# Differential Item Functioning By Language on the PROMIS® Physical Functioning Items for Children and Adolescents

**Ron D. Hays, Ph.D.**,
Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 911 Broxton Ave, Los Angeles, CA 90024

**José Luis Calderón, M.D.**,
Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 911 Broxton Ave, Los Angeles, CA 90024

**Karen L. Spritzer, B.S.**,
Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 911 Broxton Ave, Los Angeles, CA 90024

**Steve P. Reise, Ph.D.**, and
UCLA Department of Psychology, 3583 Franz Hall, Los Angeles, CA 90095-1563

**Sylvia H. Paz, Ph.D.**
Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 911 Broxton Ave

## Abstract

**Purpose**—To assess the equivalence of self-reports of physical functioning between pediatric respondents to the English- and Spanish-language Patient Reported Outcomes Measurement Information System (PROMIS®) physical functioning item banks.

**Methods**—The PROMIS pediatric physical functioning item bank include 29 upper extremity items and 23 mobility items. A sample of 5,091 children and adolescents (mean age = 12 years old, range: 8–17; 49% male) completed the English-language version of the items. A sample of 605 children and adolescents (mean age = 12 years old, range: 8–17; 55% male; 96% Hispanic) completed the Spanish-language version of the items.

**Results**—We found language (English versus Spanish) differential item functioning (DIF) for 4 upper extremity items and 7 mobility items. Product-moment correlations between estimated upper extremity and mobility scores using the English versus the equated Spanish item parameters for Spanish-language respondents were 0.98 and 0.99, respectively. After excluding cases with

significant person misfit, we found DIF for the same 4 upper extremity items that had DIF in the full sample and for 12 mobility items (including the same 7 mobility items that had DIF in the full sample). The identification of DIF items between English and Spanish-language respondents was affected slightly by excluding respondents displaying person misfit.

**Conclusions**—The results of this study provide support for measurement equivalence of self-reports of physical functioning by children and adolescents who completed the English- and Spanish-language surveys. Future analyses are needed to replicate the results of this study in other samples.

The Patient-Reported Outcomes Measurement Information System (PROMIS®) is a National Institutes of Health initiative to develop state-of-the-science measures that assess functioning and well-being in the physical, mental and social domains of health. PROMIS goals include using these measures as indicators of health care outcomes that may guide reduction of health care disparities and improvement of population health in the U.S. [1]. The PROMIS project has developed a collection of item banks for adults, adolescents and children. The PROMIS measures are intended to be used to monitor health of populations and as outcome measures in intervention studies. The focus of this paper is the PROMIS physical functioning item banks for adolescents and children.

Physical functioning includes behavioral factors such as the capacity to engage in activities of daily living (performance) as well as musculoskeletal factors such as dexterity and strength. While physical functioning is inversely associated with age, there are congenital and acquired childhood conditions (e.g., cerebral palsy, spina bifida, seizure disorders, asthma) that may severely affect physical functioning. Given that chronic diseases such as obesity, sleep apnea and diabetes may also negatively affect physical functioning, the PROMIS pediatric physical functioning item banks are of great public health importance since these conditions were once mostly diagnosed in adults and are now being increasing diagnosed in children and adolescents.

The PROMIS pediatric physical functioning item banks consist of 29 upper extremity items and 23 mobility items. The development and evaluation of the English-language version of the pediatric physical functioning item banks was previously reported [2]. Responses to the Spanish-language version of the pediatric physical functioning item banks have not yet been evaluated for equivalence to those of the English-language version. It is important to assess whether responses to items in both language versions are equivalent or if differential item functioning (DIF) exists. DIF is present if the probability of selecting a particular response varies by group when controlling for the underlying level of the concept being measured [3]. For example, at the same level of underlying depression, women are more likely to report crying than men.

It is also important to evaluate the degree to which different people respond to items in a way that is consistent with the underlying model used to score the PROMIS pediatric physical functioning item banks (i.e., person fit). An example of a lack of person fit (misfit) in the adult PROMIS physical functioning item bank is someone reporting being able to run 5 miles without any difficulty and also reporting a little difficulty being out of bed most of the day [4]. Person misfit may be suggestive of response carelessness or cognitive errors due

to survey items being difficult to comprehend [5–6]. By evaluating the extent to which an individual's pattern of item responses is consistent with the scoring model, person fit is essentially a micro-level evaluation of DIF.

We evaluate person fit on the Patient Reported Outcomes Measurement Information System (PROMIS®) English and Spanish language versions of the pediatric (children and adolescents) physical functioning upper extremity and mobility item banks. We compare estimates of DIF before and after excluding respondents with significant person misfit. The public health significance of this project is underscored by its focus on Latinos, the fastest growing and youngest minority subgroup in the U. S. More than 20% of those 5 to 17 years of age in the U.S. speak a language other than English at home and 62% of these are Spanish speakers [7]. Ensuring equivalence between English and Spanish versions of the PROMIS item banks is crucial for guiding improvement of health care for Latinos and for informing public health stakeholders and policy makers interested in mitigating health care disparities.

## Methods

### Sample

**English—**The English-language sample was 5,091 children and adolescents 8–17 years old recruited from medical clinics in North Carolina and Texas, and from North Carolina community schools. The survey was administered on laptop computers and participants received a $10 gift card for their time and effort. The sample had a mean age of 12 and 49% were male. Forty percent of the overall sample was drawn from the schools and the other 60% was from the medical clinics targeting obesity, cancer, kidney disease, rehabilitation, rheumatic disease, asthma, and sickle cell disease. Item calibrations were reported previously [2].

**Spanish—**Spanish-speaking Hispanic adults who were members of the Greenfield/Toluna online panel [8] and had a child 8–17 years old were asked to complete sociodemographic questions about their child by computer and when a transition screen appeared, they were asked to give the computer to their child so the child could answer the physical functioning questions. A sample of 605 children and adolescents (mean age = 12 years old, range: 8–17; 55% male; 96% Hispanic based on parental report; see Table 1) was included. The Spanish-language sample of children had an average score on the Short Acculturation Scale for Hispanic youth [9] of 2.6 (SD = 1.2), indicating low levels of acculturation. The study participants received nominal incentives from the online panel company to complete the survey (value did not exceed $10).

### Spanish Translation

All items were translated using the FACIT translation methodology [10] that is consistent with the International Society for Pharmacoeconomic and Outcomes Research guidelines [11]. A universal Spanish translation was created using an iterative process of two simultaneous forward translations, reconciled single translation, back-translation by a native English-speaking translator fluent in Spanish, back-translation review, review by three experts who are native Spanish speakers, pre-finalization review, revision by a native

Spanish-speaker, cognitive debriefing with 5 native Spanish-speaking children and adolescents, and finalization. Applying a universal approach to translations results in one version of the same language and requires that translators from different regions or dialects contribute to the process. The process aims to avoid colloquial expressions and enable comparisons across subgroups of Spanish speaking populations.

### Analysis Plan

A psychometric evaluation of the PROMIS pediatric physical functioning items in the English-language sample was reported previously [2]. We assessed unidimensionality (the items represent a single construct) of the items in the Spanish-language sample by fitting a one-factor categorical confirmatory factor analysis model in Mplus Version 6 [12]. We evaluated model fit using the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA); CFI>=0.95 and RMSEA<0.06 are considered acceptable [13]. Local independence was evaluated by inspection of residual correlations among items in the one-factor model with correlations of 0.20 suggestive of a violation of the assumption of local independence (items being unrelated after conditioning on the single factor).

We assessed language DIF using ordinal logistic regression with item response theory (IRT) trait scores estimated from DIF-free "anchor" items (iterative purification) as the conditioning variable using *lordif version 0.3–3* software [14]. A pseudo R-squared difference of <0.02 between nested models was used to identify potential anchor items. For items with DIF, we evaluated whether they had uniform DIF, in which DIF is in the same direction across the entire continuum or non-uniform DIF where the probability of endorsing an item response is higher for one group at lower levels of the concept but higher for the other group at higher levels of the concept. We put the Spanish-language item parameters (slopes and thresholds) on the same metric as the English-language parameters using Stocking and Lord [15] linking constants.

First, IRT scores were estimated using a graded response model. Then, these scores were used as a conditioning variable in an ordinal logistic analysis. We estimated three models for upper extremity (used as example below) and mobility:

$$Ordinal\ LR\ ; p\ (item\ response\ is\ correct) = \beta_0 + \beta_1\ (upper\ extremity) \quad (1)$$

$$Ordinal\ LR\ ; p\ (item\ response\ is\ correct) = \beta_0 + \beta_1\ (upper\ extremity) + \beta_2(group) \quad (2)$$

$$Ordinal\ LR\ ; p\ (item\ response\ is\ correct) = \beta_0 + \beta_1\ (upper\ extremity) + \beta_2(group) + \beta_3\ (upper\ extremity * group)$$

$$(3)$$

The comparison of Model 3 versus Model 1 indicates whether there is any DIF. The comparison of Model 3 versus Model 2 tests if there is non-uniform DIF, and the comparison of Model 2 versus Model 1 indicates uniform DIF. Three parameters are estimated from the models: β, pseudo $R^2$, and the likelihood-ratio $\chi^2$ statistics. We use the likelihood-ratio $\chi^2$ statistic as the DIF detection criterion ($\alpha < 0.01$) and the pseudo $R^2$ measure as the measure of magnitude ( 0.02).

Once DIF items are identified, we evaluate the magnitude of DIF using test characteristic curves separately for all items in a scale and for the items identified as having DIF. We assess DIF at the individual level by plotting trait-level estimates ignoring DIF versus trait-level estimates accounting for DIF. DIF is considered noteworthy if it equals or exceeds a small effect size (i.e., 0.20 SD).

We estimated person fit using the standardized Z(L) fit index. Large negative Z(L) values indicate misfit [16]. Large positive Z(L) values indicate response patterns that are higher in likelihood than the model predicts. To produce a potentially more powerful test of DIF, we again estimated it after removing people who displayed statistically significant misfit (p < 0.05).

We estimate readability of items using the Flesch-Kincaid readability formula [17] to see if items with DIF require higher education to understand than other items. Most formulae used to evaluate the readability of written text are based on the number of syllables per word and the number of words per sentence. The Flesch–Kincaid readability index yields an estimate of the grade level needed to read and comprehend the material. Readability estimation for survey items is challenging because the items do not necessarily conform to the grammatical structure of complete sentences or questions. Furthermore, response options influence readability but are not sentences and were excluded from readability estimates in this study.

## Results

The mean PROMIS upper extremity (mobility) scores using the existing (English-language) parameters was 50 (50) for the English-language sample and 44 (48) for the Spanish-language sample.

### Upper Extremity

The one-factor categorical confirmatory factor analysis of the 29 upper extremity items fit the data well in the Spanish-language sample (CFI, = 0.998; RMSEA, = 0.036). Standardized factor loadings ranged from 0.824 ("I used a pencil with a special grip to write") to 0.962 ("I could dial a phone") and were all statistically significant at p =0.000 (Table 2). The largest residual correlation (r = 0.038) was between "I could pour a drink from a full pitcher" and "I could dry my back with a towel."

We found 4 upper extremity items with DIF between the Spanish and English responses: 1) F1_UE3: "I could hold an empty cup"; 2) F3_UE9: "I could pull open heavy doors"; 3) F4_UE1: "I could open a jar by myself"; and 4) F4_UE10: "I could pour a drink from a full pitcher." All of the 4 items displayed uniform DIF. The mean Flesch-Kincaid [17] estimated

grade level to read these 4 item stems is 1.6 (versus 1.4 for all 29 upper extremity items). The item parameters estimates for the English and Spanish-language respondents for the 4 upper extremity items with DIF are shown in Table 3.

The impact on the total score for the DIF items is noticeable because the curves in the right side of Figure 1 are not superimposed on one another but the curves on the left side indicate DIF impact was minimal when all items were included. As seen in the scatterplot of upper extremity scores estimated using English-language parameters (x axis) by the difference between this estimate and the score estimated taking into account DIF (y axis), the largest DIF impact at the individual level for upper extremity was about 0.30 of a standard deviation (Figure 2). Stocking-Lord linking constants were used to transform linearly the Spanish item parameter estimates to the English metric (Spanish slopes = Spanish calibrated slope/ 1.377145; Spanish thresholds = (Spanish calibrated threshold * 1.37714) −2.372854). The product-moment correlation between trait-level estimates using the English versus the equated Spanish parameters for the Spanish sample was very high (Figure 3) at r = 0.98 (intraclass correlation = 0.96).

We identified 53 English- and 38 Spanish-language cases with significant misfit (p < 0.05) for the 29 upper extremity items. After excluding these cases, we found DIF for the same 4 items as in the full scale.

### Mobility

A one-factor categorical confirmatory factor analysis of the 23 mobility items in the Spanish-language sample fit the data well (CFI = 0.996; RMSEA = 0.054). Standardized factor loadings ranged from 0.815 ("I could run a mile") to 0.967 ("I could walk across the room") and were all statistically significant at p =0.000. (Table 4) The largest residual correlation was 0.042 between "I have been physically able to do the activities I enjoy most" and "I could do sports and exercise that other kids my age could do."

We found that 7 of the 23 mobility items had language DIF: 1) F1_MOB2: "I could ride a bike"; 2) F1_MOB3 "I could do sports and exercise that other kids my age could do"; 3) F2_MOB1: "I could run a mile"; 4) F2_MOB4: "I could walk upstairs without holding on to anything"; 5) F3_MOB5: "I used a walker, cane or crutches to get around"; 6) F3_MOB10: "I could turn my head all the way to the side"; and 7) F4_MOB4: "I could keep up when I played with other kids." Two of these items displayed non-uniform DIF (F3_MOB5 and F3_MOB10). The mean Flesch-Kincaid estimated grade level to read these 7 item stems is 2.0 (same as 2.0 for all 23 mobility items). The item parameters estimates for the English and Spanish-language respondents for the 7 mobility items with DIF are shown in Table 5.

The impact for the DIF items was noticeable (right-hand side of Figure 4); some small impact is seen at trait-levels slightly below the average when all items are included (left-hand side of Figure 4). The scatterplot of mobility scores estimated using English-language parameters (x axis) compared to the difference between this estimate and the score estimated taking DIF into account (Figure 5) shows that the largest DIF impact at the individual level for mobility was about 0.40 of a standard deviation (Figure 5). Stocking-Lord linking constants were used to transform linearly the Spanish item parameter estimates to the

English metric (Spanish slopes = Spanish calibrated slope/1.725001; Spanish thresholds = (Spanish calibrated threshold * 1.725001) −1.693995). The product-moment correlation between trait-level estimates using the English versus the equated Spanish parameters for the Spanish sample was very high (Figure 6) at r = 0.99 (intraclass correlation = 0.96).

We identified 84 English- and 37 Spanish-language cases with significant person misfit for the 23 mobility items. After excluding these cases, we found 12 mobility items with DIF including items 1–7 above plus five other items (F1_MOB1: I have been physically able to do the activities I enjoy most; F1_MOB6: I could go up one step; F2_MOB6: I could get up from a regular toilet; F3_MOB9: I could get up from the floor; and F4_MOB7: I used a wheelchair to get around).

## Discussion

As the U.S. Latino subgroup continues to grow, it is important to ensure that physical functioning survey measures perform equivalently in Spanish-language and English-language children and adolescents. We found some items in the PROMIS pediatric physical functioning items had language (English versus Spanish) DIF. This means that people with the same level of underlying physical functioning respond differently to these items if they respond to the English or Spanish-language version of the survey. Impact at the individual level for some respondents exceeded a small effect size (0.20 SD). This is potentially problematic because one of the original goals of the PROMIS initiative was to develop item banks that could be used across different subgroups. However, language DIF on estimated scores was inconsequential. Product-moment correlations between estimated upper extremity and mobility scores using the English versus equated Spanish item parameters for Spanish-language respondents were 0.98 and 0.99, respectively.

One of the advances of PROMIS® is the use of computer adaptive testing (CAT) to measure health outcomes. In CAT items are selectively administered depending on a respondent's position on the latent trait continuum. Thus, with CAT typically only a subset of the items in the bank is used to arrive at a trait-level estimate for an individual and the impact of DIF items in the bank will vary depending on the total number of items administered and whether the items with DIF are selected. Hence, without knowing the item set to be used for a respondent a-priori, the impact of DIF among the items in a bank is impossible to predict. Language-specific item parameters can be used for items with DIF in estimating scores. But, the impact of DIF on CAT estimated scores was inconsequential as estimates of upper extremity and mobility for those who completed the Spanish-language survey were similar whether English parameters or Spanish-specific parameters were used for items displaying DIF (Figures 3 and 6).

DeWitt et al. [2] suggested 8-item short forms for upper extremity and mobility. Of the 8 upper extremity items they recommended (F2_UE2, F2_UE3, F2_UE4, F3_UE7, **F3_UE9**, F3_UE11, **F4_UE1**, **F4_UE10**), 3 of them were among the 4 items with language DIF (bolded). Of the 8 short-form mobility items in their suggested short form (F1_MOB1, **F1_MOB3**, **F2_MOB4**, F2_MOB7, F3_MOB3, F3_MOB8, F3_MOB9, **F4_MOB4**), 3 of them were among the 7 items with language DIF (bolded). The product-moment correlations

between estimated scores using the English-language parameters for the 8 upper extremity items and 8 mobility items versus using equated Spanish parameters were 0.997 and 0.999, respectively.

Including persons who answered items in a manner that does not correspond to the underlying IRT model (person misfit) tends to reduce the item discrimination parameter estimates because person misfit reflects inconsistency in responding to different items in a unidimensional scale. The effect of excluding respondents displaying substantial person misfit on DIF was to alter some of the items identified as having DIF between English and Spanish-language responses to the physical functioning item banks, but the overall level of DIF was essentially unchanged.

### Limitations

It is important to acknowledge limitations of the study. Although they were instructed to give the computer to their child to answer the survey, it is possible that some parents completed the questions rather than following the directions. In addition, the results of this study may not generalize to the U.S. Spanish-language pediatric and adolescent population in general. Convenience internet panels such as those in the current study are known to differ in education and other characteristics from those in the adult general population [18]. These differences may affect responses to the PROMIS physical functioning items. In addition, the small amount of DIF detected by language might have been due to differences between the samples on characteristic other than language. Matching the Spanish-language and English-language samples on variables such as age and gender could have reduced or eliminated DIF altogether. Moreover, future studies should target individuals that are more representative of those whose primary language is Spanish in the U.S. Finally, future analyses are needed to examine the variability in patterns of person misfit to help elucidate the lack of impact on DIF in this study.

### Acknowledgments

### References

1. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Young S, et al. Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005–2008. Journal of Clinical Epidemiology. 2010; 63(11):1179–1194. [PubMed: 20685078]

2. DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, Lai JS, Yeatts KB, DeWalt DA. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. Journal of Clinical Epidemiology. 2011; 64(7):794–804. [PubMed: 21292444]

3. Yang FM, Heslin KC, Mehta KM, Yang C-W, Oceptek-Welikson K, Kleinman M, Morales LS, Hays RD, Stewart AL, Mungas D, Jones RN, Teresi JA. A comparison of item response theory-based methods for examining differential item functioning in object naming test by language of assessment among older Latinos. Psychological Test and Assessment Modeling. 2011; 53:440–460. [PubMed: 23471423]

4. Hays, RD. Response 1 to Reeve's chapter: Applying Item response theory for questionnaire evaluation. In: Madans, J.Miller, K.Maitland, A., Willis, G., editors. Question Evaluation Methods: Contributing to the Science of Data Quality. Hoboken, New Jersey: Wiley & Sons Inc; 2011. p. 125-135.

5. Paz SH, Spritzer KL, Morales LS, Hays RD. Evaluation of the Patient-Reported Outcomes Information System (PROMIS®) Spanish Physical Functioning Items. Quality of Life Research. 2013; 22(7):1819–1830. [PubMed: 23124505]

6. Reise SP. A comparison of item- and person-fit methods of assessing model-data fit in IRT. Applied Psychological Measurement. 1990; 14:127–137.

7. Skinner, C., Wright, VR., Aratani, Y., et al. English Language Proficiency, Family Economic Security, and Child Development. National Center for Children in Poverty. 2010. website. http://www.nccp.org/publications/pub_948.html

8. Toluna Group Ltd. https://us.toluna.com/

9. Barona A, Miller JA. Short acculturation scale for Hispanic youth (SAS-Y): A preliminary report. Hispanic Journal of Behavioral Sciences. 1994; 16:155–162.

10. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. Eval Health Prof. 2005; 28(2):212–32. [PubMed: 15851774]

11. Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendation on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR patient reported outcome translation and linguistic validation good practice task force report. Value in Health. 2008; 12:430–440. [PubMed: 19138309]

12. Muthén, LK., Muthén, BO. Mplus User's Guide. Sixth. Los Angeles, CA: Muthén & Muthén; p. 1998-2011.

13. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS). Medical Care. 2007; 45:S22–31. [PubMed: 17443115]

14. Choi S, Gibbons L, Crane P. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. Journal of Statistical Software. 2011; 39(8):1–30.

15. Stocking ML, Lord FM. Developing a common metric in item response theory. Applied Psychological Measurement. 1983; 7(2):201–210.

16. Drasgow F, Levine MV, Williams EA. Appropriateness measurement with polytomous item response models and standardized indices. The British Journal of Mathematical and Statistical Psychology. 1985; 38:67–68.

17. Kincaid, J., Fishburne, R., Rodgers, R., Chissom, B. Derivation of new readability formulas for Navy enlisted personnel (Branch Report 8–75). Millington, TN: Chief of Naval Training; 1975.

18. Hays RD, Liu H, Kapteyn A. Use of Internet panels to conduct surveys. Behav Res Methods. 2015 Sep; 47(3):685–90. [PubMed: 26170052]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
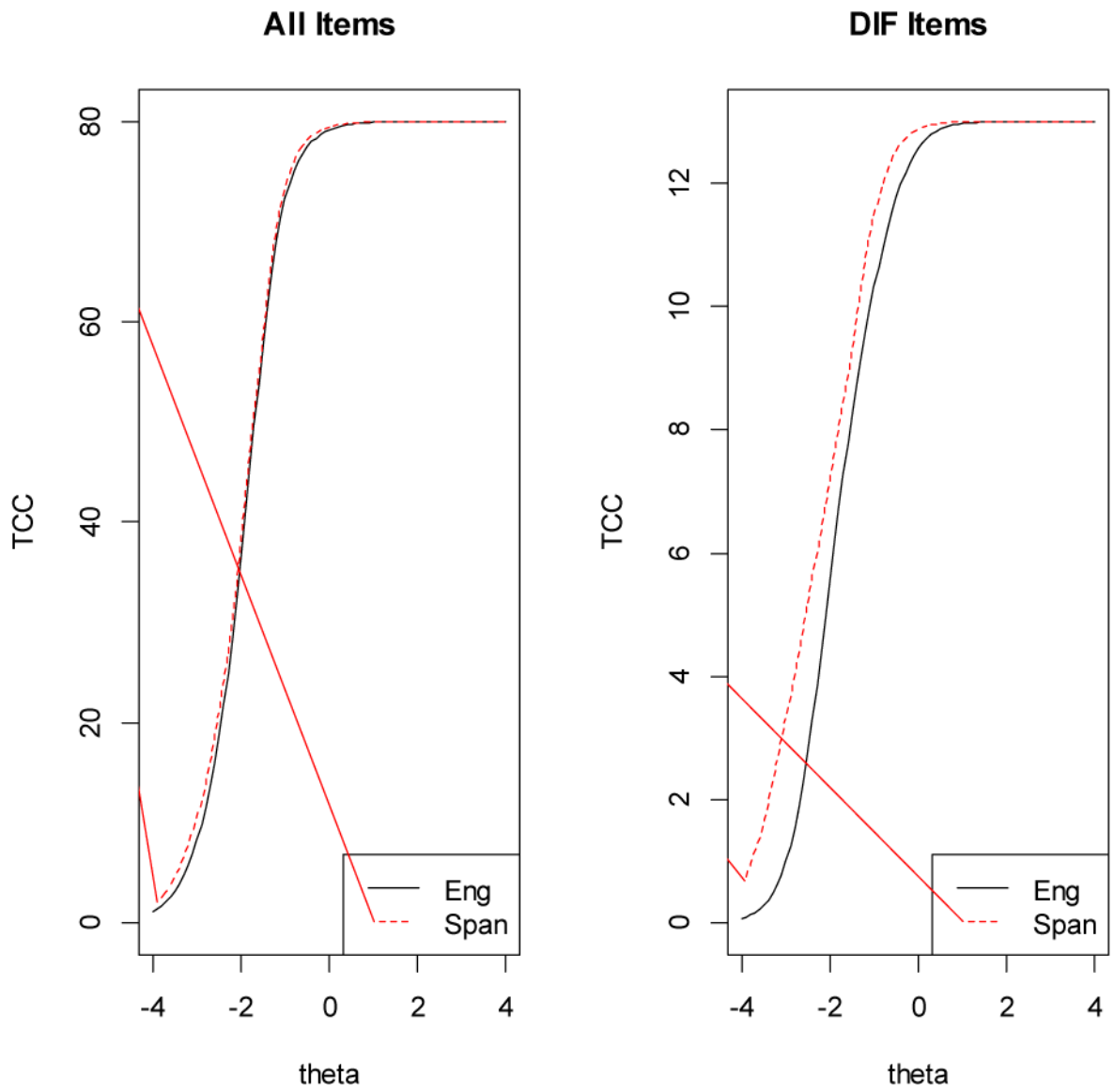
**Figure 1.**
DIF Impact for Upper Extremity Items

**Figure 2.**
DIF Impact at Individual Level – Upper Extremity Items

**Figure 3.**
Correlation of CAT-based Trait-level Estimates Using English (x-axis) and Spanish (y-axis)
Parameters for All 29 Upper Extremity Items in Spanish Sample (n=605)
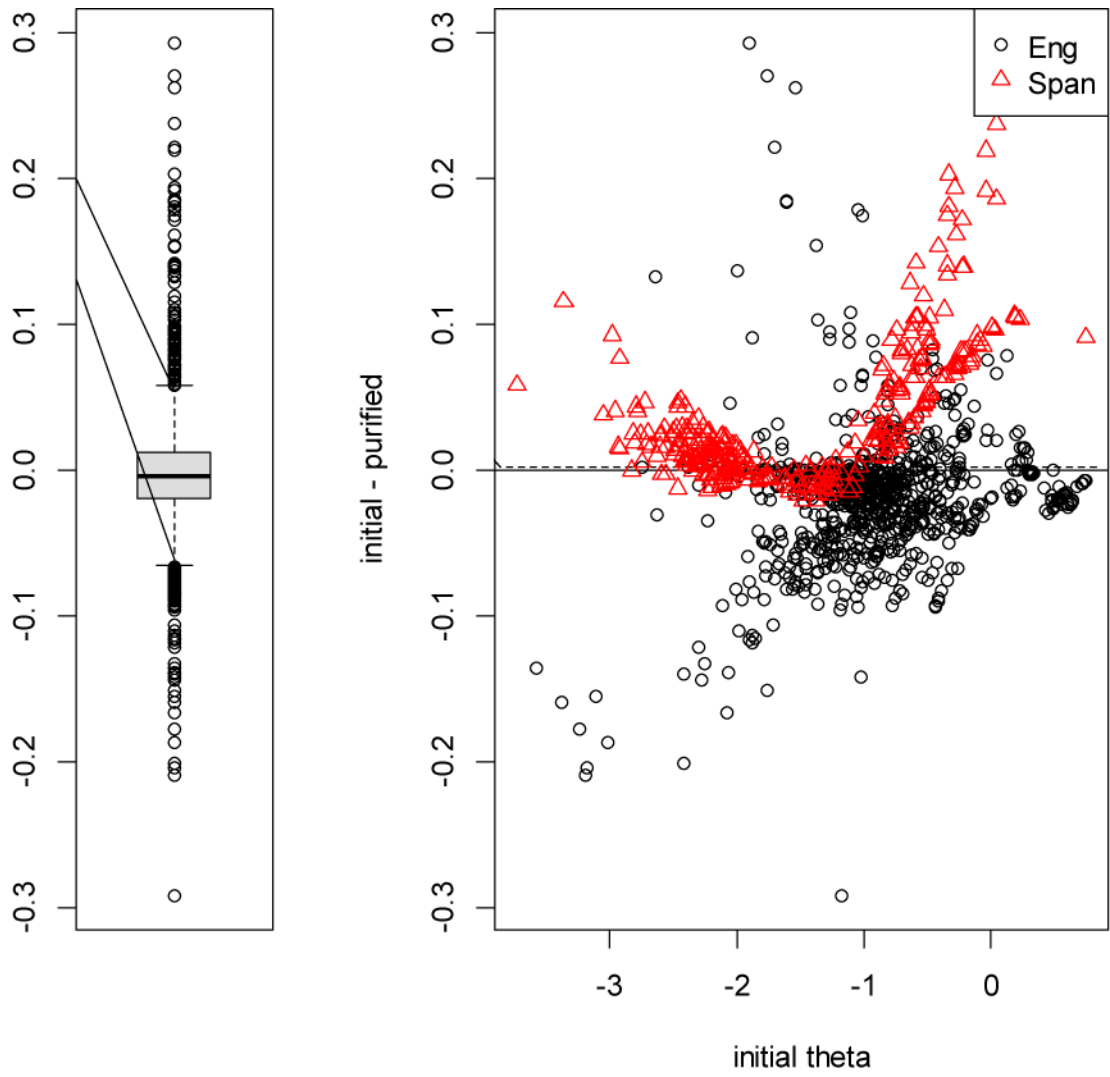
**Figure 4.**
DIF Impact for Mobility Items

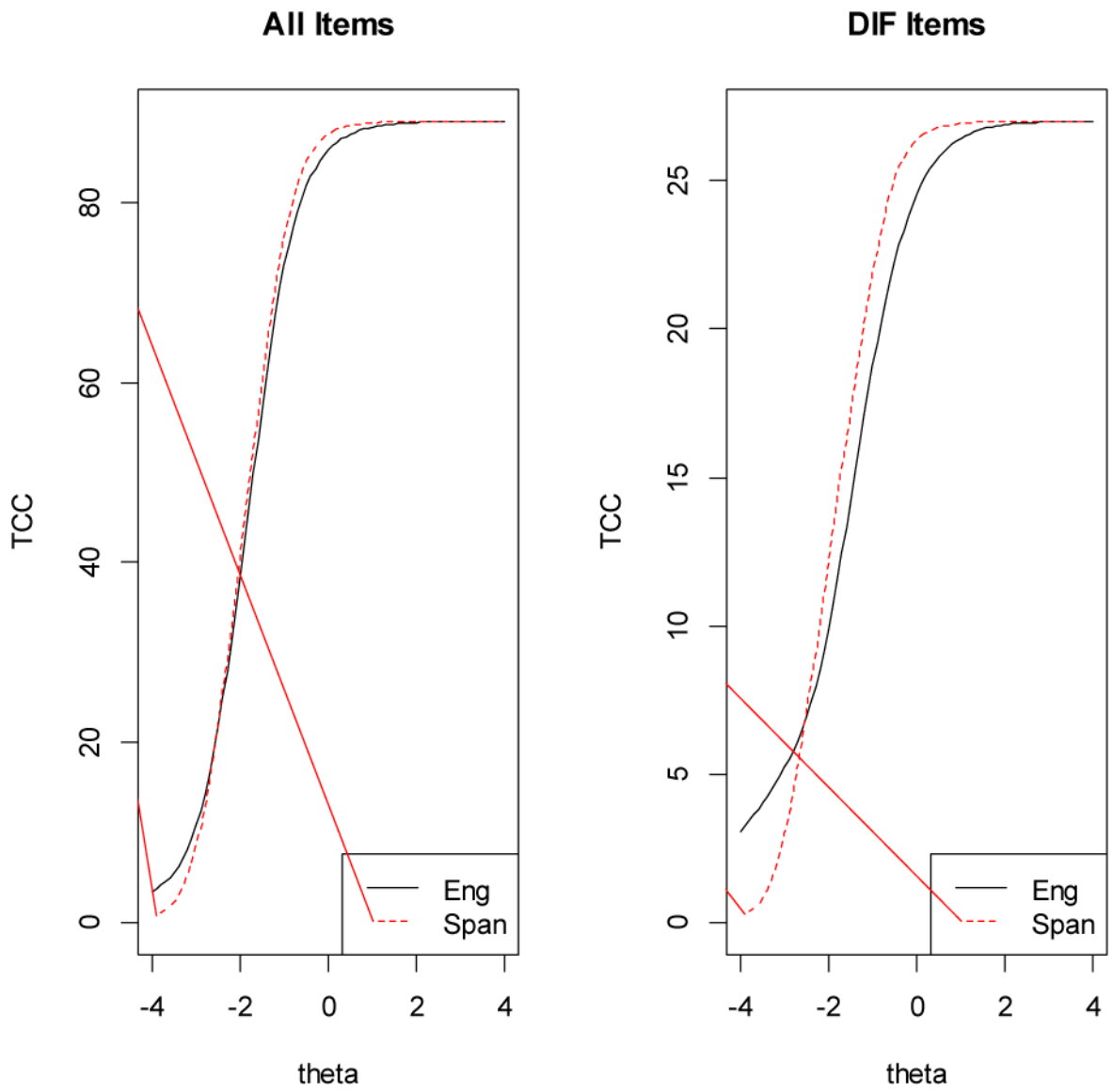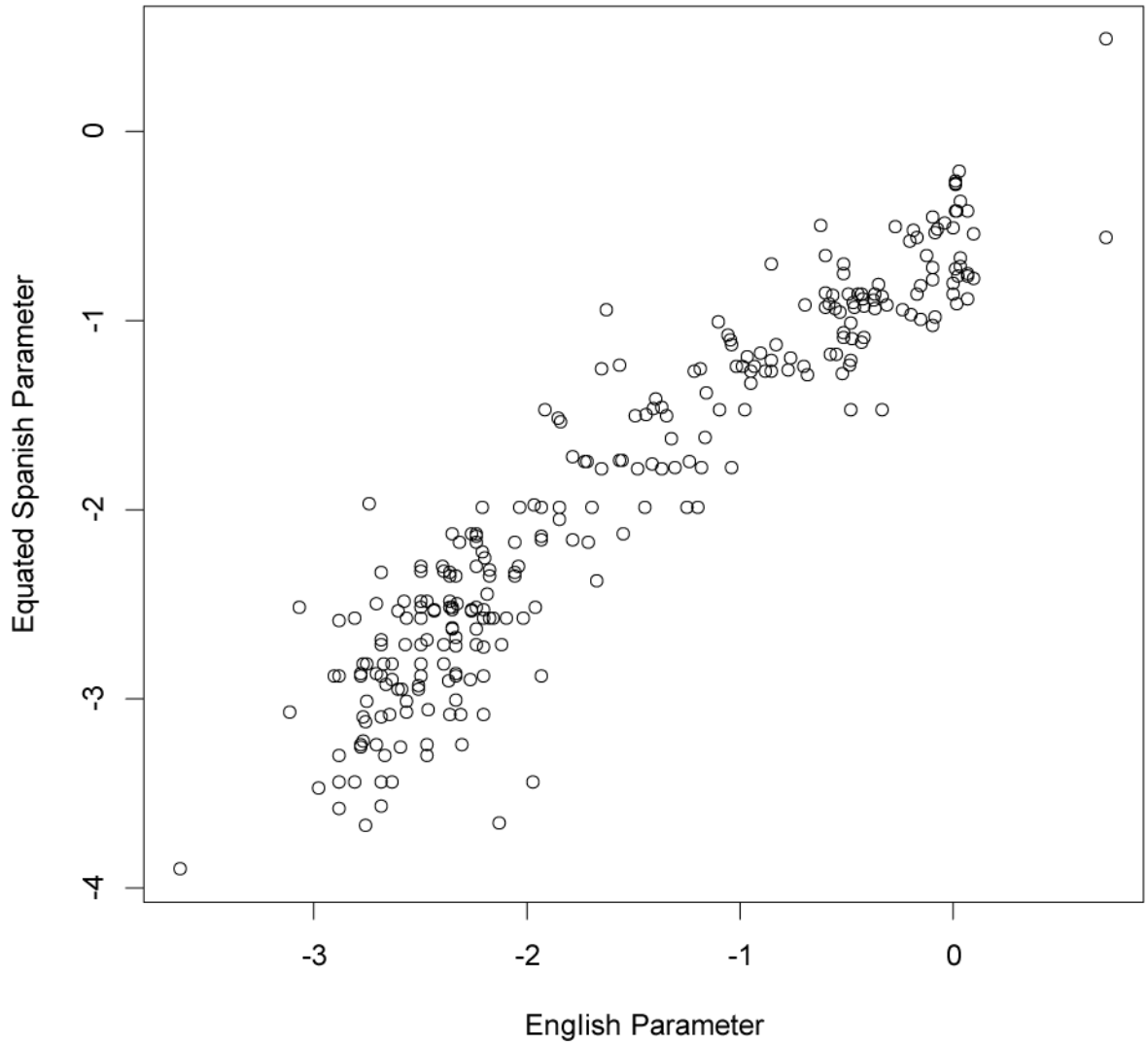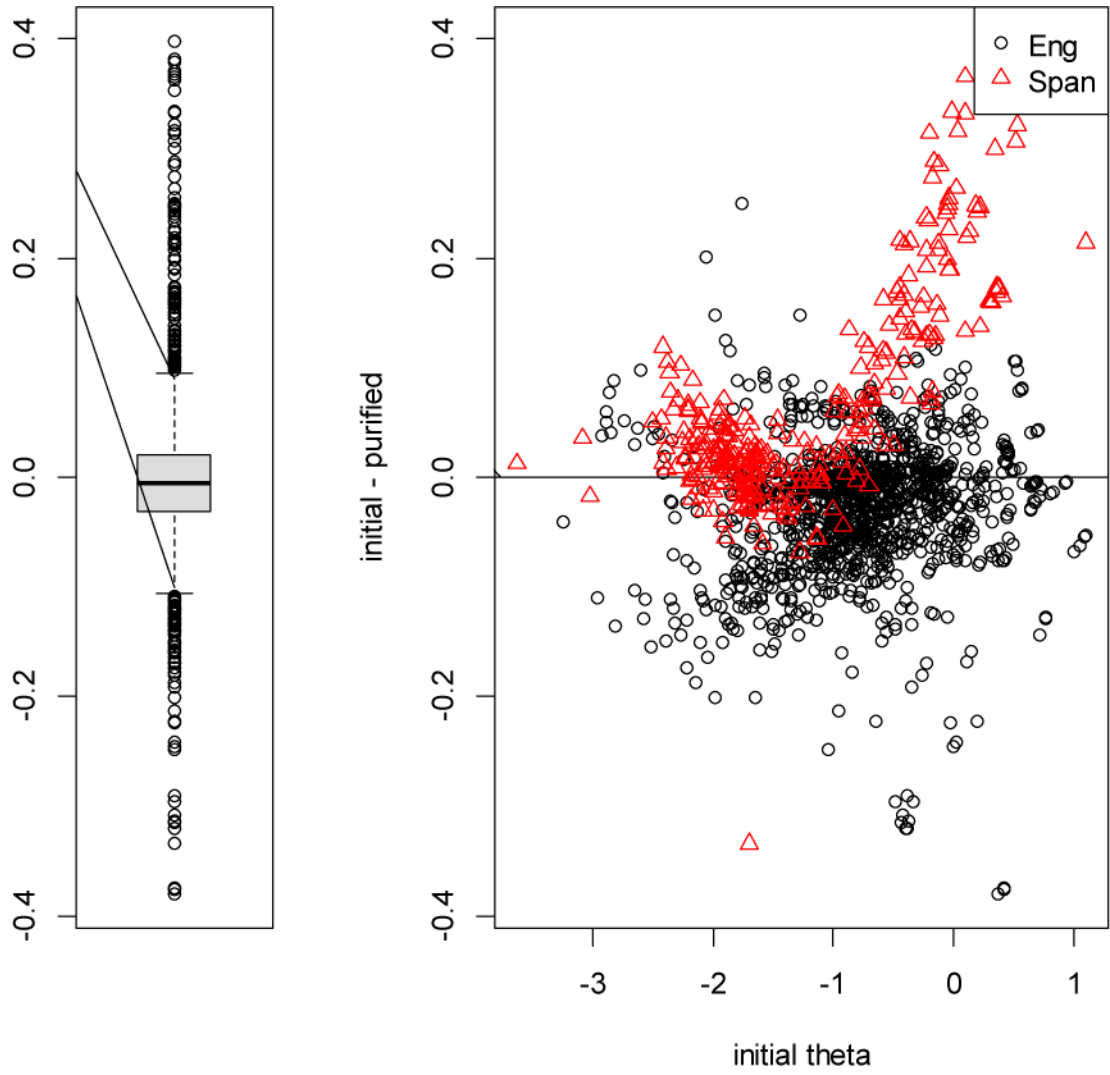**Figure 5.**
DIF Impact at Individual Level – Mobility Items
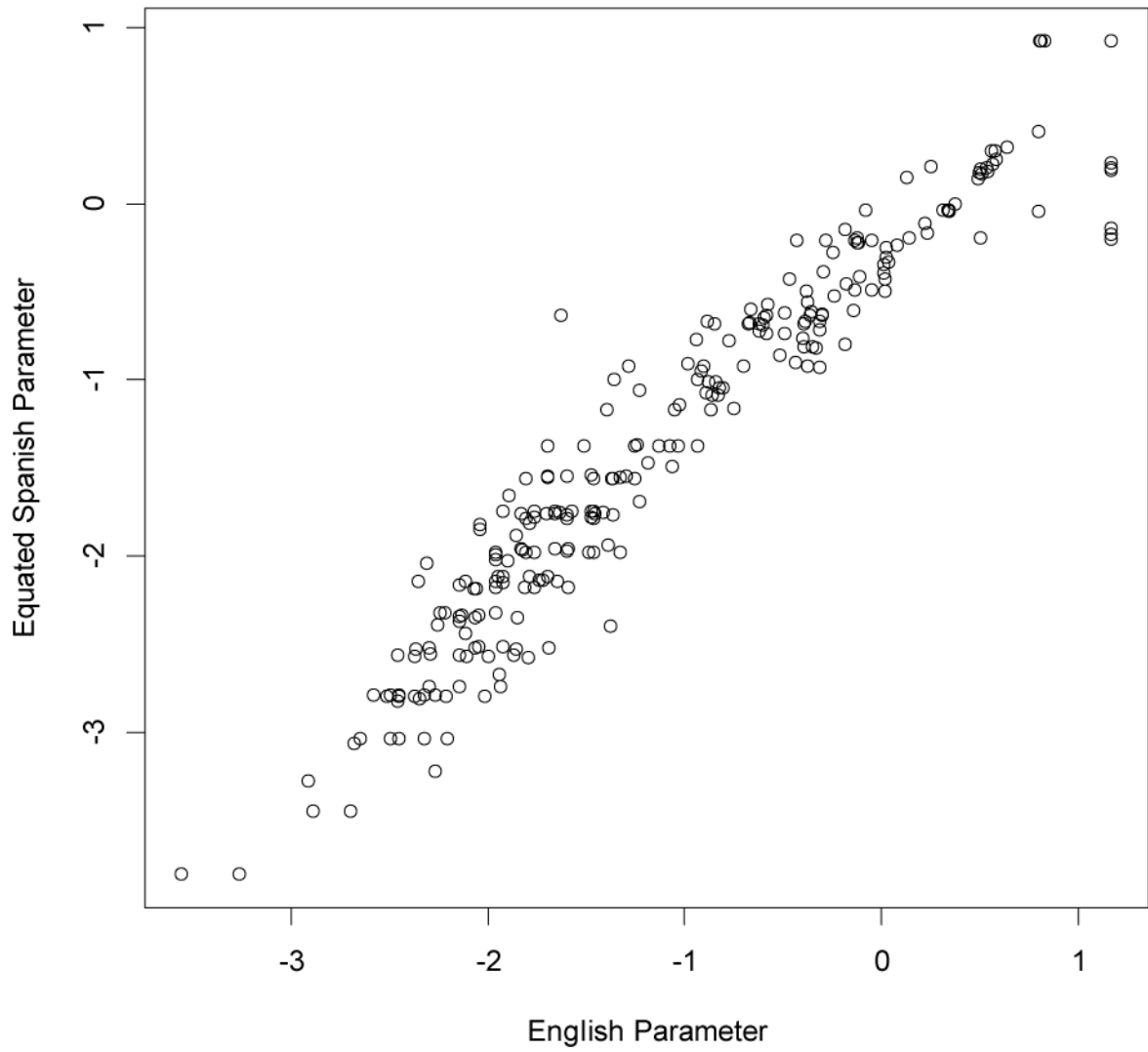
**Figure 6.**
Correlation of CAT-based Trait-level Estimates Using English (x-axis) and Spanish (y-axis)
Parameters for All 23 Mobility Items in Spanish Sample (n=605)

**Table 1**

Sociodemographic and Clinical Characteristics of PROMIS® Pediatric Spanish-Language Physical Function Sample (n=605)

| | *n* | *%* |
|---|---|---|
| Age categories | | |
|   8–12 | 302 | 50 |
|   13–17 | 303 | 50 |
| Gender | | |
|   Male | 330 | 55 |
|   Female | 275 | 45 |
| Race/Ethnicity | | |
|   Hispanic | 580 | 96 |
|   Non-Hispanic White | 21 | 3 |
|   Non-Hispanic Black or African American | 3 | 1 |
|   Non-Hispanic American Indian/Alaskan Native | 1 | 0.2 |
| Education | | |
| Highest grade completed for those not currently attending school (n = 16) | | |
|   Never attended school | 3 | 19 |
|   7th grade | 1 | 6 |
|   8th grade | 1 | 6 |
|   9th grade | 2 | 13 |
|   10th grade | 2 | 13 |
|   HS graduate | 7 | 44 |
| Grade currently in for those attending school (n = 589) | | |
|   Kindergarten | 1 | 0.2 |
|   1st grade | 30 | 5 |
|   2nd grade | 38 | 6 |
|   3rd grade | 54 | 9 |
|   4th grade | 57 | 10 |
|   5th grade | 35 | 6 |
|   6th grade | 41 | 7 |
|   7th grade | 42 | 7 |
|   8th grade | 63 | 11 |
|   9th grade | 54 | 9 |
|   10th grade | 71 | 12 |
|   11th grade | 54 | 9 |
|   12th grade | 49 | 8 |
| Comorbidities | | |
| Ever told you have … | *n* | % |
|   Arthritis or rheumatism | 2 | 0.3 |

| | n | % |
|---|---|---|
| Asthma | 83 | 14 |
| Diabetes or high blood sugar or sugar in urine | 8 | 1 |
| Cancer other than non-melanoma skin cancer | 23 | 4 |
| Depression | 15 | 2 |
| Anxiety | 21 | 3 |
| Alcohol or drug problem | 11 | 2 |
| Sleep disorder | 38 | 6 |
| Multiple sclerosis | 2 | 0.3 |
| Epilepsy | 16 | 3 |
| Muscular dystrophy | 17 | 3 |
| None of the above | 410 | 68 |

**Table 2**

Categorical Factor Loadings For 29 Upper Extremity Items In Spanish Sample

| Item | English item wording; In the past 7 days. | Spanish item wording: En los Últimos 7 días. | Factor Loading | Standard Error | t-statistic |
|---|---|---|---|---|---|
| F1_UE1 | I could tie shoelaces by myself. | Pude atarme los cordones de los zapatos sin ayuda. | 0.926 | 0.010 | 93.529 |
| F1_UE2 | I could put on my clothes by myself. | Pude ponerme la ropa sin ayuda. | 0.931 | 0.011 | 84.848 |
| F1_UE3 | I could hold an empty cup. | Pude sostener un vaso vacío. | 0.955 | 0.007 | 145.933 |
| F1_UE4 | I could pull on and fasten my seatbelt. | Pude tirar de (jalar) mi cinturón de seguridad y abrochármelo. | 0.941 | 0.007 | 129.303 |
| F2_UE1 | I could move my hands or fingers. | Pude mover las manos o los dedos. | 0.945 | 0.007 | 131.131 |
| F2_UE2 | I could put on my shoes by myself. | Pude ponerme los zapatos sin ayuda. | 0.955 | 0.007 | 142.111 |
| F2_UE3 | I could button my shirt or pants. | Me pude abotonar la camisa o los pantalones. | 0.956 | 0.007 | 143.050 |
| F2_UE4 | I could use a mouse or touch pad for the computer. | Pude usar el ratón o el *touch pad* de la computadora. | 0.935 | 0.010 | 94.535 |
| F2_UE5 | I could lift a cup to drink. | Pude levantar un vaso para beber. | 0.954 | 0.007 | 144.531 |
| F2_UE7 | I could cut paper with scissors. | Pude cortar papel con unas tijeras. | 0.959 | 0.006 | 156.471 |
| F2_UE8 | I could wash my face with a cloth. | Pude limpiarme la cara con un paño. | 0.958 | 0.007 | 143.198 |
| F3_MOB4 | I needed help with a bath. | Necesité ayuda para bañarme. | 0.877 | 0.014 | 62.871 |
| F3_UE2 | I could put on my socks by myself. | Pude ponerme los calcetines sin ayuda. | 0.960 | 0.006 | 154.568 |
| F3_UE3 | I could put toothpaste on my toothbrush by myself. | Pude poner pasta dental en mi cepillo de dientes sin ayuda. | 0.940 | 0.009 | 105.360 |
| F3_UE4 | I could pull a shirt on over my head by myself. | Pude ponerme un suéter por la cabeza sin ayuda. | 0.948 | 0.007 | 130.358 |
| F3_UE5 | I could hold a full cup. | Pude sostener un vaso lleno. | 0.957 | 0.006 | 147.717 |
| F3_UE6 | I could zip up my clothes. | Pude cerrar las cremalleras (cierres) de mi ropa. | 0.958 | 0.006 | 163.609 |
| F3_UE7 | I could use a key to unlock a door. | Pude usar una llave para abrir una puerta. | 0.913 | 0.013 | 71.372 |
| F3_UE8 | I could dial a phone. | Pude marcar los números del teléfono. | 0.962 | 0.006 | 154.884 |
| F3_UE9 | I could pull open heavy doors. | Pude abrir puertas pesadas tirando (jalando) de las mismas. | 0.912 | 0.011 | 80.334 |
| F3_UE11 | I could open the rings in school binders. | Pude abrir las anillas *(rings)* de las carpetas escolares. | 0.942 | 0.009 | 105.903 |
| F3_UE12 | I used a pencil with a special grip to write. | Para escribir, usé un lápiz que tenía un agarre especial. | 0.824 | 0.017 | 47.844 |
| F4_UE1 | I could open a jar by myself. | Pude abrir un frasco sin ayuda. | 0.920 | 0.009 | 97.939 |
| F4_UE2 | I could write with a pen or pencil. | Pude escribir con un lápiz o una pluma (bolígrafo). | 0.951 | 0.007 | 141.332 |
| F4_UE3 | I could brush my teeth by myself. | Pude cepillarme los dientes sin ayuda. | 0.950 | 0.007 | 132.608 |
| F4_UE4 | I could turn door handles by myself. | Pude girar las perillas (pomos) de las puertas sin ayuda. | 0.950 | 0.007 | 131.530 |
| F4_UE9 | I could open my clothing drawers. | Pude abrir mis cajones de ropa. | 0.949 | 0.007 | 131.737 |

| Item | English item wording: In the past 7 days. | Spanish item wording: En los Ultimos 7 dias. | Factor Loading | Standard Error | t-statistic |
|---|---|---|---|---|---|
| F4_UE10 | I could pour a drink from a full pitcher. | Pude servirme una bebida de una jarra llena. | 0.935 | 0.009 | 102.134 |
| F4_UE12 | I could dry my back with a towel. | Pude secarme la espalda con una toalla. | 0.936 | 0.009 | 105.919 |

**Table 3**

Item Parameter Estimates for 4 Upper Extremity Items with Differential Item Functioning by Language

| Item | Discrimination | 1st Threshold | 2nd Threshold | 3rd Threshold | 4th Threshold |
|---|---|---|---|---|---|
| F1_UE3: I could hold an empty cup | 2.48 | −3.55 | −2.99 | −2.69 | — |
| | 4.04 | −3.15 | −2.48 | −1.76 | — |
| F3_UE9: I could pull open heavy doors | 2.25 | −2.93 | −2.50 | −1.92 | −0.70 |
| | 2.66 | −4.52 | −3.42 | −2.45 | −1.46 |
| F4_UE1: I could open a jar by myself | 1.67 | −3.85 | −2.97 | −2.26 | −0.75 |
| | 2.54 | −4.32 | −3.23 | −2.43 | −1.36 |
| F4_UE10 I could pour a drink from a full pitcher. | 1.89 | −3.65 | −2.95 | −2.40 | −1.22 |
| | 2.96 | −4.11 | −3.25 | −2.50 | −1.62 |

Note: English parameters from DeWitt et al. (2011) are shown followed by parameters for Spanish sample after transformation to English-language metric. 1st Threshold = Theta needed for 50% chance of responding *with a lot of trouble*, *with some trouble*, *with a little trouble*, or *with no trouble*; 2nd Threshold = Theta needed for 50% chance of responding *with some trouble*, *with a little trouble*, or *with no trouble*; 3rd Threshold = Theta needed for 50% chance of responding *with a little trouble*, or *with no trouble*; 4th Threshold = Theta needed for 50% chance of responding *with no trouble*. *never*, and *never*.

3rd Threshold = Theta needed for 50% chance of responding *with a little trouble*, or *with no trouble*; 4th Threshold = Theta needed for 50% chance of responding *with no trouble*. *never*, and *never*.

**Table 4**

Categorical Factor Loadings For 23 Mobility Items In Spanish Sample

| Item | English item wording: In the past 7 days. | Spanish item wording: En los últimos 7 días. | Factor Loading | Standard Error | t-statistic |
|---|---|---|---|---|---|
| F1_MOB1 | I have been physically able to do the activities I enjoy most. | He podido hacer físicamente las actividades que más me gustan. | 0.938 | 0.007 | 128.381 |
| F1_MOB2 | I could ride a bike. | Pude montar en bicicleta. | 0.902 | 0.012 | 77.388 |
| F1_MOB3 | I could do sports and exercise that other kids my age could do. | Pude practicar los mismos deportes y ejercicios que hacían otros niños/as de mi edad. | 0.918 | 0.010 | 91.924 |
| F1_MOB5 | I could get down on my knees without holding on to something. | Pude arrodillarme sin agarrarme a nada. | 0.940 | 0.007 | 135.721 |
| F1_MOB6 | I could go up one step. | Pude subir un escalón. | 0.948 | 0.007 | 127.317 |
| F2_MOB 1 | I could run a mile | Pude correr una milla. | 0.815 | 0.017 | 48.445 |
| F2_MOB4 | I could walk up stairs without holding on to anything. | Pude subir escaleras sin agarrarme a nada. | 0.947 | 0.006 | 150.955 |
| F2_MOB6 | I could get up from a regular toilet. | Pude levantarme de un inodoro *(toilet)* común. | 0.947 | 0.008 | 123.353 |
| F2_MOB7 | I could stand up on my tiptoes. | Pude ponerme de puntillas (sobre las puntas de los pies). | 0.931 | 0.010 | 95.068 |
| F3_MOB2 | I could get into bed by myself. | Pude acostarme en la cama sin ayuda. | 0.941 | 0.009 | 108.986 |
| F3_MOB3 | I could stand up by myself. | Pude ponerme de pie sin ayuda. | 0.949 | 0.006 | 150.116 |
| F3_MOB5 | I used a walker, cane or crutches to get around. | Usé un andador, un bastón o unas muletas para moverme. | 0.930 | 0.008 | 115.879 |
| F3_MOB8 | I could move my legs. | Pude mover las piernas. | 0.939 | 0.010 | 98.541 |
| F3_MOB9 | I could get up from the floor. | Pude levantarme del piso (suelo). | 0.952 | 0.007 | 139.123 |
| F3_MOB10 | I could turn my head all the way to the side. | Pude girar (voltear) la cabeza totalmente hacia un lado. | 0.913 | 0.010 | 94.124 |
| F3_MOB12 | I could walk across the room. | Pude caminar a través de la habitatión. | 0.967 | 0.005 | 189.324 |
| F4_MOB2 | I could get in and out of a car. | Pude subir y bajar de un automóvil. | 0.937 | 0.009 | 100.898 |
| F4_MOB3 | I could walk more than one block. | Pude caminar más de una manzana (cuadra). | 0.926 | 0.010 | 94.279 |
| F4_MOB4 | I could keep up when I played with other kids. | Pude mantener el mismo ritmo cuando jugaba con otros niños/as. | 0.919 | 0.009 | 103.732 |
| F4_MOB6 | I could get out of bed by myself. | Pude levantarme de la cama sin ayuda. | 0.920 | 0.011 | 85.458 |
| F4_MOB7 | I used a wheelchair to get around. | Usé una silla de ruedas para moverme. | 0.912 | 0.011 | 81.978 |
| F4_MOB9 | I could carry my books in my backpack. | Pude llevar mis libros en la mochila. | 0.940 | 0.007 | 125.443 |
| F4_MOB10 | I could bend over to pick something up. | Pude inclinarme para recoger algo. | 0.946 | 0.007 | 137.950 |

**Table 5**

Item Parameter Estimates for 7 Mobility Items with Differential Item Functioning by Language

| Item | Discrimination | 1st Threshold | 2nd Threshold | 3rd Threshold | 4th Threshold |
|---|---|---|---|---|---|
| F1_MOB2: I could ride a bike. | 1.67 | −2.27 | −2.03 | −1.68 | −1.16 |
|  | 2.32 | −3.76 | −2.93 | −1.88 | −0.96 |
| F1_MOB3: I could do sports and exercise that other kids my age could do. | 3.11 | −1.92 | −1.75 | −1.14 | −0.45 |
|  | 2.54 | −3.73 | −2.85 | −1.68 | −0.76 |
| F2_MOB 1: I could run a mile | 1.13 | −2.64 | −1.78 | −0.52 | 0.88 |
|  | 1.63 | −3.80 | −2.87 | −1.66 | −0.25 |
| F2 MOB4: I could walk upstairs without holding on to anything. | 1.97 | −2.79 | −2.40 | −1.96 | −1.28 |
|  | 2.96 | −3.92 | −2.81 | −1.98 | −0.96 |
| F3 MOB5: I used a walker, cane or crutches to get around. | 1.67 | −3.56 | −3.38 | −2.88 | −2.48 |
|  | 2.66 | −4.02 | −2.73 | −1.84 | −1.00 |
| F3 MOB10: I could turn my head all the way to the side. | 1.16 | −4.45 | −3.97 | −3.32 | −2.28 |
|  | 2.70 | −3.94 | −2.92 | −1.97 | −1.10 |
| F4 MOB4: I could keep up when I played with other kids. | 1.96 | −2.90 | −2.41 | −1.63 | −0.49 |
|  | 2.52 | −3.84 | −2.81 | −1.80 | −0.72 |

Note: English parameters from DeWitt et al. (2011) are shown followed by parameters for Spanish sample after transformation to English-language metric. 1st Threshold = Theta needed for 50% chance of responding *with a lot of trouble*, *with some trouble*, *with a little trouble*, or *with no trouble*; 2nd Threshold = Theta needed for 50% chance of responding *with some trouble*, *with a little trouble*, or *with no trouble*; 3rd Threshold = Theta needed for 50% chance of responding *with a little trouble*, or *with no trouble*; 4th Threshold = Theta needed for 50% chance of responding *with no trouble*.

F3_MOB5 uses a different response scale: *almost always*, *often*, *sometimes*, *almost never*, and *never*.