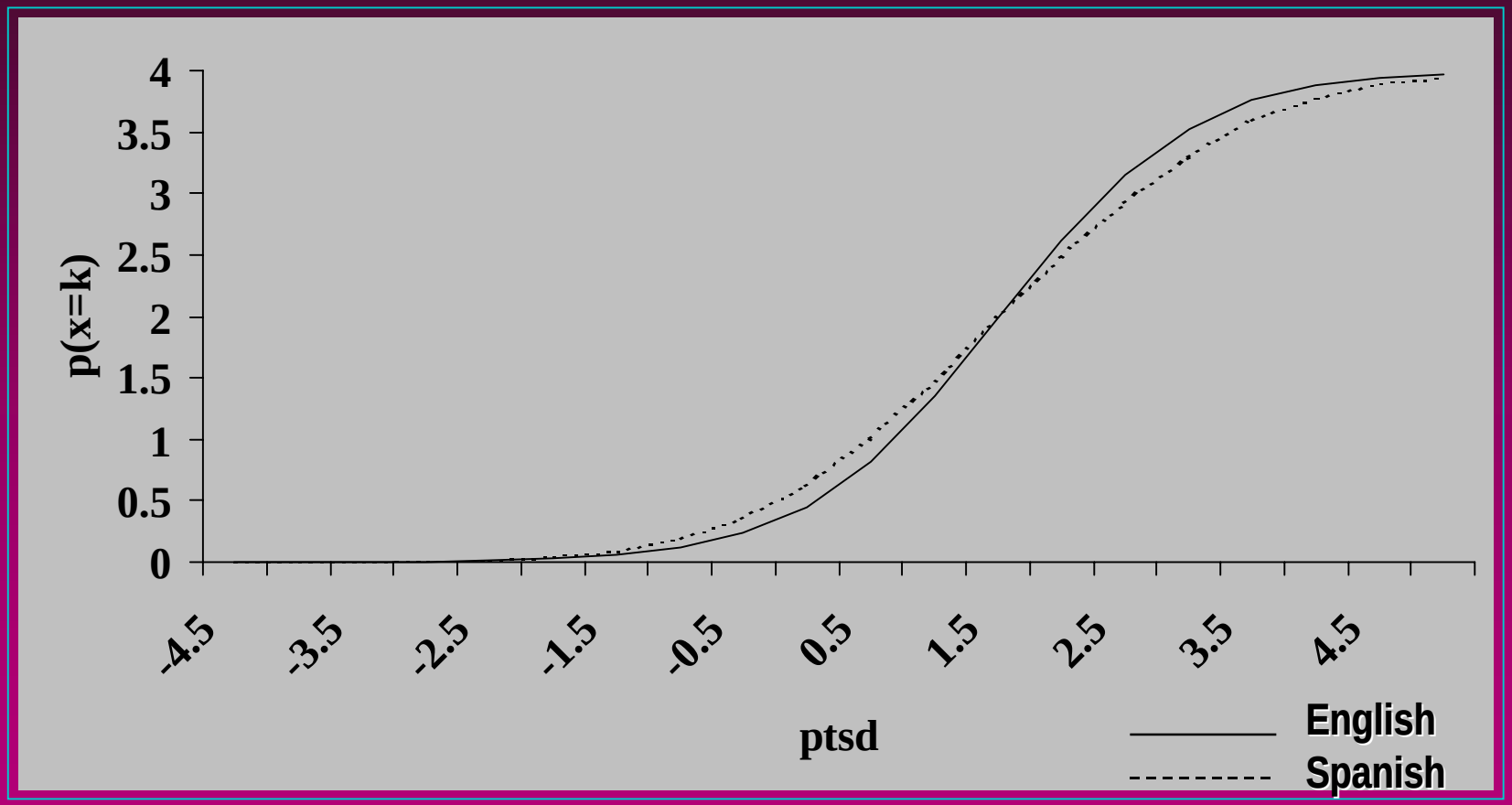
IRF - "Trouble remembering"



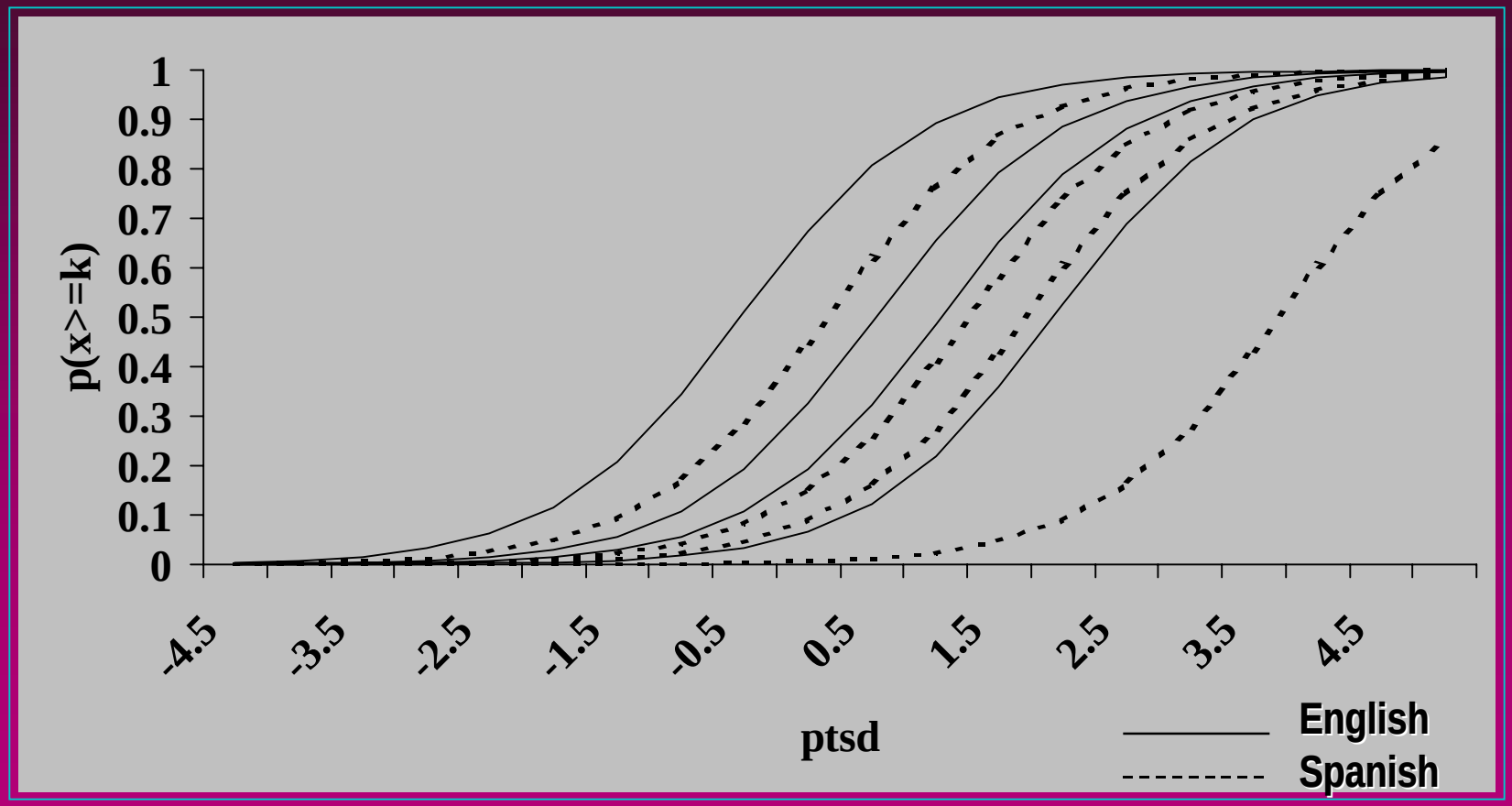
BRF - "Emotionally numb"



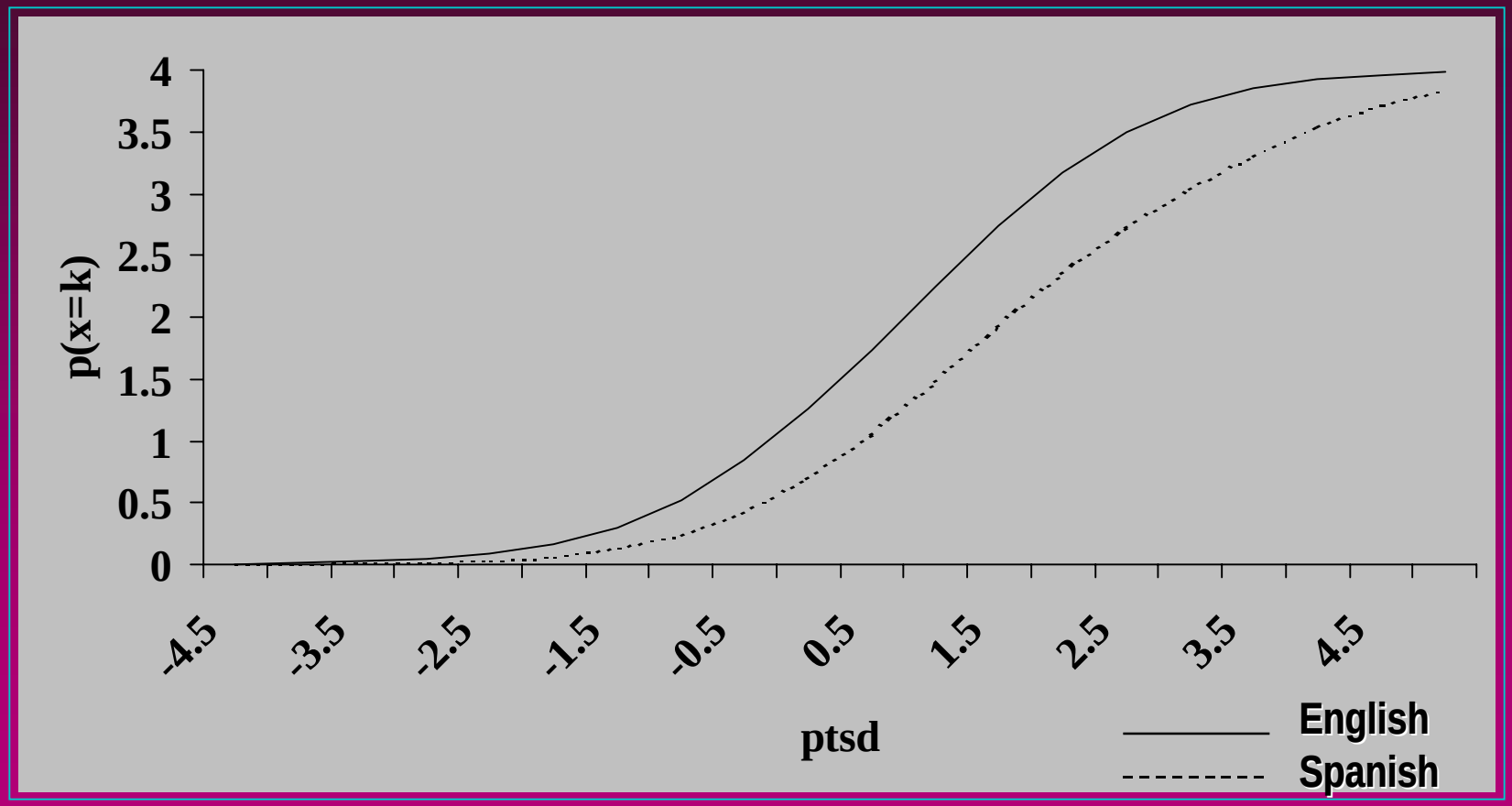
IRF - “Emotionally numb”



BRF - “Irritable, angry outbursts”



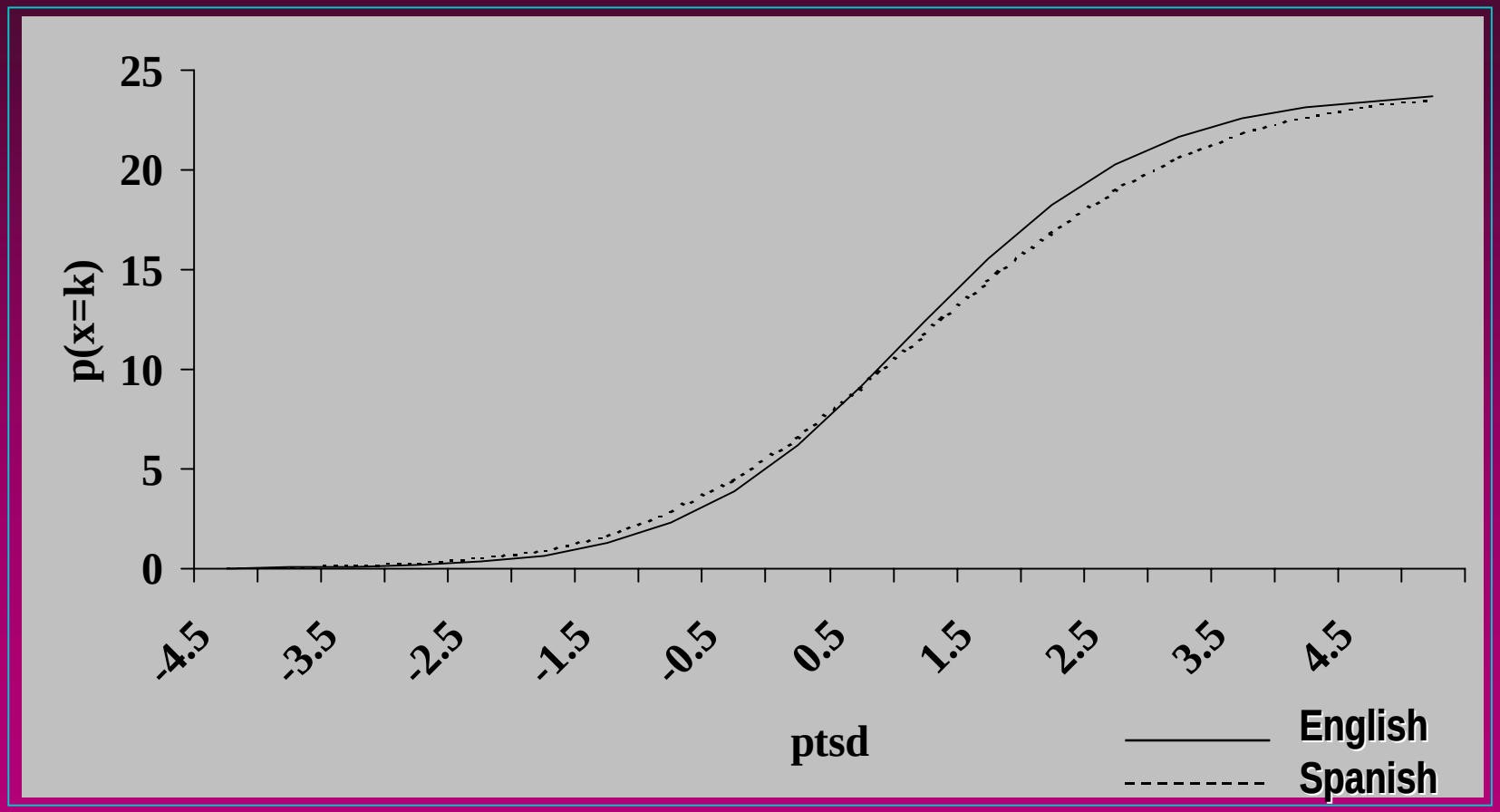
IRF - “Irritable, angry outbursts”



Evaluating Impact

- How different are the groups at the scale level (as opposed to the item level)?
  - » Compare TRFs of two groups
  - » Compare group means

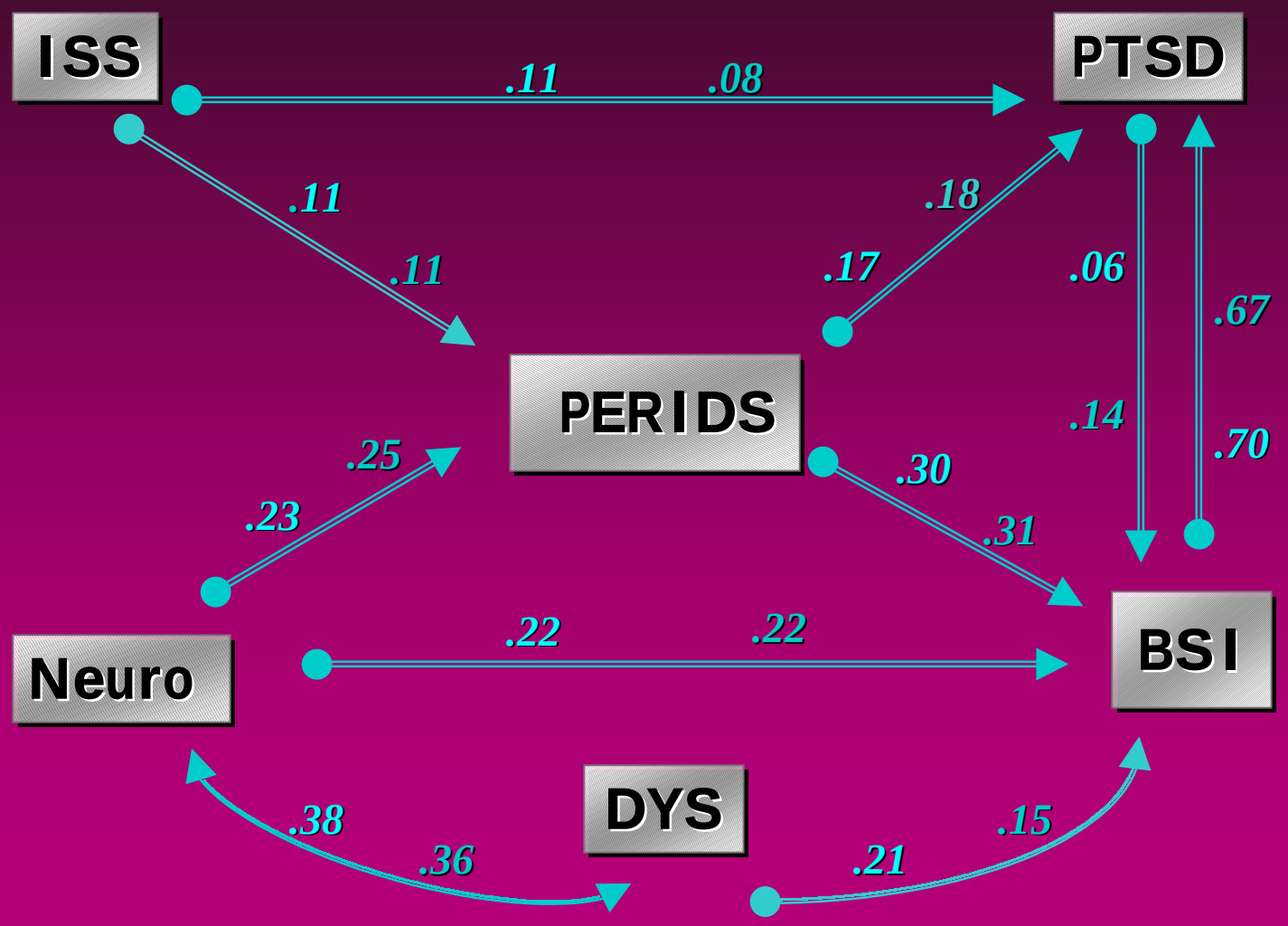
Test response function for 6 dif items



Evaluating Impact (cont.)

- Will adjusting for dif change substantive results?
  - » Example – path model

Path Model Results



General discussion

- Choice of method
- Anchor items/equating
- Impact/decisions/use of test

General discussion (cont.)

- Practical barriers to implementation
  - » Software is unfamiliar and not always readily available
  - » Appropriate use of software requires extensive knowledge of IRT