Does a Standardized Spanish Translation of USDA's U.S. Household Food Security Survey Module Affect Differential Item Functioning Between Hispanic and White Non-Hispanic Households?

Matthew P. Rabbitt*[1] and Alisha Coleman-Jensen[1]

Economic Research Service
U. S Department of Agriculture
1400 Independence Avenue, Southwest
Mailstop 1800
Washington, DC 20250-0002
*Contact Author: matthew.rabbitt@ers.usda.gov
Phone: (202)-694-5593
Fax: (202)-245-4779

Proceeding of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

This paper uses data from the 2009-2014 CPS-FSS to estimate difference-in-difference Rasch and logistic regression models of food security to examine whether the implementation of a standardized Spanish translation of the Household Food Security Survey Module (HFSSM) affected differential item functioning (DIF) between Hispanic and white, non-Hispanic households. A key assumption of the Rasch model, on which the HFSSM is based, is that the severity and discrimination of items are invariant across subpopulations. Violation of this assumption is referred to as differential item function. The practical effect of DIF is that severity of food security at a given raw score may not be the same in two subpopulations, and prevalence estimates based on classification by raw score may give a biased comparison between subpopulations. Separate Rasch models are estimated for white, non-Hispanic and Hispanic households interviewed before and after the Spanish translation was standardized. Comparisons of the item-severity parameters from the Rasch models suggest three items (and one frequency follow-up) exhibited DIF: *Food bought did not last*, *Cut or skipped meals*, and *Ate less than felt should*. However, DIF was small and all items were classified as "A" items according to Educational Testing Service guidelines. Logistic regression model results for the effect on the standardized Spanish translation on items that exhibited meaningful DIF were generally imprecisely estimated. The practical implications of DIF were assessed using bias estimates for the prevalence of food insecurity and very low food security. The USDA standardized Spanish translation may have reduced the bias on the prevalence of food insecurity by as much as 0.16 percentage points and increased the bias on very low food security by as much as 0.48 percentage points.

Keywords: food security, Hispanics, Rasch model, differential item functioning

---

[1] The views expressed in this paper are those of the authors and do not necessarily reflect those of the USDA or Economic Research Service. The Authors gratefully acknowledge Mark Nord, Paul De Boeck, Yan Zhang, and audience members at the 2015 FCSM research conference for helpful comments.

Introduction

The Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA) measures the incidence and severity of food insecurity in U.S. households with the annual Current Population Survey Food Security Supplement (CPS-FSS). The Household Food Security Survey Module (HFSSM), administered as a part of the supplement, consists of a series of 18 questions about the conditions and behaviors characterizing households having difficulty meeting their basic food needs. Responses to these questions are used to construct the U.S. food security scale, which is based on the Rasch model (Rasch, 1960). A key assumption of the Rasch model is that the severity and discrimination of items are invariant across subpopulations. Violation of this assumption is referred to as differential item functioning (DIF). The practical effect of DIF is that severity of food insecurity at a given raw score (the count of affirmed food-insecure items) may not be the same in two subpopulations.

Ideally respondents to Federal surveys understand survey items and respond to those items consistently regardless of their characteristics, or the survey context. However, differences between households or across survey administrations may affect responses to survey items. This analysis examines whether and to what extent a standardized Spanish translation of USDA's HFSSM affects responses between Hispanic and white, non-Hispanic respondents. Separate Rasch models are estimated for white, non-Hispanic households and Hispanic households interviewed before and after the USDA's standardization of the Spanish translation for the HFSSM. Rasch model item-severity parameters were compared using tables and figures to assess the extent of DIF between white, non-Hispanic and Hispanic households. Logistic regression models were estimated for items that exhibited DIF based on the Rasch model analysis and Education Testing Service (ETS) guidelines for classifying DIF (Zwick, 2012) to explore what respondent characteristics may be related to differences between white, non-Hispanic and Hispanic households. Bias estimates for the prevalence of food insecurity and very low food security were also calculated based on the Rasch model analysis.

Background

USDA has measured the prevalence of food insecurity in U.S. households since 1995. Food-insecure households are those that had difficulty at some time during the year providing enough food for all their members due to a lack of resources. In 2014, 14.0 percent of U.S. households were food insecure (Coleman-Jensen et al., 2015). At USDA's request, the Committee on National Statistics (CNSTAT) convened an expert panel to review the food security measure and offer recommendations for improving the measure and food security statistics (for more information, see National Research Council, 2006). One recommendation from the CNSTAT panel was to conduct analysis "fitting models that allow for different item parameters for households with and without children for the questions that are appropriate for all households in order to study the possibility and effects of differential item function" (National Research Council, 2006, p.98). As a response to that recommendation ERS conducted research to examine the extent to which responses differ among population subgroups, including Hispanic and non-Hispanic respondents (Nord, 2012).

The underlying measurement theory behind the HFSSM is the Rasch model. The Rasch model is part of the family of Item Response Theory (IRT) models that originated in educational testing. The items used to assess household food security (10 items for adult only households, 18 items for households with children) are ordered in terms of severity. A greater number of affirmative responses to the items indicates a greater level of severity of food insecurity. The number of affirmative responses, or raw score on the scale, is used to assign household food security status. When developing the food security measure and with each subsequent year of data, analyses are conducted to ensure that the data fit the Rasch model (single-parameter logistic IRT measurement model) and that raw score is an adequate indicator of food security status[2]. A key characteristic of the Rasch model is that the way households respond to the items is dependent only on the latent trait (food security), not the characteristics of the households. As explained succinctly by Nord (2012):

> "The multiple-item Rasch model, on which the U.S. Household Food Security Scale is based, assumes that the severity and discrimination of items are invariant across subpopulations. Violation of this assumption between two subpopulations is described as differential item function. The underlying cause of DIF could

[2] Under the assumptions of the Rasch model the raw score is a minimal sufficient statistic for the latent trait being measured.

be either that respondents in two subpopulations understand the question to refer to different objective conditions, or that the way the two subpopulations experience or manage food insecurity differs so that the relative severity of items differs between the subpopulations. The practical effect of DIF is that severity of food insecurity at a given raw score may not be the same in the two subpopulations, and prevalence statistics based on classification by raw score may give a biased comparison between the two subpopulations" (Nord, 2012, p. 59).

Nord (2012) found that DIF was present for some items between Hispanic and white, non-Hispanic households with children. DIF by Hispanic origin was not found for households without children. Differential item function was not substantial but large enough that it could introduce some bias to comparisons of food insecurity prevalence in white, non-Hispanic and Hispanic households. The item *Could not afford balanced meals* was the item with the largest DIF. The estimated impact of the DIF was a 0.14 percentage point downward bias in the prevalence of food insecurity in Hispanic relative to white, non-Hispanic households. Adjusting or correcting for the inconsistencies between the two populations would not affect any substantive conclusions because the prevalence of food insecurity in Hispanic households is substantially higher than the prevalence of food insecurity in white, non-Hispanic headed households.

Although many Hispanic respondents are English speakers, one potential explanation for differences between Hispanic and non-Hispanic respondents is linguistic issues. As ERS investigated potential reasons for DIF between Hispanic and white, non-Hispanic household reference persons, we learned that multi-lingual Census interviewers translate survey content "on-the-fly". In an effort to standardize the implementation of Spanish-language interviews, ERS worked with Census to instrument both an English and Spanish language version of the HFSSM (see Appendix A). This Spanish translation was first developed by Harrison and colleagues (2003) and then revised by USDA researchers and multi-lingual Census interviewers. The Spanish language version of the food security questionnaire was first implemented in CPS-FSS interviews in 2012. In 2012 and 2013, a paper copy of the Spanish translation of the food security questions was provided to multi-lingual Census interviewers that they could read from. The computer instrument of the CPS-FSS was only available in English. In 2014, the Spanish language version of the full Food Security Supplement questionnaire was incorporated into the computer interview instrument. Interviewers could use either the English or Spanish version of the interview instrument. This paper is the first analysis of responses obtained through this standardized translation. The motivation is to examine DIF between Hispanic and white, non-Hispanic households with the implementation of the Spanish-language module.

Data

The analyses use data from the 2009-2014 Current Population Survey Food Security Supplement (CPS-FSS). The CPS is the official source of government statistics on employment status and poverty. Approximately 55,000 households are interviewed each month with data collected on labor force participation status, income, household demographics, and state identifiers. The CPS is representative of the state- and national-levels of the civilian, noninstitutionalized population after weighting.

The Food Security Supplement is conducted as a supplement to the CPS for the Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA). The purpose of the CPS-FSS is to estimate the prevalence and severity of food insecurity in U.S. households. Each year, ERS estimates national prevalence rates of food insecurity in its *Household Food Security in the United States* series (for example, see Coleman-Jensen et al., 2015). Households are asked questions about their food expenditures, basic food needs, and participation in public- and private-food assistance programs. To reduce respondent burden, households with income above 185 percent of the Federal poverty line[3] that showed no signs of food stress were not asked the food security questions.

Dependent Variables

The analyses examine responses to a series of food security questions taken from the HFSSM as dependent variables. The HFSSM consists of a series of 18 questions about the conditions and behaviors characterizing households having difficulty meeting their basic food needs. All households are asked the first 10 question about food needs among the households and adults. If children are present in the household, then an additional 8 questions

[3] The Federal poverty line was $24,008 for a family of four (two adults and two children) in 2014.

are asked about the food needs of children. The severity of food hardships experienced by the household ranges from anxiety over food running out to shortages of the amounts and kinds of foods eaten to episodes of adults and children (if present) going without food for an entire day. All of the questions reference the previous 12 months and are framed in terms of either shortages of money or affordability. For a complete listing of the English and Spanish versions of the food insecurity questions and the methods used to convert them into binary indicators see Appendix B.

The 18 questions that comprise the HFSSM form the basis of the U.S. household food security scale. This scale consist of 18 items, including 8 adult and household items, 7 child items, and 3 follow up (2 adult/household items and 1 child item). The analyses accounted for dependencies between follow-up and base items by modeling them as ordered polytomous items rather than as two independent items. Analyses were restricted to the 10-item (commonly referred to as the adult scale) U.S. food security scale to mitigate concerns about bias due to a departure from unidimensionality in the food security scale due to the presence of children (Nord and Coleman-Jensen, 2014) and to reduce the likelihood of detecting DIF between households with and without children unrelated to the language of interview (Wilde, 2004; Nord, 2012).

Independent Variables

The logistic regression analyses include additional controls for respondent's demographics, household characteristics, and household resources to explain the sources of DIF. The principle explanatory variables are an indicator for a Hispanic household reference person, the household's Rasch score, and interactions with indicators for the time periods pre- and post-USDA standardization of the HFSSM Spanish translation, where the pre- and post-USDA standardization time periods are characterized by the years 2009-2011 and 2012-2014[4], respectively.

Additional control variables were included in the logistic regression models to explain sources of DIF. An indicator for households in which Spanish was the only language spoken by adults age 15 and older is included to control for linguistic differences between Hispanic and white, non-Hispanic households. The analyses also controlled for respondents who reported they were non-citizen immigrants and the number of years since they entered the U.S. interacted with the indicator for non-citizen immigrant. Citizenship status and the number of years since entry into the U.S. are likely highly correlated with acculturation. The analyses also include controls for the proportion of adults who are female (a proxy for female respondent), age of the oldest adult (and it's square), income relative to the federal poverty line (and it's square), educational attainment of the most highly educated adult in the household, metropolitan residence status, and region- and time- fixed effects. These controls are commonly found to be associated with food security status (Coleman-Jensen et al., 2015).

Analysis Sample

The analyses consider households with a Hispanic or white, non-Hispanic reference person and incomes below 185 percent of the federal poverty line. The majority of households with incomes above 185 percent of the federal poverty line were screened out of the food security questions. Households with incomplete or missing responses to any of the food security questions were omitted from the analysis sample. Additionally, households with extreme raw scores (0 or 10) on the 10-item U.S. food security scale were omitted from the sample because households with these raw scores cannot be identified in the Conditional Maximum Likelihood (CML) Rasch model, resulting in a sample of 23,257 households.

Households were disaggregated into three mutually exclusive and exhaustive samples for the Rasch DIF analyses based on Hispanic ethnicity of the household reference person and whether or not the household was interviewed before or after USDA standardized the HFSSM Spanish translation. The first subsample consisted of households with a white, non-Hispanic reference person interviewed in 2009-2014, which will serve as the control group in the analyses. White, non-Hispanic households were not further disaggregated into pre- and post-implementation groups because the food security scale for white, non-Hispanic households is relatively consistent over time and they are unlikely to be affected by standardization of the standardized Spanish translation. The second sample includes

[4] Additional analysis not shown examined DIF for Hispanics interviewed in 2012/2013 and 2014 because of differences in the implementation of the standardized Spanish translation. No substantive differences were found. All Hispanic interviews from 2012, 2013, and 2014 are pooled for analyses.

households with a Hispanic reference person interviewed in 2009-2011. These households represent the pre-treatment group since they include Hispanic households prior to USDA's implementation of the standardized Spanish translation and will be used to establish a baseline food security scale for Hispanic households. The final group includes households with a Hispanic reference person interviewed in 2012-2014. Hispanic households interviewed in 2012-2014 would have been affected by the USDA's implementation of the Spanish translation and are considered the post-treatment group.

The final sample sizes for the Rasch analyses were 23,257 households, including 17,095 white, non-Hispanic households, and 3,113 and 3,049 Hispanic households interviewed before and after standardized the Spanish translation, respectively.

Data on Hispanic households at all income levels interviewed in 2009-2014 are used to construct a reference distribution of households across all raw scores for the food insecurity and very low food security prevalence estimate bias calculations. The sample consisted of 24,638 Hispanic households.

Econometric Methods

The U.S. household food security scale is based on an IRT model known as the Rasch model (Rasch, 1960). The Rasch model assumes multiple observable outcomes (such as food hardships) all derive from a single underlying variable, such as food security. Let household i's underlying continuous index of food insecurity be denoted by θ_i, with the property that higher values of the index correspond to greater levels of hardship. While the researcher is unable to observe θ_i directly, suppose he has j continuous indicators, Y_{ij}^*, which are related to θ_i such that each depends on the index and some random measurement error, ν_j. The relationship between the observable indicators and the underlying latent index can be expressed as

$$Y_{ij}^*=\theta_i+\nu_j, \tag{1}$$

which describes a factor analytic relationship in which the food-insecurity index, θ_i, is the underlying factor and the factor loadings (discrimination parameters) are constrained to be equal across all items and normalized to one.

Currently, the model is expressed in terms of a set of continuous indicators of food insecurity (the Y_{ij}^*'s); however, the U.S. food security scale consists of a series of binary variables. The Rasch model assumes the continuous indicators, Y_{ij}^*, are related to the binary responses of survey respondents as follows:

$$Y_{ij}=\begin{cases}1 & \text{if} \quad Y_{ij}^*>\delta_j\\ 0 & \text{if} \quad Y_{ij}^*\le\delta_j\end{cases}, \tag{2}$$

where δ_j is the severity threshold (calibration) parameter. Equation (2) specifies the same relationship between the latent continuous indicators and the observed categorical responses as that in standard probit and logistic models. The severity thresholds (δ_j) are estimated as part of the multivariate model and take on different values for each food-insecure condition. Higher values of the severity thresholds are associated with items that capture greater severity of food insecurity. Given equations (1) and (2), and the assumption that random measurement error (ν_j) follows a logistic distribution, the probability that household i's respondent affirms the j^{th} food-insecure condition is

$$P\left(Y_{ij}=1\mid\theta_i,\delta_j\right)=\frac{\exp\left(\theta_i-\delta_j\right)}{1+\exp\left(\theta_i-\delta_j\right)}, \tag{3}$$

where exp( ) is the exponential function.

The Rasch model assumes the errors in the responses are conditionally independent, which implies the probability of an affirmative response to a given food-insecure condition for a value of θ_i, the latent trait, does not depend on the response to another question. This conditional independence assumption implies the conditional probability of a given response vector is the product of the probabilities for each item. By stacking the household's responses (Y_{ij}) into a vector, Y_i, the probability of observing a given response pattern is given by

$$P\left(Y_i = y_i \mid \theta_i, \delta_j\right) = \prod_{j=1}^{J}\left[\frac{\exp\left(\theta_i - \delta_j\right)}{1+\exp\left(\theta_i - \delta_j\right)}\right]^{Y_{ij}}\left[1-\frac{\exp\left(\theta_i - \delta_j\right)}{1+\exp\left(\theta_i - \delta_j\right)}\right]^{\left(1-Y_{ij}\right)}, \quad (4)$$

where j runs over the number of food-insecure conditions for each household. Assuming observations are independent and identically distributed, the likelihood function is the product of the probabilities of observing a given response pattern for all observations and can be estimated via maximum likelihood methods. Rasch models for the analyses were estimated via Conditional Maximum Likelihood (CML) using SAS programs developed by ERS. The programs model items with follow-up questions as trichotomies and account for internal screening in the HFSSM.

Assessing Differential Item Functioning in the Rasch Model

Another key assumption of the Rasch model is parameter invariance, which assumes the severity and discrimination of items are invariant across subpopulations. When this assumption is violated for the severity thresholds, δ_j, the model will exhibit uniform DIF, or more formally,

$$\delta_j = \delta_j + \alpha_j G, \quad (5)$$

where G is an indicator for membership in a specific subpopulation and α_j, an unknown parameter to be estimated, is the j^{th} item specific effect of belonging to subpopulation group G compared to a reference subpopulation. If, after estimation, the α_j's are not statistically significantly different from zero, then membership in a subpopulation does not differentially impact the probability of affirming the food-insecurity items and the items do not exhibit DIF. However, if the α_j's can be distinguished from zero, then group membership explains latent food insecurity and the items exhibit DIF.

While numerous methods exist for detecting DIF (i.e., logistic regression, mixed logistic regression, Mantel-Haenszel chi-square statistic, etc.) (De Ayala, 2013; Van den Noortgate and De Boeck, 2005), the current analyses use CML Rasch models, estimated separately, for white non-Hispanic households, and Hispanics households before and after standardization of the Spanish translation. The result of estimating Rasch models for each subsample is three sets of item-severity parameters and their standard errors. However, the severity parameter estimates are not directly comparable because of differences in scaling. The analyses adjusted for this prior to making comparisons using a linear transformation that equated the mean and standard deviation of the Hispanic scales to the white, non-Hispanic scale. After equating the scales, estimates of the α_j's can be obtained by subtracting the severity thresholds (δ_j's) for white, non-Hispanics households from the severity thresholds for Hispanics households.

For the estimates of DIF to be meaningful, the differences in severity thresholds for the two groups being compared must be statistically significantly different from zero. To test for the presence of differences, the analyses use Welch's t-test, which is calculated as follows for item j.

$$t_j = \frac{\delta_{j,\,\mathrm{Hispanic}} - \delta_{j,\,\mathrm{WNH}}}{\sqrt{\left(SE_{j,\,\mathrm{Hispanic}}\right) + \left(SE_{j,\,\mathrm{WNH}}\right)}}, \quad (6)$$

where $\delta_{j,\,\mathrm{Hispanic}}$ and $\delta_{j,\,\mathrm{WNH}}$ are the severity thresholds for Hispanic and white, non-Hispanic households, respectively. The corresponding standard errors for the severity thresholds are given by $SE_{j,\,\mathrm{Hispanic}}$ and $SE_{j,\,\mathrm{WNH}}$.

Even small differences in severity parameters can be detected for subpopulations if sample sizes are large enough. Because of this, the National Center for Education Statistics (NCES) and ETS[5] have adopted classification rules based on (Petersen, 1988) to help researchers identify meaningful DIF. Items are classified into one of three categories (A, B, and C) based on the magnitude and statistical significance of the estimate of DIF. An "A" item is

[5] For a review of DIF classification procedures see Zwick (2012).

one in which the estimate of DIF is not significant at the 5 percent level or better or the magnitude of the DIF estimate is smaller than 0.43 logits[6] in absolute value. Items in category "A" exhibit no or negligible DIF. For an item to be classified as a "C" item, the estimate of DIF must be statistically significantly greater than 0.43 logits in absolute value at the 5 percent level or better and are larger than 0.64 logits in absolute value. These items indicate moderate or large DIF. Other items are categorized as "B" items, which are greater than 0.43 logits, but less than 0.64 logits, in absolute value and statistically distinguishable from zero.

Standards for identifying meaningful magnitudes of DIF vary from 0.4 to 0.6 logits (Draba, 1977; Longford et al., 1993; Scheuneman and Subhiyah, 1998; Wang, 2000; Paek, 2002; Elder et al., 2003). For the purposes of this analysis, meaningful DIF is defined by the ETS classification rules.

Explaining Differential Item Functioning

Logistic regression models are used in the DIF literature to detect and explain DIF. Under this approach, household i's response to the j^{th} food-insecurity condition is modeled as

$$Y_{ij} = \begin{cases} 1 \text{ if } \theta_i + \alpha G_i + \gamma\theta_i + X_i\beta + e_i > 0 \\ 0 \text{ otherwise} \end{cases}, \tag{7}$$

where θ_i is the household's food insecurity measure, G_i is an indicator for Hispanic group membership, and X_i is a matrix of control variable that includes respondent- and household-characteristics and region fixed effects, including a constant, and an error term, e_i, which is distributed logistically. The presence of uniform DIF may occur if, α, the group membership parameter is nonzero. Non-uniform DIF may be detected if an interaction between the group membership indicator (G) and the household's food insecurity measure (θ_i) is nonzero.

The sources of uniform- and non-uniform-DIF may be explained in this framework using a series of logistic regression models. The researcher will first estimate a baseline model that only includes the household's food insecurity measure, an indicator for Hispanic group membership, and their interaction. A second model is then estimated that builds on this baseline model by adding controls for the household respondent's demographics and household characteristics. Any changes in the Hispanic group membership variable and it's interaction with the household's food insecurity measure between the baseline and adjusted model indicate proportion of variation in uniform- and non-uniform-DIF that is explained by these additional characteristics.

The researcher must overcome two challenges to estimate the logistic model specified in equation (7). First, the food insecurity measure (θ_i) is an unobservable construct. A common solution to this problem is to use the household's raw food insecurity score (i.e., count of affirmed items.) as a proxy variable. However, this approach may be affected by multicollinearity. Therefore, the analyses use the household's Rasch score.

The second challenge that must be overcome in this analysis is the fact that Hispanics self-select into the standardized HFSSM Spanish translation because of observable and unobservable differences between Hispanic households or Census interviewers elect to administer the standardized translation for reasons that are unobservable to the researcher. The current analyses addresses this potential source of bias by implementing a difference-in-differences logistic regression model. Under this new parameterization, equation (7) is now

$$Y_{ijt} = \begin{cases} 1 \text{ if } \alpha_1 G_i Post_t + \gamma_1 G_i Post_t \theta_i + \lambda Post_t + \theta_i + \alpha_0 G_i + \gamma_0 \theta_i G_1 + X_{ijt}\beta + \tau_t + e_{ijt} > 0 \\ 0 \text{ otherwise} \end{cases}, \tag{8}$$

where i, j, and t are indices for the household, item, and year. The dummy variable $Post_t$ is equal to 1 for households surveyed in 2012-2014. G_i and X_i are defined above, τ_t represents year fixed-effects, and e_{ijt} is the error term. The

[6] The classification procedures proposed by Petersen (1988) and used the National Center for Education Statistics and Educational Testing Service use delta units because it uses the Mantel-Haenszel delta difference. One logit is equal to 2.35 deltas.

effect of administering the USDA standardized HFSSM translation will be unbiased as long as unobservable characteristics are time invariant.

Estimating the Effect of Differential Item Functioning on Food Security Prevalence Estimates

The extent of bias due to DIF between Hispanic and white, non-Hispanic households on the prevalence estimates of food insecurity and very low food security was assessed based on the methodology outlined in Nord (2012), which used the item severity parameters and household measures estimated separately for the populations of interest (Hispanic and white, non-Hispanic households in the current analysis) and then adjusted to a common scale using a linear transformation.

Nord calculated the bias on the prevalence of food insecurity (and very low food security) using a multistep iterative procedure. The steps for this procedure are as follows.

1. A reference distribution across food security raw scores was calculated using data for Hispanic households, including extreme scores (0 and 10) on the 10-item food security scale and all income levels. The observed (measured) prevalence of food insecurity (very low food security) in the reference distribution was calculated by determining the percentage of the sample with a raw score of 3 (6) or higher.
2. An initial value for the threshold for food insecurity (very low food security) was selected using the midpoint of the mean severity of raw scores 2 and 3 (5 and 6).
3. The proportion of households in each raw score group with "true" severity higher than the initial value of the threshold (i.e., truly food insecure (very low food secure) was calculated using the household parameters for white, non-Hispanic households and a normal distribution with mean and standard deviation corresponding to those estimated for that raw score group. Households in raw score group 0 were assumed to be food secure and households parameters for households with raw score 10 were calculated as if the raw score was 9.5.
4. The proportion of "truly" food insecure (very low food secure) in each raw score group were weighted by the proportion of the reference distribution with that raw score.
5. The threshold was then iteratively adjusted (repeating steps 3-5) until the "true" prevalence equaled the observed prevalence, which is the threshold at which there is zero measurement bias.
6. Step 3 was repeated for Hispanic households using their household parameters.
7. The proportion of "truly" food insecure (very low food secure) in each raw score group were weighted by the proportion of the reference distribution with that raw score. The weighted sum across raw scores represented the "true" prevalence of food insecurity (very low food security) in Hispanic households using the threshold at which measurement bias was zero for white, non-Hispanic households.
8. The difference between the observed prevalence of food insecurity (very low food security) and the true prevalence of food insecurity (very low food security) in Hispanic households is an estimate of the bias of measured food insecurity (very low food security) in Hispanic households relative to white, non-Hispanic households.

The analyses employed the procedure above to determine how standardization of the HFSSM Spanish translation affected food insecurity and very low food security prevalence estimates. Bias for the prevalence of food insecurity and very low food security were calculated, separately, using the samples of Hispanic households before and after standardization of the Spanish translation.

Results

Item-severity parameters from the CML Rasch models, estimated separately for white, non-Hispanic and Hispanic households before and after USDA standardized the HFSSM Spanish translation are listed in Table 1. The top panel contains item-severity parameters while the bottom panel lists the mean and standard deviation of the estimated 10-item food security scale. Column one reports Rasch model estimates for the samples of white, non-Hispanic households interviewed in 2009-2014. The remaining columns contain Rasch model item-severity parameter estimates for the samples of Hispanic households who were interviewed before and after the USDA implemented the standardized Spanish translation of the HFSSM.

[Table 1 about here.]

The relative severity of the items in the 10-item U.S. food security scale is as expected and similar for white, non-Hispanic and Hispanic households. Comparisons of the item-severity parameters between white, non-Hispanic and Hispanics households suggest, *Worried food would run out*, *Food bought didn't last*, *Adult cut or skipped meals (1 or 2 vs. 0)*, *Adult cut or skipped meals (2 vs. 0 or 1)*, and *Ate less than felt should* differed in severity between the two groups, regardless of which Hispanic household sample was used (pre- and post-USDA standardization). All other differences in item-severity parameters cannot be statistically distinguished from zero. When factoring in the classification guidelines proposed by ETS and adopted by NCES, the differences in item-severity parameters between white, non-Hispanic and Hispanic households interviewed before standardization of the Spanish translation fall in the "A" item grouping. Items in the "A" group typically exhibit little or no meaningful level of DIF. The items *Worried food would run out* and *Food bought did not last* were relatively more severe while the items *Adult cut or skipped a meal (1 or 2 vs. 0)*, *Adult cut or skipped a meal (2 vs. 0 or 1)*, and *Ate less than felt should* were less severe for white, non-Hispanics compared to Hispanics interviewed before standardization.

The classification of differences in item-severity parameters between white, non-Hispanic and Hispanic households interviewed after the USDA standardized the Spanish translation generally followed the patterns described above for white, non-Hispanics and Hispanics interviewed before the USDA standardization of the Spanish translation, with one exception. The difference in the item-severity parameter of *Adult cut or skipped meals (1 or 2 vs. 0)* for white, non-Hispanics and Hispanics interviewed after the USDA standardized the Spanish translation is now classified as a "B" item, rather than an "A" item, however, the difference-in-difference (i.e., change in DIF before and after standardization) for this item is not statistically significantly different from zero. Items within this group are characterized by weak DIF.

The effect of the USDA standardization on DIF between white, non-Hispanic and Hispanic households is assessed in the far right column of Table 1. This column represents the difference in DIF before and after the USDA standardized the Spanish translation. Intuitively, this is equivalent to a difference-in-difference Rasch model where the control group (white, non-Hispanics) is assumed to be unaffected by the treatment (the USDA's standardization of the Spanish translation). While this may appear to be a strong assumption, the U.S. food security scale is remarkably consistent over time, especially for white, non-Hispanic households. The item *Food bought did not last* is the only item where the difference in DIF was statistically significant, albeit at the 10 percent level. The difference in DIF estimate for this item suggests DIF increased by 0.13 logits in absolute magnitude and decreased in severity relative to white, non-Hispanics as a result of the USDA's standardization. Even with this increase in DIF the item remains classified as an "A" item, albeit within one standard deviation of being considered a "B" item.

The bottom panel of Table 1 presents the mean and standard deviation of the 10-item food security scale for white, non-Hispanic and Hispanic households. The standard deviations for Hispanics interviewed before and after the USDA standardized the Spanish translation are 17 percent and 15 percent larger, respectively, than in white, non-Hispanic households when all models were estimated on a logistic metric (discrimination parameters equal to 1.0). This indicates a stronger fit of the data to the Rasch model for Hispanic households, that is, moderately higher average discrimination of items. Parameters were adjusted for this difference prior to comparison using a linear transformation.

Items may differ in severity and/or discrimination between white, non-Hispanic and Hispanic households. Any potential concerns about DIF in discrimination (non-uniform DIF) are relieved by item-infit statistics. Item discrimination estimates may be obtained by estimating 2-parameter Rasch models, however, the current analysis does not consider these models because the aim of this analysis was to keep the modelling strategies as close as possible to the existing methodology used to construct the U.S. food security scale. Item-infit statistics provide a proxy (although inversely) for item discrimination. Table 2 presents item-infit statistics from CML Rasch models estimated separately for white, non-Hispanic and Hispanic households interviewed before and after USDA standardized the Spanish translation of the HFSSM.

[Table 2 about here.]

Item-infit statistics were generally similar for white, non-Hispanic and Hispanic households, regardless of which Hispanic sample was used. Ideally, all item-infit statistics should be 1.0, indicating all items discriminate equally well, as assumed by the Rasch model. In practice, item-infit statistics ranging from 0.7 to 1.3 are generally considered to indicate an acceptable fit of the Rasch model. Household's responses generally fit the Rasch model well, with the majority of item-infit statistics ranging from 0.9 to 1.3, with the exception of *Adult cut or skipped meals (overall)*, which had an item-infit of 1.39. Because the item-infit statistic is a chi-square-like statistic, higher values indicate the item is more weakly (i.e., poorly discriminates) associated with latent food insecurity.

Difference-in-difference logistic regression models for food insecurity items that exhibited meaningful DIF were estimated to assess the extent to which the standardized Spanish translation and other respondent- and household-characteristics explain DIF between white, non-Hispanic and Hispanic households. Table 3 contains logistic regression coefficients and standard errors from models for the items *Food bought did not last*, *Adult cut or skipped meals (1 or 2 vs. 0)*, *Adult cut or skipped meals (2 vs. 0 or 1)*[7], and *Ate less than felt should*. The top panel contains estimates for uniform DIF coefficients, while the middle and bottom panels list estimates for non-uniform DIF and respondent/household characteristic coefficients, respectively. Unadjusted models include controls for uniform- and non-uniform-DIF and time fixed effects. The adjusted models build upon the unadjusted models by adding controls for respondent and household characteristics (see the data section for a listing of the control variables), and region fixed effects.

[Table 3 about here.]

Uniform DIF, represented by the indicator for a household with a Hispanic reference person, was detected for all items and all specifications. Households headed by a Hispanic reference person appear to be less likely to affirm each item compared to households headed by white, non-Hispanics. For each item, the coefficient on Hispanic reference person interacted with the post-USDA standardization time period is imprecisely estimated. The second panel lists coefficients and standard errors that control for non-uniform DIF. Non-uniform DIF is captured by including a control for the interaction between the indicator for households with a Hispanic reference person and the household's Rasch score. Non-uniform DIF was detected for all items with the exception of *Adult cut or skipped meals (1 or 2 vs. 0)*. Since the USDA standardized Spanish translation may also have an effect on non-uniform DIF, the analyses included an interaction between Hispanic reference person, the post-USDA standardization time period, and the household's Rasch score, which was insignificant for all items and specifications, with the exception of the item *Ate less than felt should*. After controlling for respondent- and household- characteristics, the USDA standardized Spanish translation was associated with a reduction in the probability of affirming *Ate less than felt should* at different values of the household's Rasch score compared to white, non-Hispanic households.

Controlling for respondent and household characteristics increased uniform DIF between Hispanics and white, non-Hispanics by 14.8 percent for the item *Food bought did not last*, while decreasing DIF for *Adult cut or skipped meals (1 or 2 vs. 0)*, *Adult cut or skipped meals (2 vs. 0 or 1)*, and *Ate less than felt should* by 17.4 percent, 5.1 percent, and 1.9 percent, respectively. In the cases of *Adult cut or skipped meals (1 or 2 vs. 0)* and *Ate less than felt should*, the decline in the uniform DIF was accounted for mainly by educational attainment of the highest educated person in the household. For *Food bought did not last*, the increase in uniform DIF was almost entirely accounted for by educational attainment of the highest educated adult and households with children. The decline in uniform DIF for *Adult cut or skipped meals (2 vs. 0 or 1)* was strongly associated with the female headed households.

Differences in items severity for *Food bought did not last*, *Cut or skipped meals*, and *Ate less than felt should* for white, non-Hispanic and Hispanic households may bias the prevalence of food insecurity and very low food security. The bias for food insecurity prevalence estimates is expected to be upwards because of the lower severity of *Food bought did not last* in Hispanic households relative to white, non-Hispanic households (Figure 1). The USDA's classification system for a household's food insecurity (very low food security) status is based on a raw score of three or more (six or more). The items in question are the second, fourth, fifth, and sixth on the U.S. food security scale in order of severity.

[Figure 1 about here.]

The prevalence of very low food security is expected to be biased downwards for Hispanic households relative to white, non-Hispanic households because of the higher severity of *Cut or skipped meals* and *Ate less than felt should*

[7] Ordered logistic models were also estimated for *Adult cut or skipped meals* to model the dependencies between the base and follow-up question, however, the results produced very similar results.

in Hispanic households relative to white, non-Hispanic households. Two of the items in question, *Cut or skipped meals* and *Ate less than felt should*, are close to the threshold and will likely be the strongest influences on the bias for very low food security prevalence estimates. These items are the fourth, fifth, and sixth on the U.S. food security scale based on severity.

The estimated bias on the prevalence of food insecurity is in the opposite direction in spite of the lower severity of *Food bought did not last*, regardless of which Hispanic household parameters were used. The bias analysis was conducted based on the mean and standard deviation of household measures for Hispanic households interviewed before and after standardization of the Spanish translation. For Hispanic households interviewed before the Spanish translation was standardized, the prevalence of food insecurity was biased downwards by 0.62 percentage points (2.8 percent) for Hispanic households relative to white, non-Hispanic households. The downwards bias on the prevalence of food insecurity was smaller for Hispanics interviewed after standardization. Specifically, the prevalence of food insecurity was downwards biased by 0.46 percentage points (2.06 percent) for Hispanic households relative to white, non-Hispanics. Differencing the bias estimates using the household measures of Hispanics interviewed before and after standardization suggests the standardization of the Spanish translation may have reduced the bias on food insecurity induced by DIF between Hispanics and white, non-Hispanics by as much as 0.16 percentage points.

Bias estimates on the prevalence of very low food security are consistent with expectations. When the household measures is used for Hispanics interviewed before standardization of the Spanish interview, the prevalence of very low food security is downwards biased by 0.38 percentage points (4.9 percent). However, the downwards bias on very low food security roughly doubles to 0.86 percentage points (11.2 percent) when using household measures from Hispanics interviewed after standardization. While at face value the standardized Spanish translation appears to bias very low food insecurity prevalence estimates, these findings are similar to the results reported in Nord (2012). In his analysis he estimated that the measured prevalence of very low food security for Hispanic households with children was biased downwards by 0.72 percentage points or 14 percent. After differencing the bias estimates using the household measures of Hispanics interviewed before and after standardization, the USDA's standardization of the Spanish translation appears to have increased the bias on very low food insecurity induced by DIF between Hispanics and white, non-Hispanics by as much as 0.48 percentage points.

Conclusion

This analyses use nationally representative data from the 2009-2014 CPS-FSS to examine whether a standardized Spanish translation of the HFSSM affects DIF between Hispanic and white, non-Hispanic households. Comparisons of Rasch model item-severity parameters suggest DIF exists between Hispanic and white, non-Hispanic households. Small differences in item-severity parameters were detected for *Worried food would run out*, *Food bought didn't last*, *Adult cut or skipped meals* (at the different thresholds), and *Ate less than felt should*, however, none of the items exhibited meaningful levels of DIF according to the guidelines proposed by ETS and used by NCES. Effect sizes of DIF for each item resulted in the majority of items being classified as "A" items. Under ETS guidelines, "A" items exhibit no DIF.

The analyses employed a difference-in-difference approach to estimate the effect of the USDA's implementation of a standardized Spanish translation of HFSSM on DIF between Hispanic and white, non-Hispanic households. While the Rasch results suggest DIF exists between Hispanics and white, non-Hispanics, differences in DIF before and after the standardization were imprecisely estimated for all items except *Food bought did not last*. For this item, standardization of the Spanish translation increased DIF by 0.13 logits. While standardization of the Spanish translation increased the effect size of DIF for this item, it remains classified as an "A" item.

Results from the logistic models for items that exhibited meaningful DIF consistently indicated standardization of the Spanish translation had no effect on uniform DIF. While the standardized Spanish translation does not appear to have had an effect on uniform DIF, respondent- and household-characteristics explained nearly one eight and one fifth of DIF for *Food bought did not last* and *Adult cut or skipped meals*. Non-uniform DIF was detected in the logistic regression models for Hispanics relative to white, non-Hispanics households, however, standardization of the Spanish translation had little to no effect on non-uniform DIF for the items except for *Ate less than felt should*.

While the results suggest modest DIF exists between Hispanic and white, non-Hispanic households, it is important to assess the practical implications of these findings. The implication of DIF on the measurement of food insecurity in these subpopulations appears to be rather minimal. To place this in perspective, the estimated bias on the prevalence of food insecurity ranges from -0.46 (2.8 percent) to -0.62 (2.1 percent) percentage points for Hispanic households relative to white, non-Hispanic households. The lower end estimate of bias is associated with the interview period after the USDA standardized the Spanish translation, and suggests standardization may have reduced the bias on the prevalence of food insecurity by as much as 0.16 percentage points.

Standardization of the Spanish translation of the HFSSM has important practical effects for the measurement of very low food security. Bias for the prevalence of very low food security ranges from -0.38 percentage points (4.9 percent) to -0.86 percentage points (-11.2 percent) when using the household measures for Hispanics interviewed before and after standardization, respectively. Bias on the prevalence of very low food security appears to have increased by 0.48 percentage points as a result of the USDA's standardization of the Spanish translation. The increase in bias for the prevalence of very low food security is likely being caused by positive DIF for *Adult cut or skipped meals* and *Ate less than felt should*, which are the fourth, fifth, and sixth items on the U.S. food security scale and fall directly below the threshold for very low food security. Overall, these bias estimates are relatively small compared to the most recent prevalence estimates when one considers that 22.4 percent of all Hispanic households were food insecure compared to 10.5 percent of white, non-Hispanic households in 2014 (Coleman-Jensen et al., 2015). In that same year 6.9 percent and 4.5 percent of Hispanic and white, non-Hispanic households experienced very low food security, respectively.

References

Bickel, Gary, Mark Nord, Cristofer Price, William Hamilton, and John Cook. 2000. Guide to Measuring Household Food Security Revised 2000. U.S. Department of Agriculture, Food and Nutrition Service.

Coleman-Jensen, Alisha, Matthew P. Rabbitt, Christian Gregory, and Anita Singh. 2015. Household Food Security in the United States in 2014. ERR-194. U.S. Department of Agriculture, Economic Research Service.

De Ayala, Rafael. J. (2013). The Theory and Practice of Item Response Theory. New York, NY. Guilford Publications.

Draba, R. E. (1977). The Identification and Interpretation of Item Bias. Research Memorandum, (25).

Elder, C., McNamara, T., and Congdon, P. (2002). Rasch Techniques for Detecting Bias in Performance Assessments: An Example Comparing the Performance of Native and Non-Native Speakers on a Test of Academic English. Journal of Applied Measurement, 4(2), 181-197.

Harrison, Gail. G., Ame Stormer, Dena R. Herman, and Donna M. Winham. 2003. Development of a Spanish-Language Version of the US Household Food Security Survey Module. The Journal of Nutrition, 133(4), 1192-1197.

Kilanowski, Jill. F., and Li Lin. 2012. Rasch Analysis of US Household Food Security Survey Module in Latino Migrant Farmworkers. Journal of Hunger & Environmental Nutrition, 7(2-3), 178-191.

Longford, Nicholas. T., Holland, P. W., and Thayer, D. T. (1993). Stability of the MH D-DIF Statistics Across Populations. In: Wainer, PWHH., editor. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates, 67-113.

National Research Council. 2006. Food Insecurity and Hunger in the United States: An Assessment of the Measure. Panel to Review the U.S. Department of Agriculture's Measurement of Food Insecurity and Hunger, Gooloo S. Wunderlich and Janet L. Norwood, Editors, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nord, Mark. 2012. Assessing Potential Technical Enhancements to the U.S. Household Food Security Measures. TB-1936. U.S. Department of Agriculture, Economic Research Service.

Nord, Mark, and Coleman-Jensen, A. (2014). Improving Food Security Classification of Households with Children. Journal of Hunger and Environmental Nutrition, 9(3), 318-333.

Paek, I. (2002). Investigations of Differential Item Functioning: Comparisons Among Approaches, and Extension to a Multidimensional Context. University of California, Berkeley.

Petersen, N. (1988). DIF Procedures for Use in Statistical Analysis. Unpublished Memorandum.

Rasch, Georg. 1960. Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: University of Chicago Press.

Scheuneman, Janice D., and Subhiyah, R. G. (1997). Evidence for the Validity of a Rasch Model Technique for Identifying Differential Item Functioning. Journal of Outcome Measurement, 2(1), 33-42.

Van den Noortgate, Wim, and Paul De Boeck. 2005. Assessing and Explaining Differential Item Functioning Using Logistic Mixed Models. Journal of Educational and Behavioral Statistics, 30(4), 443-464.

Wang, W. C. (1999). Modeling Effects of Differential Item Functioning in Polytomous Items. Journal of Applied Measurement, *1*(1), 63-82.

Wilde, P. E. (2004). Differential Response Patterns Affect Food-Security Prevalence Estimates for Households with and without Children. The Journal of Nutrition, 134(8), 1910-1915.

Zwick, Rebecca. (2012). A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement. ETS Research Report Series, 2012(1), i-30.

Tables

Table 1: CML Rasch Model Item Severity Parameters Estimated in Separate Models for Hispanic and White Non-Hispanic Households Pre- and Post-USDA Standardization of the HFSSM Spanish Translation

		Hispanic households				
	White non-Hispanic households	Pre-USDA Standardized Spanish Translation		Post-USDA Standardized Spanish Translation		
Item	Severity parameter	Severity parameter	Difference from White non-Hispanic households	Severity parameter	Difference from White non-Hispanic households	Difference-in-Difference
Worried food would run out	4.01	3.77	-0.24*** (0.067)	3.75	-0.26*** (0.065)	-0.02 (0.087)
Food bought did not last	4.97	4.74	-0.23*** (0.056)	4.61	-0.36*** (0.057)	-0.13* (0.074)
Could not afford balanced meals	5.07	5.10	0.03 (0.055)	5.00	-0.07 (0.055)	-0.10 (0.072)
Adult cut or skipped meals (1 or 2 vs. 0)	6.39	6.71	0.32*** (0.094)	6.90	0.50*** (0.097)	0.19 (0.126)
Adult cut or skipped meals (2 vs. 0 or 1)	6.84	7.03	0.19* (0.109)	7.24	0.39*** (0.106)	0.21 (0.143)
Ate less than felt should	6.43	6.69	0.26*** (0.059)	6.80	0.37*** (0.060)	0.12 (0.078)
Hungry but did not eat	8.10	8.18	0.08 (0.073)	8.11	0.01 (0.072)	-0.07 (0.096)
Lost weight	8.96	8.98	0.02 (0.087)	8.97	0.01 (0.088)	-0.01 (0.116)
Adult did not eat for whole day (1 or 2 vs. 0)	9.50	9.32	-0.18 (0.191)	9.22	-0.28 (0.191)	-0.10 (0.255)
Adult did not eat for whole day (2 vs. 0 or 1)	9.73	9.49	-0.24 (0.236)	9.41	-0.32 (0.241)	-0.08 (0.391)
Mean—all items	7.00	7.00		7.00		
Std. deviation—all items	1.90	1.90		1.90		
Discrimination parameter	1.00	1.17		1.15		
Number of cases	17,095	3,113		3,049		

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS. Standard errors appear in parenthesis. All models were estimated on a logistic metric (discrimination parameter = 1.00). The discrimination parameter above for Hispanic and White, non-Hispanic households are those required to equate the standard deviation to be equivalent in Hispanic and White, non-Hispanic households for the purposes of that comparison. Severity parameters for the trichotomous items, *Adult cut or skipped meals* and *Adult did not eat for whole day*, are the "Rasch-Thurstone" or "50-percent probability" values.

* Significant at the 0.10 level. ** Significant at the 0.05 level. *** Significant at the 0.01 level.

Table 2: CML Rasch Model Item-Infit Statistics for Separate Models for Hispanic and White Non-Hispanic Households

		Hispanic Households	
Item	White non-Hispanic households	Pre-USDA Standardized Spanish Translation	Post-USDA Standardized Spanish Translation
Worried food would run out	0.96	1.00	0.97
Food bought did not last	0.94	0.89	0.88
Could not afford balanced meals	1.22	1.16	1.21
Adult cut or skipped meals (1 vs. 0)	0.98	1.06	0.96
Adult cut or skipped meals (2 vs. 1)	1.15	1.24	1.30
Adult cut or skipped meals (overall)	1.15	1.39	1.31
Ate less than felt should	0.75	0.70	0.76
Hungry but did not eat	0.78	0.85	0.80
Lost weight	0.91	0.87	0.86
Adult did not eat for whole day (1 vs. 0)	1.07	0.98	0.91
Adult did not eat for whole day (2 vs. 1)	1.04	1.13	1.31
Adult did not eat for whole day (overall)	1.07	1.12	1.25
Number of cases	17,095	3,113	3,049

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS. Fit statistics for the trichotomous items, *Adult cut or skipped meals* and *Adult did not eat for whole day*, were calculated from mean square differences between observed responses and the expected mean response that the response was in one of the two categories of that threshold. Overall fit statistics for trichotomous items were based on observed value (0, 1, or 2) versus expected value in the 0-2 range.

Table 3: Difference-in-Difference Logistic Regression Models of Responses to Items that Exhibit Meaningful Differential Item Functioning for Hispanic and White Non-Hispanic Households

	Food bought did not last		Adult cut or skipped meals (1 or 2 vs. 0)		Adult cut or skipped meals (2 vs. 0 or 1)		Ate less than felt should	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
<u>Uniform differential item functioning parameters:</u>								
Hispanic reference person interacted with post-USDA standardization of Spanish translation	-0.094	-0.067	-0.215	-0.206	0.862	0.837	0.867	0.884
	(0.340)	(0.341)	(0.821)	(0.822)	(0.883)	(0.885)	(0.612)	(0.612)
Hispanic reference person	-0.824***	-0.967***	-1.215**	-1.035*	-1.611**	-1.533**	-1.722***	-1.690***
	(0.243)	(0.249)	(0.547)	(0.551)	(0.642)	(0.648)	(0.445)	(0.450)
Post-USDA standardization of Spanish translation	0.145	0.174	-0.809*	-0.846*	-0.254	-0.272	-0.465	-0.465
	(0.179)	(0.180)	(0.444)	(0.445)	(0.469)	(0.470)	(0.326)	(0.326)
<u>Non-uniform differential item functioning parameters:</u>								
Hispanic reference person interacted with post-USDA standardization of Spanish translation and Rasch score	0.049	0.048	0.003	-0.009	-0.183	-0.181	-0.192	-0.198*
	(0.101)	(0.101)	(0.155)	(0.155)	(0.145)	(0.146)	(0.118)	(0.118)
Hispanic reference person interacted with Rasch score	0.396***	0.395***	0.124	0.125	0.230**	0.229**	0.292***	0.292***
	(0.071)	(0.071)	(0.104)	(0.105)	(0.106)	(0.106)	(0.087)	(0.087)
Rasch score	0.808***	0.813***	1.594***	1.604***	1.488***	1.492***	1.246***	1.253***
	(0.030)	(0.030)	(0.056)	(0.056)	(0.053)	(0.053)	(0.041)	(0.041)
Post-USDA standardization of Spanish translation interacted with Rasch Score	0.008	0.006	0.132	0.141*	0.033	0.034	0.067	0.068
	(0.042)	(0.042)	(0.084)	(0.084)	(0.075)	(0.075)	(0.059)	(0.059)
<u>Respondent/household characteristics:</u>								
Respondent is a non-citizen immigrant		0.219		-0.324		-0.167		-0.089
		(0.151)		(0.215)		(0.224)		(0.186)
Non-citizen immigrant interacted with years since U.S. entry		-0.004		0.010		0.012		-0.002
		(0.007)		(0.010)		(0.010)		(0.009)
Spanish is the only languages spoken by all adults		-0.041		-0.062		0.059		0.025
		(0.124)		(0.168)		(0.177)		(0.146)
Household with child		0.208***		-0.023		-0.097		0.165**
		(0.060)		(0.085)		(0.085)		(0.074)
Proportion of adults who are female		0.117		0.064		0.307***		-0.086
		(0.084)		(0.118)		(0.116)		(0.103)

Table 3 (Cont.): Difference-in-Difference Logistic Regression Models of Response to Items that Exhibit Meaningful Differential Item Functioning for Hispanic and White Non-Hispanic Households

	Food bought did not last		Adult cut or skipped meals (1 or 2 vs. 0)		Adult cut or skipped meals (2 vs. 0 or 1)		Ate less than felt should	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Income to poverty ratio		-0.318		0.273		0.531		-0.149
		(0.242)		(0.341)		(0.340)		(0.297)
Income to poverty ratio, squared		0.084		-0.058		-0.260		0.106
		(0.120)		(0.172)		(0.173)		(0.150)
Age of oldest adult		0.024**		-0.026*		0.006		-0.010
		(0.010)		(0.014)		(0.014)		(0.012)
Age of oldest adult, squared		-0.000*		0.000		-0.000		0.000
		(0.000)		(0.000)		(0.000)		(0.000)
Most highly educated adult less than high school		0.097		-0.123		-0.064		0.024
		(0.080)		(0.115)		(0.116)		(0.099)
Most highly educated adult some college, less than bachelor’s degree		-0.179***		0.164*		0.076		0.126
		(0.066)		(0.094)		(0.095)		(0.082)
Most highly educated adult bachelor or 4-year degree		-0.388***		0.392***		0.052		0.215*
		(0.101)		(0.146)		(0.149)		(0.128)
Most highly educated adult graduate or professional degree		-0.375**		0.453*		0.180		0.209
		(0.175)		(0.257)		(0.260)		(0.226)
Resident in principal city of metropolitan statistical area		-0.065		-0.096		-0.180*		0.065
		(0.072)		(0.102)		(0.104)		(0.089)
Resident in metropolitan statistical area not identified		-0.086		-0.011		-0.074		0.189*
		(0.082)		(0.118)		(0.117)		(0.103)
Resident outside of metropolitan statistical area		-0.015		-0.154		-0.126		0.080
		(0.076)		(0.109)		(0.110)		(0.095)
Constant	-2.625***	-3.203***	-8.184***	-7.540***	-8.971***	-9.410***	-6.512***	-6.304***
	(0.129)	(0.304)	(0.297)	(0.485)	(0.331)	(0.518)	(0.225)	(0.407)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Log-Likelihood	-4,194.157	-4,155.258	-2,250.262	-2,226.626	-2,204.574	-2,192.293	-2,898.414	-2,887.090
N	23,257	23,257	23,257	23,257	23,257	23,257	23,257	23,257

Note: Logistic models were estimated using weighted household data for households with income below 185% of the Federal poverty line from the 2009-2014 CPS-FSS. Standard errors appear in parenthesis.

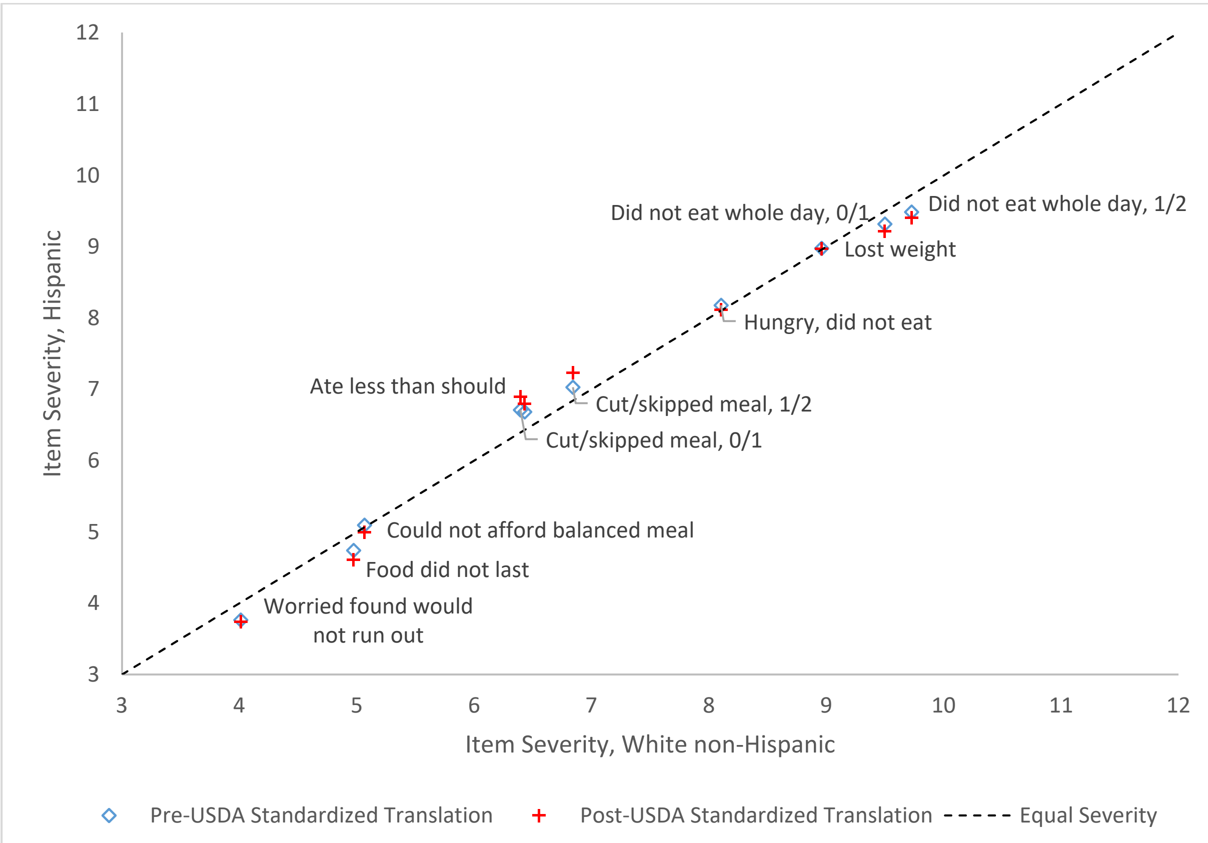
Table 4: Bias for the Estimated Prevalence of Food Insecurity and Very Low Food Security of Hispanic and White Non-Hispanic Households Due to Differential Item Functioning

Method and Characteristic	Measured Prevalence	True Prevalence		Bias, Measured versus True		White non-Hispanic versus Hispanic	
		White non-Hispanic Household	Hispanic Household	White non-Hispanic Household	Hispanic Household	Bias, Percent of All Households	Percent of Observed Prevalence
	- - - - - - - - percent - - - - - - - -			*- percentage points -*		*- - - percent - - -*	
Pre-USDA Standardized Spanish Translation							
Food Insecurity	22.15	22.15	22.77	0.00	-0.62	-0.62	-2.80
Very Low Food Security	7.77	7.77	8.15	0.00	-0.38	-0.38	-4.89
Post-USDA Standardized Spanish Translation							
Food Insecurity	22.15	22.15	22.61	0.00	-0.46	-0.46	-2.06
Very Low Food Security	7.77	7.77	8.64	0.00	-0.86	-0.86	-11.19

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS.

Figures

Figure 1: CML Rasch Model Item-Severity Parameter Comparison, Hispanic versus White, non-Hispanic Households



Note: Item-severity parameters were estimated from separate CML Rasch models for Hispanic and White, non-Hispanic households using data from the 2009-2014 CPS-FSS.

Appendix A. Analysis of Differential Item Functioning between Hispanic and White Non-Hispanic Households with and without Children

Table A.1: CML Rasch Model Item Severity Parameters Estimated in Separate Models for Hispanic and White Non-Hispanic Households without Children Pre- and Post-USDA Standardization of the HFSSM Spanish Translation

	White non-Hispanic households	Hispanic households				
		Pre-USDA Standardized Spanish Translation		Post-USDA Standardized Spanish Translation		
Item	Severity parameter	Severity parameter	Difference from White non-Hispanic households	Severity parameter	Difference from White non-Hispanic households	Difference-in-Difference
Worried food would run out	4.23	3.91	-0.31***	3.93	-0.30***	0.01
			(0.110)		(0.101)	(0.147)
Food bought did not last	5.02	4.72	-0.29***	4.71	-0.31***	-0.01
			(0.097)		(0.092)	(0.131)
Could not afford balanced meals	4.77	4.96	0.19**	4.75	-0.01	-0.21
			(0.095)		(0.091)	(0.129)
Adult cut or skipped meals (1 or 2 vs. 0)	6.40	6.71	0.31**	6.86	0.46***	0.15
			(0.158)		(0.152)	(0.214)
Adult cut or skipped meals (2 vs. 0 or 1)	6.81	7.06	0.25	7.26	0.44***	0.20
			(0.174)		(0.158)	(0.231)
Ate less than felt should	6.47	6.64	0.17*	6.76	0.28	0.11
			(0.098)		(0.095)	(0.133)
Hungry but did not eat	8.18	8.25	0.07	8.14	-0.04	-0.10
			(0.114)		(0.109)	(0.155)
Lost weight	8.90	8.95	0.05	8.80	-0.10	-0.15
			(0.128)		(0.124)	(0.175)
Adult did not eat for whole day (1 or 2 vs. 0)	9.50	9.28	-0.22	9.29	-0.21	0.02
			(0.254)		(0.282)	(0.371)
Adult did not eat for whole day (2 vs. 0 or 1)	9.72	9.51	-0.21	9.50	-0.23	-0.02
			(0.320)		(0.336)	(0.454)
Mean—all items	7.00	7.00		7.00		
Std. deviation—all items	1.90	1.90		1.90		
Discrimination parameter	1.00	1.07		1.06		
Number of cases	10,645	1,027		1,068		

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS. Standard errors appear in parenthesis. All models were estimated on a logistic metric (discrimination parameter = 1.00). The discrimination parameter above for Hispanic and White, non-Hispanic households are those required to equate the standard deviation to be equivalent in Hispanic and White, non-Hispanic households for the purposes of that comparison. Severity parameters for the trichotomous items, *Adult cut or skipped meals* and *Adult did not eat for whole day*, are the "Rasch-Thurstone" or "50-percent probability" values.

* Significant at the 0.10 level. ** Significant at the 0.05 level. *** Significant at the 0.01 level.

Table A.2: Bias for the Estimated Prevalence of Food Insecurity and Very Low Food Security of Hispanic and White Non-Hispanic Households without Children Due to Differential Item Functioning

		True Prevalence		Bias, Measured versus True		White non-Hispanic versus Hispanic	
Method and Characteristic	Measured Prevalence	White non-Hispanic Household	Hispanic Household	White non-Hispanic Household	Hispanic Household	Bias, Percent of All Households	Percent of Observed Prevalence
	- - - - - - - - percent - - - - - - - -			*- percentage points -*		*- - - percent - - -*	
Pre-USDA Standardized Spanish Translation							
Food Insecurity	19.61	19.61	19.90	0.00	-0.29	-0.29	-1.46
Very Low Food Security	7.95	7.95	8.41	0.00	-0.46	-0.46	-5.79
Post-USDA Standardized Spanish Translation							
Food Insecurity	19.61	19.61	19.79	0.00	-0.17	-0.17	-0.88
Very Low Food Security	7.95	7.95	8.72	0.00	-0.77	-0.77	-9.69

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS.

Table A.3: CML Rasch Model Item Severity Parameters Estimated in Separate Models for Hispanic and White Non-Hispanic Households with Children Pre- and Post-USDA Standardization of the HFSSM Spanish Translation

	White non-Hispanic households	Hispanic households				
		Pre-USDA Standardized Spanish Translation		Post-USDA Standardized Spanish Translation		
Item	Severity parameter	Severity parameter	Difference from White non-Hispanic households	Severity parameter	Difference from White non-Hispanic households	Difference-in-Difference
Adult/Household Food Security Items:						
Worried food would run out	2.64	2.48	-0.16** (0.074)	2.38	-0.27*** (0.075)	-0.10 (0.105)
Food bought did not last	3.85	3.60	-0.25*** (0.065)	3.34	-0.51*** (0.067)	-0.26*** (0.094)
Could not afford balanced meals	4.44	4.08	-0.36*** (0.064)	3.96	-0.49*** (0.066)	-0.12 (0.092)
Adult cut or skipped meals (1 or 2 vs. 0)	5.34	5.99	0.64*** (0.111)	6.17	0.83*** (0.116)	0.18 (0.161)
Adult cut or skipped meals (2 vs. 0 or 1)	5.78	6.30	0.52*** (0.126)	6.49	0.72*** (0.135)	0.20 (0.185)
Ate less than felt should	5.35	6.03	0.68*** (0.068)	6.16	0.81*** (0.073)	0.13 (0.100)
Hungry but did not eat	6.79	7.62	0.83*** (0.086)	7.61	0.82*** (0.0.91)	-0.01 (0.125)
Lost weight	7.78	8.53	0.75*** (0.104)	8.77	0.99*** (0.117)	0.24 (0.156)
Adult did not eat for whole day (1 or 2 vs. 0)	8.16	8.90	0.74*** (0.268)	8.72	0.56** (0.240)	-0.18 (0.360)
Adult did not eat for whole day (2 vs. 0 or 1)	8.36	9.01	0.65** (0.308)	8.89	0.53* (0.313)	-0.13 (0.439)
Mean—all items	7.00	7.00		7.00		
Std. deviation—all items	2.38	2.38		2.38		
Discrimination parameter	1.00	0.98		0.98		
Number of cases	7,378	2,383		2,113		

Table A.3 (Cont.): CML Rasch Model Item Severity Parameters Estimated in Separate Models for Hispanic and White Non-Hispanic Households with Children Pre- and Post-USDA Standardization of the HFSSM Spanish Translation

		Hispanic households				
	White non-Hispanic households	Pre-USDA Standardized Spanish Translation		Post-USDA Standardized Spanish Translation		
Item	Severity parameter	Severity parameter	Difference from White non-Hispanic households	Severity parameter	Difference from White non-Hispanic households	Difference-in-Difference
Child Food Security Items:						
Relied on few kinds of low-cost food to feed child(ren)	4.54	4.07	-0.47*** (0.064)	4.14	-0.40*** (0.066)	0.07 (0.092)
Couldn't feed child(ren) balanced meals	5.88	5.42	-0.46*** (0.067)	5.43	-0.45*** (0.070)	0.01 (0.097)
Child(ren) were not eating enough	7.89	6.76	-1.13*** (0.080)	7.00	-0.89*** (0.086)	0.24** (0.118)
Cut size of child(ren)'s meals	8.98	8.15	-0.84*** (0.104)	8.26	-0.73*** (0.112)	0.11 (0.153)
Child(ren) were hungry	9.24	9.15	-0.09 (0.127)	9.21	-0.03 (0.138)	0.07 (0.188)
Child(ren) cut or skipped meals (1 or 2 vs. 0)	9.33	9.00	-0.33 (0.278)	8.85	-0.48 (0.312)	-0.15 (0.418)
Child(ren) cut or skipped meals (2 vs. 0 or 1)	9.45	9.18	-0.27 (0.372)	8.99	-0.46 (0.392)	-0.19 (0.541)
Children did not eat for whole day	12.18	11.73	-0.45 (0.358)	11.62	-0.56 (0.360)	-0.11 (0.508)
Mean—all items	7.00	7.00		7.00		
Std. deviation—all items	2.38	2.38		2.38		
Discrimination parameter	1.00	0.98		0.98		
Number of cases	7,378	2,383		2,113		

Note: CML Rasch models estimated using weighted household data from the 2009-2014 CPS-FSS. Standard errors appear in parenthesis. All models were estimated on a logistic metric (discrimination parameter = 1.00). The discrimination parameter above for Hispanic and White, non-Hispanic households are those required to equate the standard deviation to be equivalent in Hispanic and White, non-Hispanic households for the purposes of that comparison. Severity parameters for the trichotomous items, *Adult cut or skipped meals*, *Adult did not eat for whole day*, and *Children cut or skipped meals*, are the "Rasch-Thurstone" or "50-percent probability" values.

* Significant at the 0.10 level. ** Significant at the 0.05 level. *** Significant at the 0.01 level.

Table A.4: Bias for the Estimated Prevalence of Food Insecurity and Very Low Food Security of Hispanic and White Non-Hispanic Households with Children Due to Differential Item Functioning

		True Prevalence		Bias, Measured versus True		White non-Hispanic versus Hispanic	
Method and Characteristic	Measured Prevalence	White non-Hispanic Household	Hispanic Household	White non-Hispanic Household	Hispanic Household	Percent of All Bias, Households	Percent of Observed Prevalence
	- - - - - - - - percent - - - - - - - -			*- percentage points -*		*- - - percent - - -*	
Pre-USDA Standardized Spanish Translation							
Food Insecurity	29.50	29.50	28.30	0.00	1.20	1.20	4.06
Very Low Food Security	13.78	13.78	14.77	0.00	-0.99	-0.99	-7.20
Post-USDA Standardized Spanish Translation							
Food Insecurity	29.50	29.50	27.96	0.00	1.54	1.54	5.21
Very Low Food Security	13.78	13.78	15.33	0.00	-1.55	-1.55	-11.25

Note: CML Rasch models estimated for the 18-item U.S. food security scale using weighted household data from the 2009-2014 CPS-FSS.

Appendix B. U.S. Adult Household Food Security Survey Module Questions

Questions asked of all households:

USDA English Adult Household Food Security Scale Module	USDA Standardized Spanish Adult Household Food Security Scale Module
"We worried whether our food would run out before we got money to buy more." Was that often, sometimes, or never true in the last 12 months?	The La primera situación es "(Me preocupó / Nos preocupamos) que la comida se podía acabar antes de tener dinero para comprar más." (Para Ud. / En su hogar), ¿ésto ocurrió frecuentemente, a veces, o nunca en los últimos 12 meses?
"The food that we bought just didn't last and we didn't have money to get more." Was that often, sometimes, or never true for you in the last 12 months?	La comida que (compré / compramos) no rindió lo suficiente, y (no tenía / no teníamos) dinero para comprar más." (Para Ud. / En su hogar), ¿ésto ocurrió frecuentemente, a veces, o nunca en los últimos 12 meses?
"We couldn't afford to eat balanced meals." Was that often, sometimes, or never true for you in the last 12 months?	"(No tenía / No teníamos) recursos suficientes para comer comida variada y nutritiva." (Para Ud. / En su hogar), ¿ésto ocurrió frecuentemente, a veces, o nunca en los últimos 12 meses?
"In the last 12 months, did you or other adults in the household ever cut the size of your meals or skip meals because there wasn't enough money for food? (Yes/No)	En los últimos 12 meses, ¿(Ud. / Ud. u otro adulto del hogar) redujo alguna vez la cantidad de sus comidas o dejó de desayunar, almorzar o cenar porque le faltaba dinero para alimentos? (Sí,No)
[If yes above, ask] How often did this happen—almost every month, some months but not every month, or in only 1 or 2 months?	[If SÍ above, ask] ¿Con qué frecuencia sucedió esto? Casi todos los meses, algunos meses pero no todos, o solamente en 1 ó 2 meses?
In the last 12 months, did you ever eat less than you felt you should because there wasn't enough money for food? (Yes/No)	En los últimos 12 meses, ¿comió Ud. alguna vez menos de lo que pensaba que debía comer porque le faltaba dinero para alimentos? (Sí,No)
In the last 12 months, were you ever hungry, but didn't eat because there wasn't enough money for food? (Yes/No)	En los últimos 12 meses, ¿Tuvo Ud. hambre alguna vez pero no comió porque le faltaba dinero para alimentos? (Sí,No)
In the last 12 months, did you lose weight because there wasn't enough money for food? (Yes/No)	En los últimos 12 meses, ¿Perdió Ud. peso porque no comió los alimentos suficientes por falta de dinero para comida? (Sí,No)
In the last 12 months did you or other adults in your household ever not eat for a whole day because there wasn't enough money for food? (Yes/No)	En los últimos 12 meses, ¿alguna vez no comió (Ud. / Ud. u otro adulto del hogar) en todo el día porque le faltaba dinero para comida? (Sí,No)
[If yes above, ask] How often did this happen—almost every month, some months but not every month, or in 1 or 2 months?	[If SÍ above, ask] ¿Con qué frecuencia sucedió esto? Casi todos los meses, algunos meses pero no todos, o solamente en 1 ó 2 meses?

Note: "Affirmative" responses indicated in bold (Bickel et al., 2000).

Appendix C: Characteristics of All Households

Table C.1: Means of Analysis Variables for Households with Income Less Than 185% of the Federal Poverty Threshold and No Extreme Responses the U.S. Adult Household Food Security Scale

		Pre-USDA Standardized Spanish Translation		Post-USDA Standardized Spanish Translation	
	All Households	White non-Hispanic Households	Hispanic Households	White non-Hispanic Households	Hispanic Households
Respondent is a non-citizen immigrant	0.149	0.017	0.426***	0.018	0.399***
	(0.003)	(0.002)	(0.10)	(0.002)	(0.010)
Non-citizen immigrant interacted with years since U.S. entry	2.307	0.224	6.475***	0.213	6.529***
	(0.054)	(0.031)	(0.193)	(0.031)	(0.201)
Spanish only language spoken by all household members	0.082	0.003	0.264***	0.003	0.221***
	(0.002)	(0.001)	(0.009)	(0.001)	(0.008)
Proportion of adults who are female	0.585	0.584	0.574	0.590	0.586
	(0.002)	(0.004)	(0.006)	(0.004)	(0.006)
Age of oldest adult	50.476	50.872	48.757***	51.768	48.769***
	(0.129)	(0.219)	(0.331)	(0.216)	(0.337)
Income/poverty ratio	0.968	1.002	0.893***	0.996	0.920***
	(0.003)	(0.006)	(0.009)	(0.006)	(0.009)
Household with child	0.480	0.405	0.665***	0.386	0.637***
	(0.004)	(0.006)	(0.009)	(0.006)	(0.010)
Most highly educated adult less than high school	0.200	0.142	0.338***	0.133	0.310***
	(0.003)	(0.004)	(0.009)	(0.004)	(0.009)
Most highly educated adult high school diploma (reference)	0.353	0.371	0.342*	0.356	0.323**
	(0.003)	(0.006)	(0.009)	(0.006)	(0.009)
Most highly educated adult some college, less than bachelor	0.334	0.363	0.245***	0.379	0.276***
	(0.003)	(0.006)	(0.008)	(0.006)	(0.009)
Most highly educated adult bachelor or other 4-year degree	0.089	0.098	0.062***	0.101	0.072***
	(0.002)	(0.004)	(0.009)	(0.004)	(0.005)
Most highly educated adult graduate or professional degree	0.024	0.026	0.013***	0.031	0.019**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Resident in principal city of metropolitan statistical area	0.281	0.190	0.465***	0.187	0.465***
	(0.003)	(0.005)	(0.010)	(0.005)	(0.010)
Resident in metropolitan statistical area, not principal city (reference)	0.339	0.344	0.333	0.334	0.346
	(0.004)	(0.006)	(0.009)	(0.006)	(0.009)
Resident in metropolitan statistical area, not identified	0.162	0.196	0.107***	0.185	0.104***
	(0.003)	(0.005)	(0.006)	(0.005)	(0.006)
Resident outside of metropolitan statistical area	0.218	0.270	0.095***	0.294	0.085***
	(0.003)	(0.005)	(0.007)	(0.005)	(0.005)
Northeast (reference)	0.156	0.158	0.147	0.156	0.160
	(0.003)	(0.005)	(0.007)	(0.005)	(0.008)
Midwest	0.212	0.283	0.074***	0.275	0.079***
	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
South	0.376	0.371	0.378	0.383	0.369
	(0.004)	(0.006)	(0.010)	(0.006)	(0.009)
West	0.256	0.188	0.400***	0.185	0.391***
	(0.003)	(0.005)	(0.09)	(0.005)	(0.009)
N	23,257	8,509	3,113	8,586	3,049

Note: Means and standard deviations (in parenthesis) estimated using weighted household data from the 2012-2014 CPS-FSS. Stars indicate statistically significant differences of means for Hispanics relative to White, non-Hispanics in each time period.
* Significant at the 0.10 level. ** Significant at the 0.05 level. *** Significant at the 0.01 level.