Evaluating English Versus Spanish Language Equivalence in the Patient-Reported Outcomes Measurement Information System Physical Functioning Item Bank

Sylvia H. Paz, Karen Spritzer, Leo Morales, Ron D. Hays

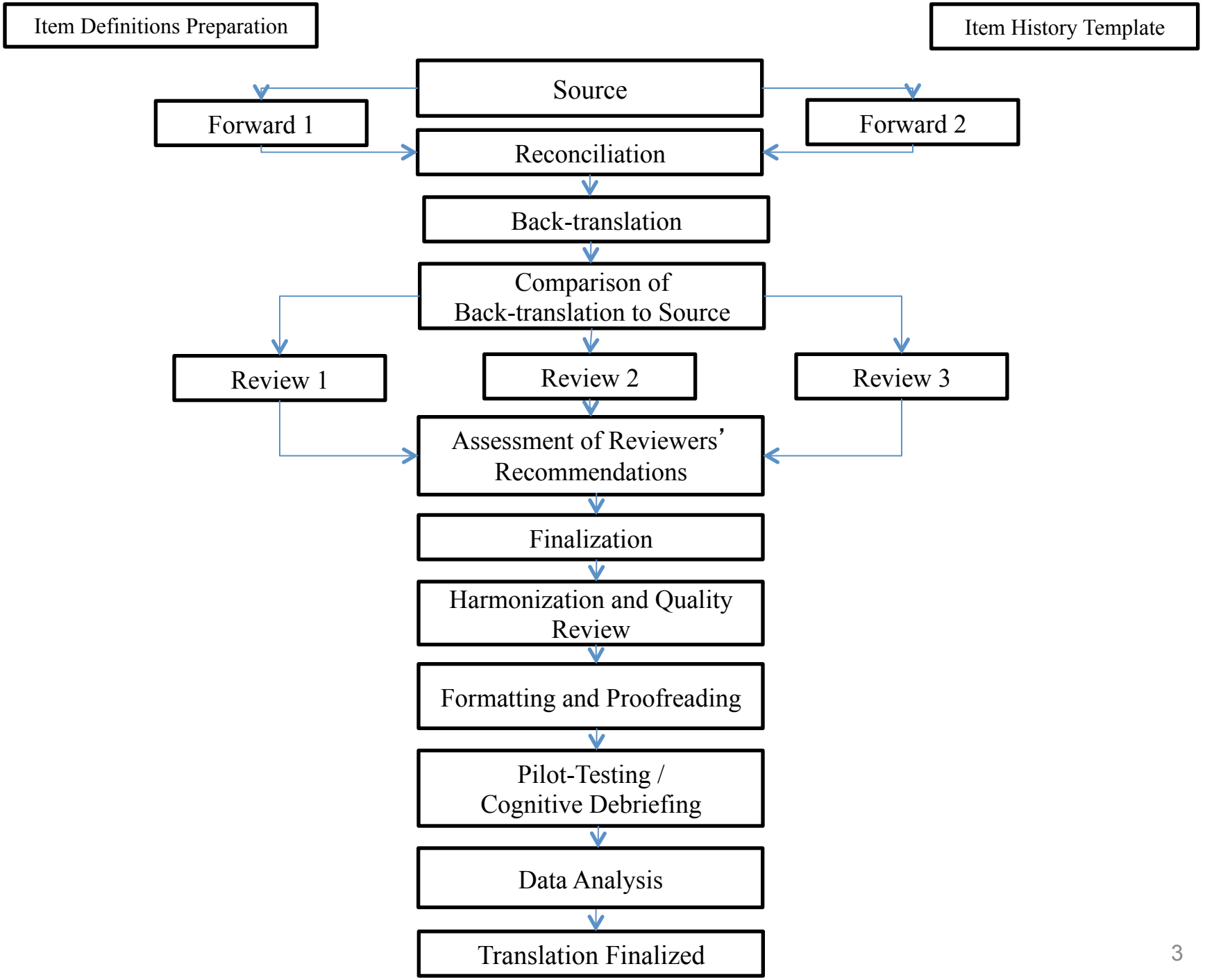
Sheraton Denver Downtown Hotel (Tower Court B)

October 29, 2011 (2:30-4:15pm)

Spanish Translation

Iterative process of forward translations, reconciliation, back translation, multiple reviews, and pre-test with cognitive debriefing (FACIT Translation Methodology). Harmonization across all languages and a universal approach to translation guided the process.

Each translated item was pre-tested and debriefed in the US with 5 Spanish-speaking subjects from the general population to ensure the translation was well understood and conceptually equivalent to the source.


Item Definitions Preparation
Item History Template
Source
Forward 1
Forward 2
Reconciliation
Back-translation
Comparison of
Back-translation to Source
Review 1
Review 2
Review 3
Assessment of Reviewers'
Recommendations
Finalization
Harmonization and Quality
Review
Formatting and Proofreading
Pilot-Testing /
Cognitive Debriefing
Data Analysis
Translation Finalized

Recruitment and Eligibility

General population Spanish-speaking Hispanics, through Toluna online panel

Eligibility:

- Hispanic Spanish speaker, age 18 or over
- Must have an average score <3.0 on Short Acculturation Scale for Hispanics (SASH)

Analyses

- Categorical confirmatory factor analyses
- 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 $P(u_i >= k) = \alpha_k + \mathbf{\beta_1 * ability}$

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

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

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

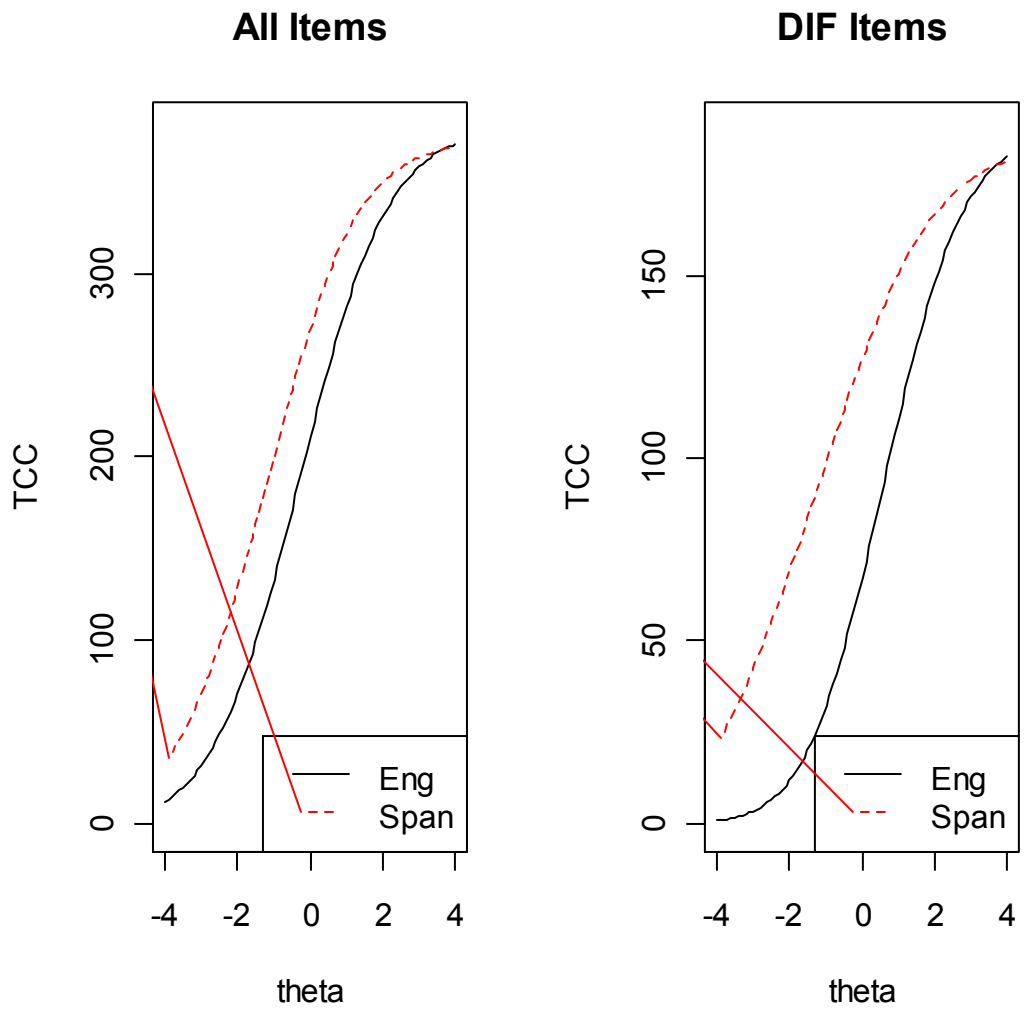
Sample Demographics

	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

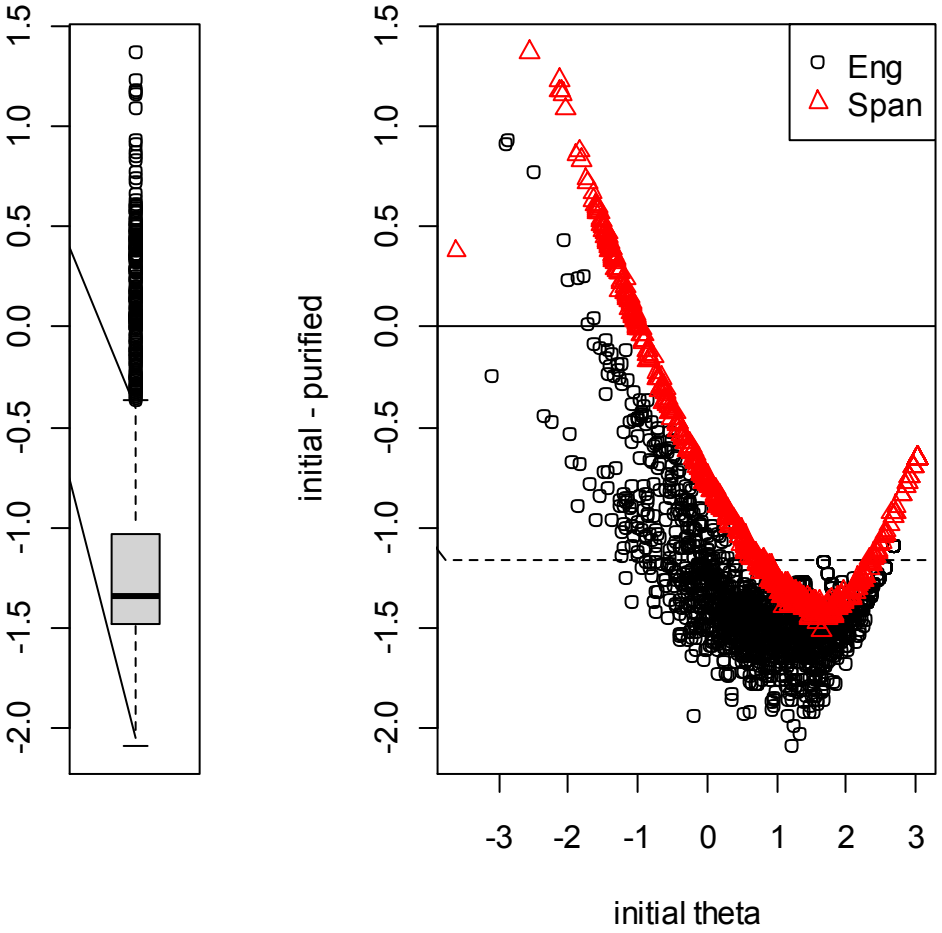
Impact of DIF on Test Characteristic Curves (TCCs)



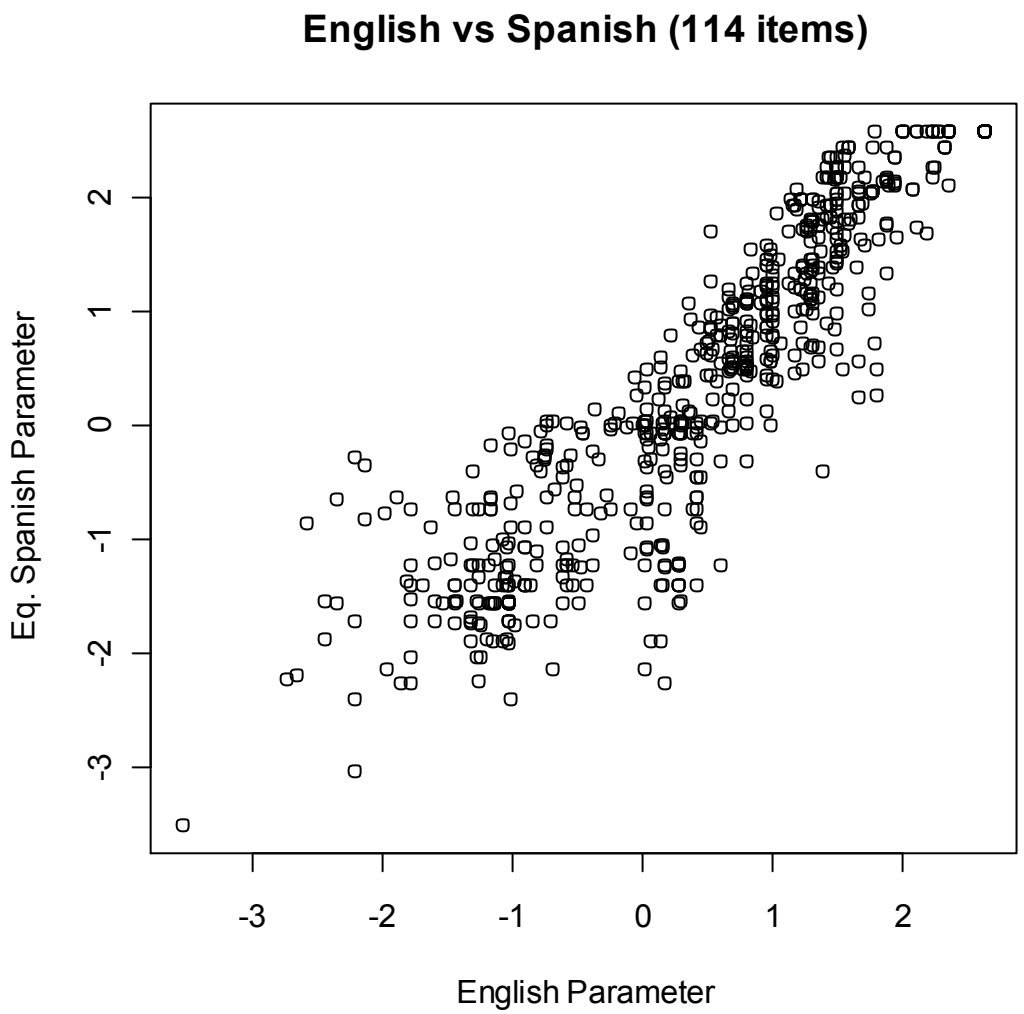
Stocking-Lord Method

- 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 minimized 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.

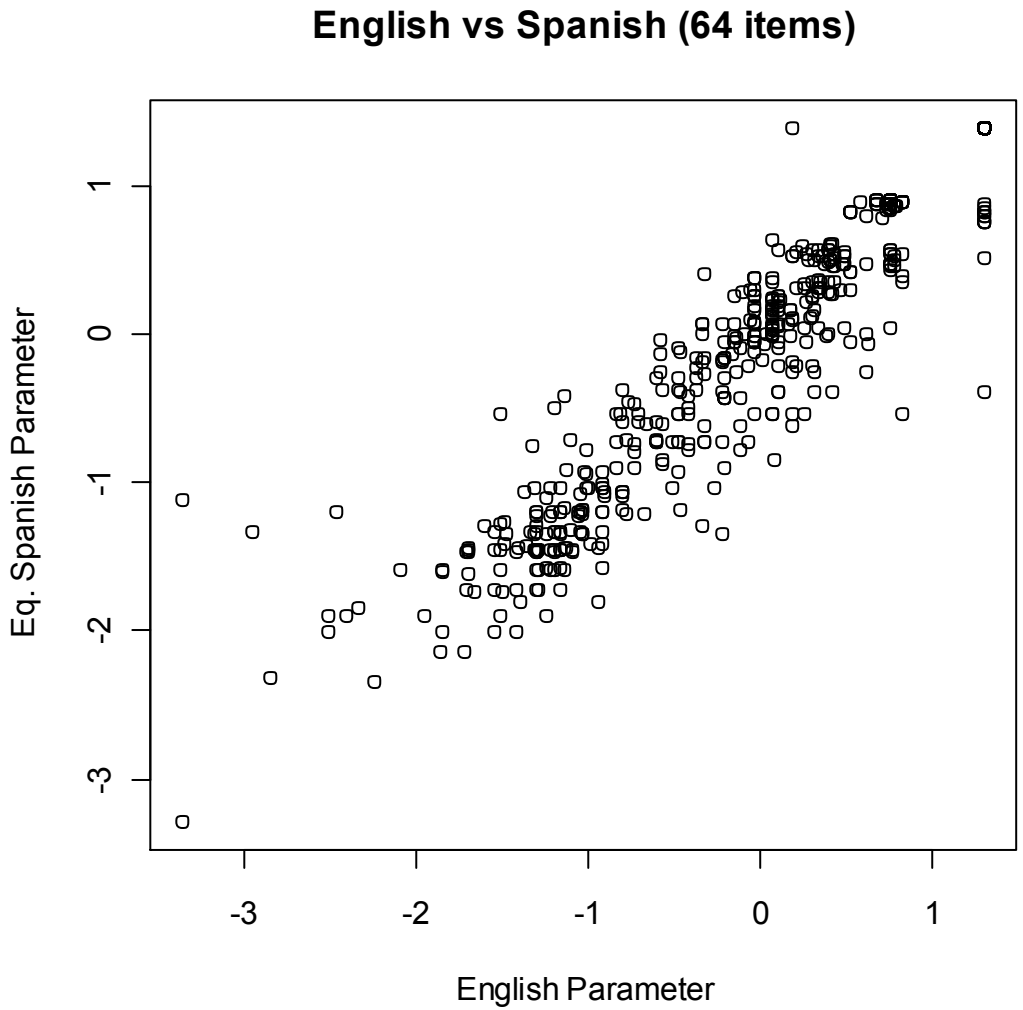
Impact of DIF at Individual Level



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)



Conclusions

- Hybrid model needed to account for language DIF
  - English calibrations for non-DIF items
  - Spanish calibrations for DIF items
- Verification based on English subgroup matched to Spanish in process