Evaluating Multi-item Scales

Ron D. Hays, Ph.D.
UCLA Division of General Internal Medicine/Health Services Research
drhays@ucla.edu
http://twitter.com/RonDHays
http://gim.med.ucla.edu/FacultyPages/Hays/

HS225A 11/04/10, 10-11:50 pm, 51-279 CHS

Example Responses to 2-Item Scale

ID	Poor	Fair	Good	Very Good	Excellent
01					2
02				1	1
03		1		1	
04			1		1
05		2			

Cronbach's Alpha

01 55
02 45
03 42
04 35
05 22

Source	df	SS	MS
Respondents (BMS)	4	11.6	2.9
Items (JMS)	1	0.1	0.1
Resp. x Items (EMS)	4	4.4	1.1
Total	9	16.1	

$$\text{Alpha} = \frac{2.9 - 1.1}{2.9} = \frac{1.8}{2.9} = \boxed{0.62}$$

Computations

- Respondents SS

  $(10^2+9^2+6^2+8^2+4^2)/2 - 37^2/10 = \underline{11.6}$

- Item SS

  $(18^2+19^2)/5 - 37^2/10 = \underline{0.1}$

- Total SS

  $(5^2+ 5^2+4^2+5^2+4^2+2^2+3^2+5^2+2^2+2^2) - 37^2/10 = \underline{16.1}$

- Res. x Item SS= Tot. SS – (Res. SS+Item SS)

Alpha for Different Numbers of Items and Average Correlation

Number of Items (k)	Average Inter-item Correlation ($\bar{r}$)					
	.0	.2	.4	.6	.8	1.0
2	.000	.333	.572	.750	.889	1.000
4	.000	.500	.727	.857	.941	1.000
6	.000	.600	.800	.900	.960	1.000
8	.000	.666	.842	.924	.970	1.000

$$\text{Alpha}_{st} = \frac{k * \bar{r}}{1 + (k - 1) * \bar{r}}$$

Spearman-Brown Prophecy Formula

$$\text{alpha}_y = \left(\frac{N \bullet \text{alpha}_x}{1 + (N - 1) * \text{alpha}_x} \right)$$

N = how much longer scale y is than scale x

Example Spearman-Brown Calculation

MHI-18

$$\frac{18/32\ (0.98)}{(1+(18/32\ -1)*0.98}$$

$$= 0.55125/0.57125 = 0.96$$

Reliability Minimum Standards

- 0.70 or above (for group comparisons)
- 0.90 or higher (for individual assessment)
  - SEM = SD (1- reliability)$^{1/2}$

Intraclass Correlation and Reliability

Model	Reliability	Intraclass Correlation
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$
Two-way fixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} + MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$

BMS = Between Ratee Mean Square
WMS = Within Mean Square
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square

Equivalence of Survey Data

- Missing data rates were significantly higher for African Americans on all CAHPS items
- Internal consistency reliability did not differ
- Plan-level reliability estimates were significantly lower for African Americans than whites

M. Fongwa et al. (2006). Comparison of data quality for reports and ratings of ambulatory care by African American and White Medicare managed care enrollees. Journal of Aging and Health, 18, 707-721.

Table 3
Estimated Number of Patients Needed to Obtain .70 Health Plan-Level Reliability on Consumer Assessment of Healthcare Providers and Systems Health Plan Scales and Global Rating Items

Scale	*White*	*African American*	*White and African American Combined*
Getting care quickly/timeliness	82	118	74
Provider communication	124	177	134
Staff helpfulness	121	128	120
Getting needed care	76	110	78
Customer service	68	98	66
Global rating items			
Rate personal doctor	124	219	136
Rate specialist	122	554	124
Rate health care	92	122	98
Rate health plan	41	67	42

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.50*	0.50	0.50
Item #2	0.50*	0.50	0.50
Item #3	0.50*	0.50	0.50
Item #4	0.50	0.50*	0.50
Item #5	0.50	0.50*	0.50
Item #6	0.50	0.50*	0.50
Item #7	0.50	0.50	0.50*
Item #8	0.50	0.50	0.50*
Item #9	0.50	0.50	0.50*

*Item-scale correlation, corrected for overlap.

Table 1.4.2 Multitrait/multi-item correlation matrix for patient satisfaction ratings

	Technical	Interpersonal	Communication	Financial
Technical				
1	0.66*	0.63@	0.67@	0.28
2	0.55*	0.54@	0.50@	0.25
3	0.48*	0.41	0.44@	0.26
4	0.59*	0.53	0.56@	0.26
5	0.55*	0.60@	0.56@	0.16
6	0.59*	0.58@	0.57@	0.23
Interpersonal				
1	0.58	0.68*	0.63@	0.24
2	0.59@	0.58*	0.61@	0.18
3	0.62@	0.65*	0.67@	0.19
4	0.53@	0.57*	0.60@	0.32
5	0.54	0.62*	0.58@	0.18
6	0.48@	0.48*	0.46@	0.24
Communication				
1	0.58@	0.59@	0.61*	0.26
2	0.47@	0.50@	0.50*	0.25
3	0.58@	0.66@	0.63*	0.23
4	0.66@	0.66@	0.67*	0.25
5	0.66@	0.71@	0.70*	0.25
Financial				
1	0.35	0.35	0.35	0.72*
2	0.17	0.14	0.15	0.65*
3	0.25	0.23	0.23	0.61*
4	0.18	0.15	0.16	0.67*
5	0.31	0.27	0.29	0.70*
6	0.24	0.23	0.22	0.73*
7	0.25	0.23	0.25	0.55*
8	0.34	0.31	0.31	0.64*
Cronbach's alpha	0.80	0.82	0.82	0.88

Note: Standard error of correlation is 0.03.

Technical = satisfaction with technical quality; Interpersonal = satisfaction with interpersonal aspects; Communication = satisfaction with communication; Financial = satisfaction with financial arrangements.

@ Correlation is within two standard errors of the correlation of the item with its hypothesized scale.

* Item-scale correlation, corrected for overlap.

Table 1.4.3 Correlations between patient satisfaction scales

	Technical	Interpersonal	Communication	Financial
Technical	1.00	0.75	0.76	0.34
Interpersonal	0.93	1.00	0.80	0.31
Communication	0.94	0.98	1.00	0.32
Financial	0.41	0.36	0.38	1.00

Note: Zero-order correlations are provided above the diagonal; correlations adjusting for unreliability of measurement are given below the diagonal.

Confirmatory Factor Analysis

- Observed covariances compared to covariances generated by hypothesized model
- Statistical and practical tests of fit
- Factor loadings
- Correlations between factors

Fit Indices

- Normed fit index: $\dfrac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$
- Non-normed fit index: $\dfrac{\frac{\chi^2_{null}}{df_{null}} - \frac{\chi^2_{model}}{df_{model}}}{\left[\frac{\chi^2_{null}}{df_{null}} - 1\right]}$
- Comparative fit index: $1 - \dfrac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}$

Hays, Cunningham, Ettl, Beck & Shapiro (1995, Assessment)

- 205 symptomatic HIV+ individuals receiving care at two west coast public hospitals
- 64 HRQOL items
- 9 access, 5 social support, 10 coping, 4 social engagement and 9 HIV symptom items

Table 4
Loadings of Observed Variables on Latent Variables in Final Model

	Physical health	Mental health	Uniqueness
Physical function	.79	–	.61
Role function	.70	–	.72
Less pain	.58	–	.82
Less disability days	.54	–	.84
Quality of sex life	.29	–	.96
Overall quality of life	–	.80	.60
Emotional well-being	–	.75	.66
Hopefulness	–	.74	.68
Freedom from loneliness	–	.70	.72
Will to function	–	.66	.75
Quality of family life	–	.56	.83
Quality of friendships	–	.54	.84
Cognitive function/distress	–	.51	.86
Current health	.48	.31	.76
Social function	.46	.39	.72
Energy/fatigue	.65	.18	.69
Quality of leisure/social	.33	.36	.83

Note. All indicators were scored so that a higher score represents better health. All estimates shown above were significant ($p < .05$). Dashed entries represent parameters fixed at zero in the

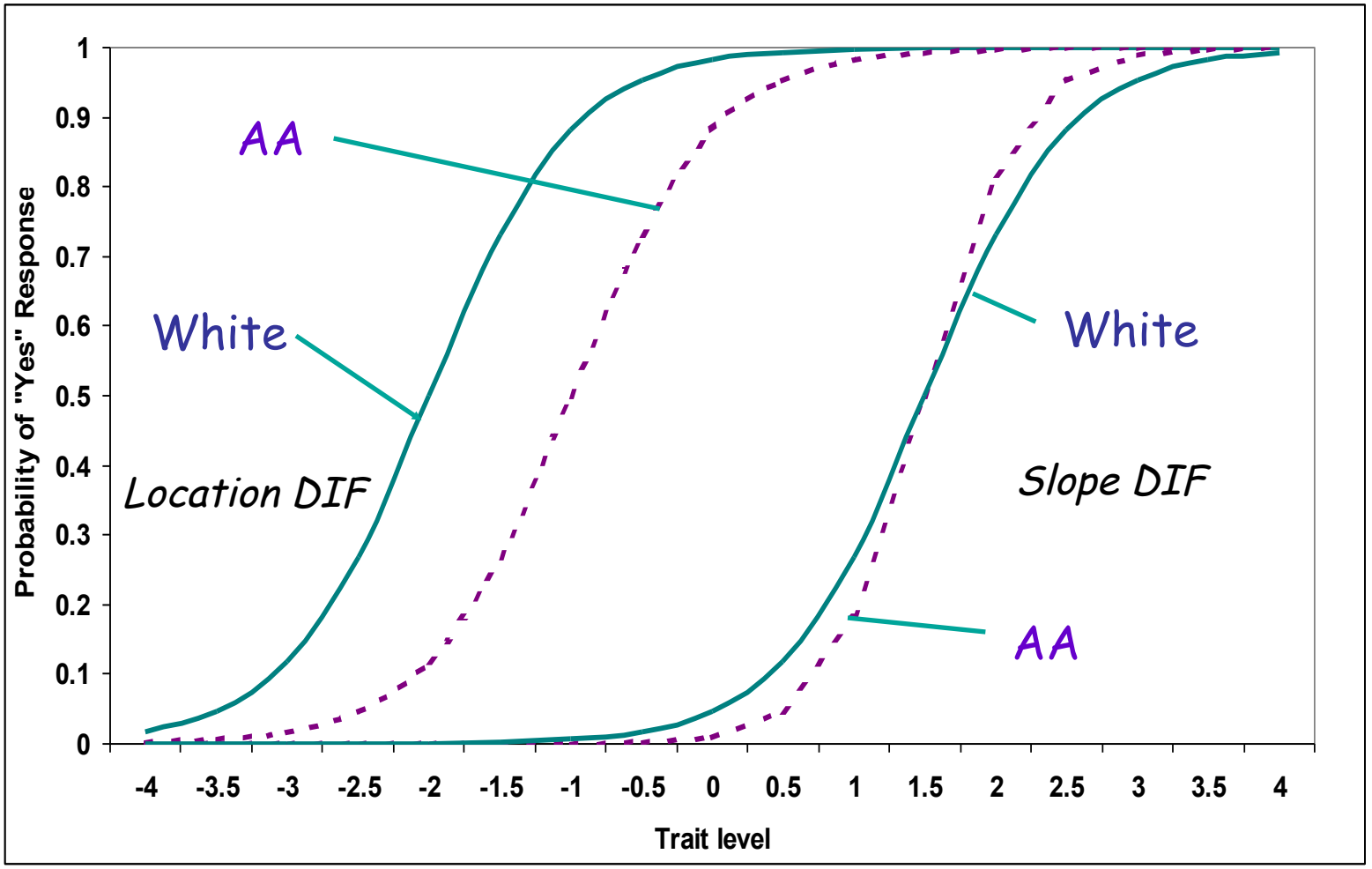
Table 5
Correlations Between Latent and Observed Variables in Final Model

	r
Physical health latent variable with:	
Less exhaustion	.58
Less myalgia	.48
Better appetite	.39
Less fever	.26
Less night sweats	.25
Less weight loss	.17
Mental health latent variable with:	
Less exhaustion	.25
Less myalgia	.22
Better appetite	.12
Social support	.54
Better coping response	.54
Less social disengagement	.40
Access to care	.35

and substantial in magnitude, ranging from 0.51 to 0.80. Although one might hope that the overall quality of life scale would load on the physical health as well as the mental health latent variable as occurred in an exploratory factor analysis of gay men with HIV disease (Burgess et al., 1993), this was not the case. Our finding indicates that overall quality of life uniquely reflects mental health, consistent with a factor analysis of HRQOL measures in a study of epilepsy patients (Vickrey et al., 1992). The significant loadings on both the physical and mental health latent variables for current health perceptions, social function, and energy/fatigue are also consistent with the results of previous studies (Hays, Sherbourne, & Mazel, 1993; Vickrey et al., 1993).

The physical and mental health latent variables correlated .31, sharing about 10% of the variance in common. This level of association is very similar to correlations of .43 and .46 reported for similar factors in previous studies (Hays & Stewart, 1990; Burgess et al., 1993). The empirical distinc-

Differential Item Functioning (2-Parameter Model)



Location = uniform; Slope = non-uniform

Language DIF Example

-
- Ordinal logistic regression to evaluate differential item functioning
  - Purified IRT trait score as matching criterion
  - McFadden's pseudo R^2 >= 0.02
- Thetas estimated in Spanish data using
  - English calibrations
  - Linearly transformed Spanish calibrations (Stocking-Lord method of equating)

Lordif

http://CRAN.R-project.org/package=lordif

Model 1 : logit $P(u_i >= k) = \alpha_k + \mathbf{\beta_1} * \mathbf{ability}$

Model 2 : logit $P(u_i >= k) = \alpha_k + \beta_1 *$ ability $+ \mathbf{\beta_2} * \mathbf{group}$

Model 3 : logit $P(u_i >= k) = \alpha_k + \beta_1 *$ ability $+ \beta_2 *$ group $+ \mathbf{\beta_3} * \mathbf{ability} * \mathbf{group}$

DIFF assessment (log likelihood values compared):

- Overall: Model 3 versus Model 1
- Non-uniform: Model 3 versus Model 2
- Uniform: Model 2 versus Model 1

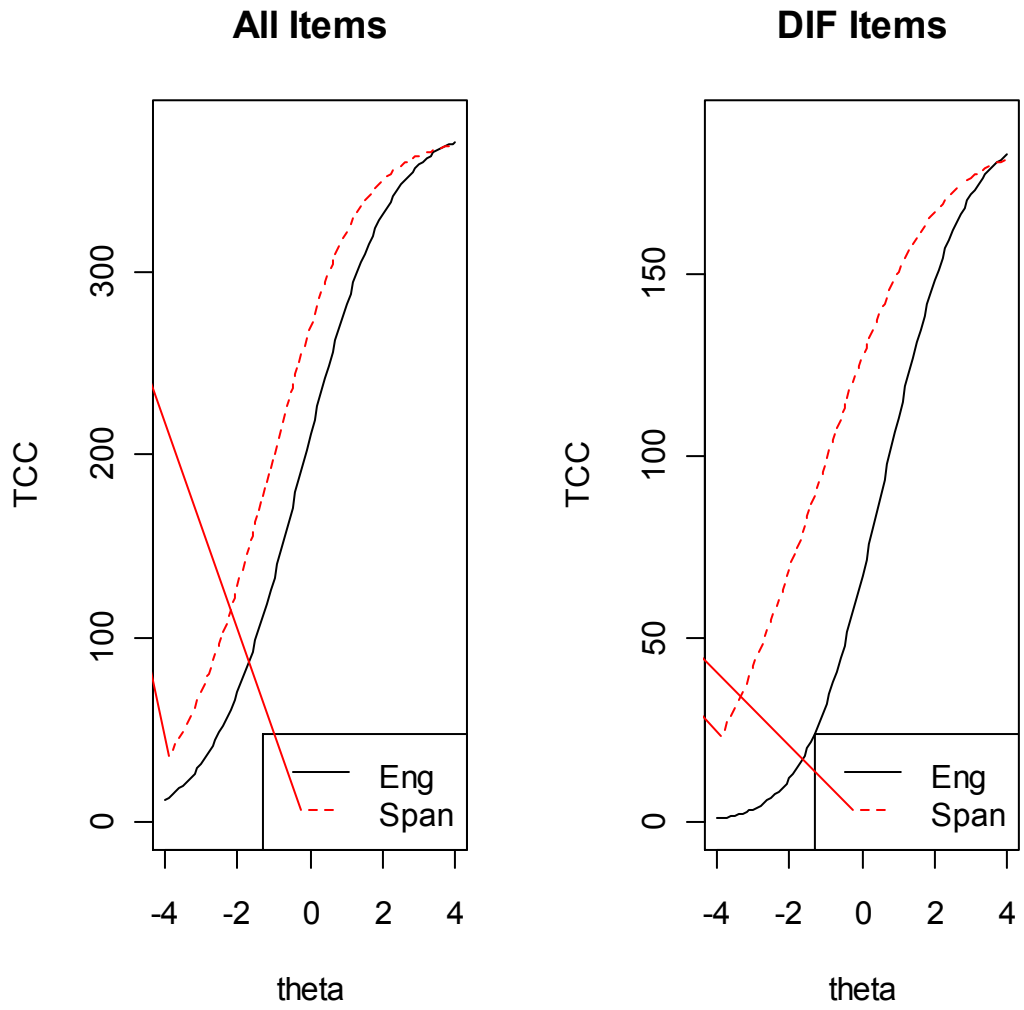
Sample Demographics

	English (n = 1504)	Spanish (n = 640)
% Female	52%	58%
% Hispanic	11%	100%
Education		
< High school	2%	14%
High school	18%	22%
Some college	39%	31%
College degree	41%	33%
Age	51 (SD = 18)	38 (SD = 11)

Results

- One-factor categorical model fit the data well (CFI=0.971, TLI=0.970, and RMSEA=0.052).
  - Large residual correlation of 0.67 between “Are you able to run ten miles” and “Are you able to run five miles?”
- 50 of the 114 items had language DIF
  - 16 uniform
  - 34 non-uniform

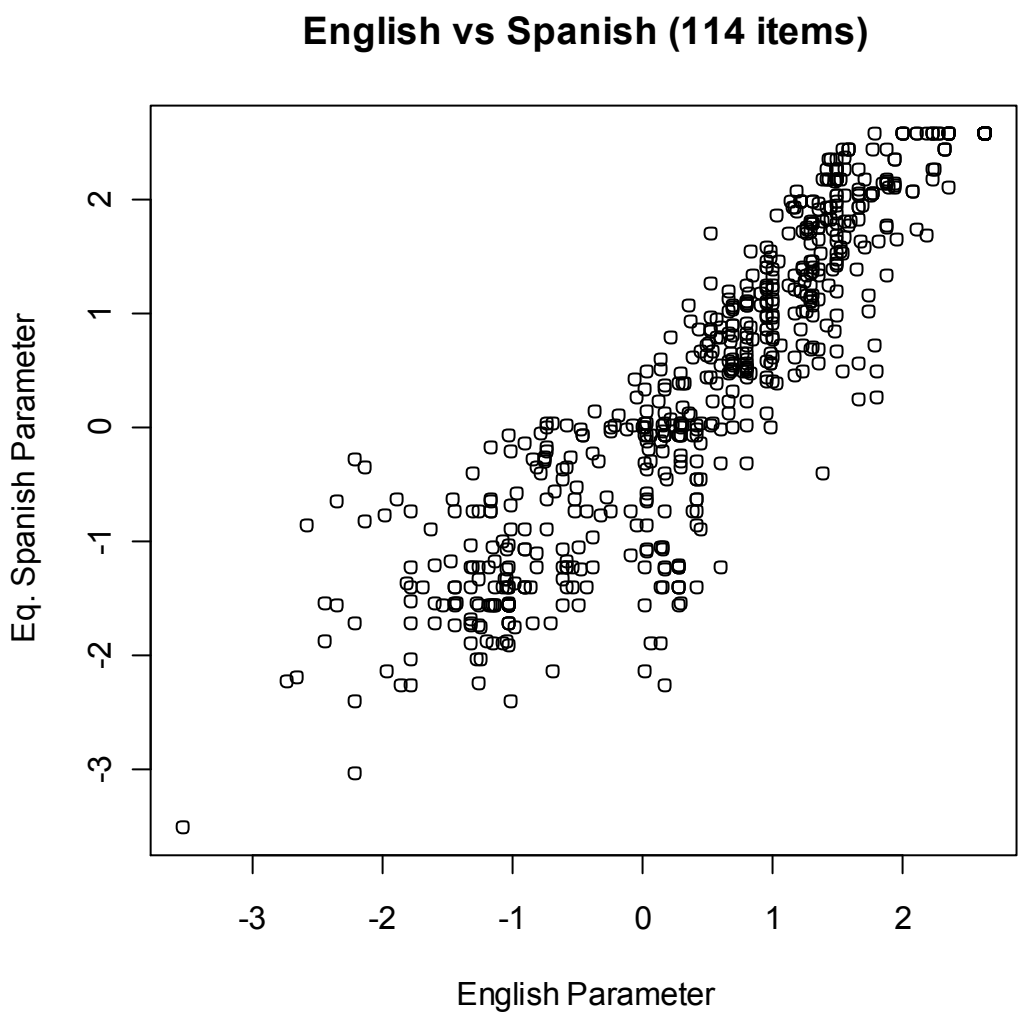
Impact of DIF on Test Characteristic Curves (TCCs)



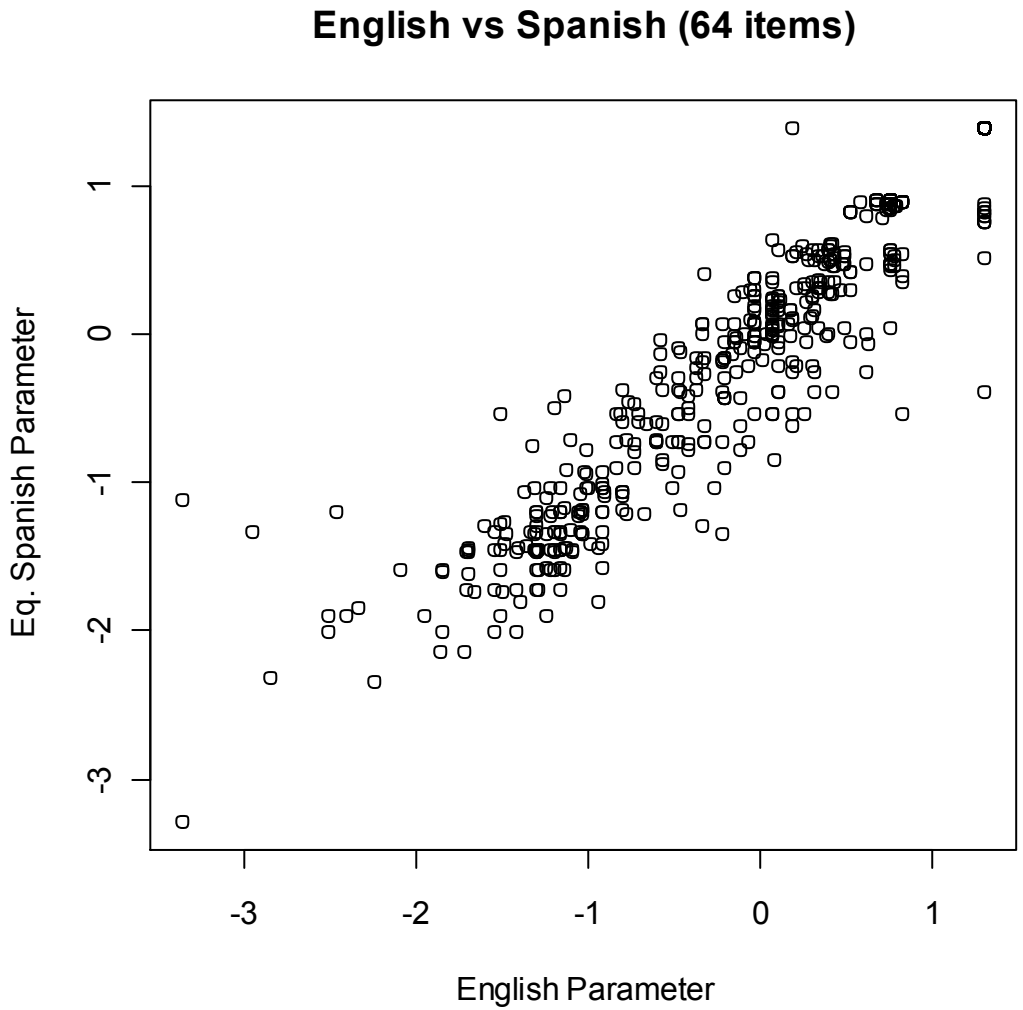
Stocking-Lord Method

- Spanish calibrations transformed so that their TCC most closely matches English TCC.
- $a^* = a/A$ and $b^* = A * b + B$
- Optimal values of A (slope) and B (intercept) transformation constants found through multivariate search to minimize weighted sum of squared distances between TCCs of English and Spanish transformed parameters
  - Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

CAT-based Theta Estimates Using English (x-axis) and Spanish (y-axis) Parameters for 114 Items in Spanish Sample (n = 640, ICC = 0.89)



CAT-based Theta Estimates Using English (x-axis) and Spanish (y-axis) Parameters for 64 non-DIF Items in Spanish Sample (n = 640, ICC = 0.96)



Implications

- Hybrid model needed to account for language DIF
- English calibrations for non-DIF items
- Spanish calibrations for DIF items

Thank you.