Using Household Fit Indices to Examine the Psychometric Quality of Food Insecurity Measures

Educational and Psychological
Measurement

2018, Vol. 78(6) 1089–1107

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164417728317

journals.sagepub.com/home/epm

**George Engelhard Jr.¹, Matthew P. Rabbitt²,
and Emily M. Engelhard³**

Abstract

This study focuses on model–data fit with a particular emphasis on household-level fit within the context of measuring household food insecurity. Household fit indices are used to examine the psychometric quality of household-level measures of food insecurity. In the United States, measures of food insecurity are commonly obtained from the U.S. Household Food Security Survey Module (HFSSM, 18 items) of the Current Population Survey Food Security Supplement (CPS-FSS). These measures, in various forms, are used to inform national programs and policies related to food insecurity. Data for low-income households with children from recent administrations of the HFSSM (2012–2014) are used in this study ($N = 7,324$). The results suggest that there are detectable levels of misfit with Infit mean square error (MSE) statistics ranging from 6.73 % to 21.33% and Outfit MSE statistics ranging from 5.31% to 9.68%. The data suggest for Outfit MSE statistics that (a) male respondents, (b) respondents with lower levels of education, and (c) respondents who did not report participating in SNAP (Supplemental Nutrition Assistance Program, formerly the Food Stamp Program) tend to have more misfit. For Infit MSE statistics, lack of homeownership appears to be a predictor of misfit. The implications of this research for future research, theory, and policy related to the measurement of household food insecurity are discussed.

Keywords

household fit, Rasch measurement theory, food insecurity

¹University of Georgia, Atlanta, GA, USA

²Economic Research Service, U.S. Department of Agriculture, Washington, DC, USA

³Feeding America, Chicago, IL, USA

Corresponding Author:

George Engelhard Jr., University of Georgia, 311 Connecticut Avenue NE, Atlanta, GA 30307, USA.

Email: gengelh@uga.edu

According to the U.S. Department of Agriculture (USDA), 12.7% of U.S. households (15.8 million) were food insecure at some time during 2015 (Coleman-Jensen, Rabbitt, Gregory, & Singh, 2016). Food insecurity is defined as limited or uncertain access to adequate food because of a lack of money or other resources. Information on food insecurity is collected from households using the U. S. Household Food Security Survey Module (HFSSM), which is administered as part of the Current Population Survey Food Security Supplement (CPS-FSS). The HFSSM is the primary instrument used to measure food insecurity in the United States (Bickel, Nord, Price, Hamilton, & Cook, 2000) and it has been adopted and modified for use in other countries, such as Canada, Mexico, Latin America, and others (M. D. Smith, Rabbitt, & Coleman-Jensen, 2017). The HFSSM is administered by the U.S. Census Bureau on behalf of the Economic Research Service of the USDA, and it has been used since 1995. Food insecurity is considered one of the major nutrition-related public health issues in the United States (Gundersen & Ziliak, 2015). Among other uses, the HFSSM is used to evaluate the effectiveness of U.S. food and nutrition assistance programs, such as the Supplemental Nutrition Assistance Program (SNAP), formerly the Food Stamp Program. These measures have been used extensively in previous studies of the causes and consequences of food insecurity (Gundersen, Kreider, & Pepper, 2011).

The classification of households as food insecure is based on 18 items in the HFSSM: 10 items relate to food hardships experienced by the general household and adults, and 8 items relate to children in the household. Households that endorse three or more items are considered food insecure under the USDA's food security status classification system (Bickel et al., 2000; Coleman-Jensen et al., 2016). These items were originally calibrated with the Rasch model (Rasch, 1960/1980) with the underlying scale maintained since that time (Coleman-Jensen, 2015; Engelhard, Engelhard, & Rabbitt, 2016). Marques, Reichenheim, de Moraes, Antunes, and Salles-Costa (2015) reviewed the literature on the measurement of food insecurity and they identified 24 measuring instruments. They concluded that the HFSSM has been evaluated the most frequently in the peer-reviewed literature. Yet none of the previous studies evaluating the psychometric quality of the HFSSM systematically examined the fit of household responses to the underlying measurement model. Previous studies focus on item fit (e.g., Nord, 2012) with the majority of these studies focused on measurement invariance across subpopulations of households.

Wells and Hambleton (2016) have described the use of residual analyses to examine model-data fit for items, and they have stressed that if data do not fit an item response theory model, then this may lead to a loss of the "all-important invariance property" (p. 395). Residual analyses of model-data fit can be summarized for both items and households. This study focuses on household fit, and the use of household fit indices to evaluate the validity of the response patterns for each household. As Messick (1995) has pointed out, validity evidence should be collected related to both items and persons responses. The validity of household response patterns is particularly important for empirical applications of food insecurity measures based on the

U.S. Food Security Scale because the USDA assigns each household a food security status based on the sum score. This approach implicitly assumes that households with the same sum scores experience the same type of food hardship. By examining household fit, our study examines whether or not this assumption is violated, and this may potentially bias comparisons between households.

In our study, households are viewed as conceptually equivalent to persons, and we draw on the literature related to person fit for our approach (Engelhard, 2009; Ferrando, 2015). Household fit provides additional information regarding the validity of the inferences for individual households that go beyond the sum scores. For more information on food security status classification, see Coleman-Jensen et al. (2016). Emerging research also supports using fit indices in explanatory models to examine patterns of misfit (Walker & Engelhard, 2015). In particular, this study includes dichotomous indices of household fit in explanatory logistic regression models with a set of respondent and household characteristics that are frequently used in economic research on food insecurity.

Purpose

The purpose of this study is to describe and illustrate the use of household fit indices to enhance our understanding of the measurement of food insecurity in the United States. In essence, the conceptual framework described in this study provides validity evidence related to estimates of food insecurity for each household based on item response patterns. The literature on person fit supports the potential for understanding food insecurity by examining how well households fit the underlying measurement model. In addition to examining household fit, this study explores the relationships between household fit indices and selected respondent and household variables that are typically used by economists to predict food insecurity. This research explores whether or not there are patterns of mismeasurement for households in some food security status categories that are essential for informing national policy on food insecurity.

Method for Assessing Model–Data Fit for Households

Rasch (1960/1980) described a measurement theory based on the requirements of specific objectivity that supports invariant measurement. Invariant measurement (Engelhard, 2013) provides meaningful scores that maintain their meaning over different contexts when appropriate model–data fit is obtained. To realize the desirable properties of Rasch measurement theory, an appropriate level of model–data fit is required.

The dichotomous Rasch model can be written as follows:

$$\phi_{ni1} = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}, \quad (1)$$

where, ϕ_{ni1} represents the probability of household n with a location of θ_n on a latent variable giving a positive response (denoted as 1) to item i with a location of δ_i on the latent variable. In our study, positive responses are endorsements of the items designed to conceptualize food insecurity.

Once estimates for household and item locations are obtained, then Equation (1) can be used to obtain the expected (predicted) probabilities, P_{ni} , as follows:

$$P_{ni1} = \frac{\exp(\hat{\theta}_n - \hat{\delta}_i)}{1 + \exp(\hat{\theta}_n - \hat{\delta}_i)}.$$

Based on the dichotomous Rasch model in Equation (1), an estimate of the response variances (statistical information), Q_{ni} , can be obtained as follows:

$$Q_{ni} = P_{ni}(1 - P_{ni}).$$

Score residuals, Y_{ni} , can be defined as

$$Y_{ni} = X_{ni} - P_{ni},$$

where X_{ni} are the observed dichotomous responses. A set of standardized residuals, Z_{ni} , can be defined as follows:

$$Z_{ni} = Y_{ni} / Q_{ni}^{1/2}.$$

Several residual-based fit indices are typically used in Rasch measurement theory: Infit and Outfit mean square error (MSE) statistics. These Infit and Outfit MSE statistics are based on residuals, and the residuals can be summarized and reported for both items and households. For example, the Outfit MSE statistic for household n can be defined as

$$\text{Outfit } MSE_n = \frac{\sum_i^L Z_{ni}^2}{L}, \quad (2)$$

where Z_{ni}^2 are the squared standardized residuals, and L is the number of items. Outfit MSE statistics are unweighted averages, and they are sensitive to extreme unexpected residuals and outliers (Smith, 2004). Infit MSE statistics provide an information-weighted version of the MSE statistic. They can be defined for household n as

$$\text{Infit } MSE_n = \frac{\sum_i^L Y_{ni}^2}{\sum_i^L Q_{ni}}, \quad (3)$$

where Y_{ni}^2 are the squared residuals, Q_{ni} is the variance of the expected response probabilities, and L is the number of items. Because Infit MSE statistics are

information-weighted averages, these statistics are less sensitive to extreme outliers (Smith, 2004). The residuals can be summarized in a similar fashion over items to yield Infit and Outfit MSE for items. When data fit the Rasch model, the expected value of Infit and Outfit MSE statistics is 1.00 (Engelhard, 2013), and the MSE statistics can range from 0 to positive infinity. MSE values greater than 1.00 indicate response patterns with more variation than expected, whereas values less than 1.00 tend to reflect less variation than expected based on the measurement model. Bond and Fox (2015), Engelhard (2013), and Smith (2004) provide more details regarding the conceptual framework and procedures for calculating Infit and Outfit MSE statistics.

Numerous studies have evaluated the utility of these MSE statistics based on theoretical and empirical sampling distributions (Karabatsos, 2000; R. M. Smith, 1991; R. M. Smith & Hedges, 1982; R. M. Smith, Schumacker, & Busch, 1982). There have been conflicting recommendations in the literature regarding the appropriate use of MSE statistics in this context. It is beyond the scope of this study to revisit the discussions of appropriate model–data fit indices in this study. We view MSE statistics as descriptive indices that can be used as effect sizes to describe various levels of misfit for items and households. For the purposes of this study, we use the following descriptive guidelines adapted from Wright and Linacre (1994) for labeling and interpreting MSE statistics:

Dichotomous fit indices of household misfit are used in this study as dependent variables in exploratory logistic regression models. The selection of these explora-

Mean square error (MSE)	Interpretation	Category
$0.50 \leq \text{MSE} < 1.50$	Productive for measurement	A
$\text{MSE} < 0.50$	Less productive for measurement, but not distorting of measures	B
$1.50 \leq \text{MSE} < 2.00$	Unproductive for measurement, but not distorting of measures	C
$\text{MSE} \geq 2.00$	Unproductive for measurement, distorting of measures	D

tory variables was based on a consideration of variables typically used in economic analyses of food insecurity. These variables capture information related to selected respondents and household characteristics. Prior studies have used these variables to explore the validity of the U.S. Food Security Scale. For example, Opsomer, Jensen, and Pan (2003) and Nord (2012) use many of these variables to explore measurement invariance (differential item functioning) related to demographic and survey-specific variables in the U.S. Food Security Scale. Among other uses, these exploratory variables have also been used in economic models of food insecurity to provide insight into the effectiveness of food assistance programs, such as SNAP. For an extensive review of this literature see Gregory, Rabbitt, and Ribar (2015). Consequently, our

logistic regression models with household misfit indices as the dependent variables have the potential to inform researchers working on measurement issues related to food insecurity. The list of the selected characteristics is shown in Table 1.

In addition to reporting summary values for the MSE fit statistics and the results of the logistic regression analyses, it is important to recognize that households can “misfit” for a variety of reasons. A productive way to diagnose individual household misfit is to use household response functions (HRFs). There are several approaches for graphically presenting model–data fit for households. In this study, our first approach is to plot standardized residuals (y -axis) based on the order of item administration (x -axis). The HFSSM items are administered in order from Item 1 to Item 18. In our second approach, HRFs are used to diagnose misfit for each household with HRFs viewed as analogous to person response functions. For the HRFs, the y -axis reflects the probability of endorsing an item and the x -axis is ordered by item difficulty. We use a nonparametric method for empirically smoothing the HRFs based on a Tukey–Hann approach for estimating HRFs (Tukey, 1977; von Hann, 1903). This approach has been previously used to analyze both item fit (Jennings & Engelhard, 2017), as well as person fit (Walker, Engelhard, Royal, & Hedgpeth, 2016). The smoothed estimates used to define the Tukey–Hann household response functions (TH-HRFs) are based on a running weighted average. Theoretically, there are an infinite number of weighted running averages that could exist with the only requirement being that the weights sum to 1. Velleman and Hoaglin (1981) suggested the following function:

$$z_i = 0.25(y_{i-1}) + 0.50(y_i) + 0.25(y_{i+1}),$$

where z_i is the new smoothed value, y_i is the current value, and (y_{i-1}) and (y_{i+1}) are the adjacent values below and above the current value, respectively.

The data can be smoothed iteratively, and this resmoothing can be used to tune the amount of roughness in the smoothed curve (Velleman & Hoaglin, 1981). This iterative running weighted average procedure is used to estimate the TH HRFs. The number of iterations is based on the total number of positive responses within each household to the HFSSM. The TH HRFs can be compared with the theoretical response function based on the Rasch model. Differences between the two functions provide diagnostic information for exploring misfit. It should be stressed that misfit statistics can be obtained from different patterns in the residuals, and that it is essential to examine a display in order to diagnose sources of misfit. This study illustrates the use of graphical displays of residuals to explore details of the fit of individual household response patterns. Wells and Hambleton (2016) highlight the importance of using graphical displays for items, and the same principle applies to households.

Table 1. Summary of Household Misfit for Selected Respondent and Household Characteristics (Cell Entries are Percentages).

Variables	All	Infit MSE		Outfit MSE	
		≥ 1.5	≥ 2.0	≥ 1.5	≥ 2.0
Respondent characteristics					
Female	68.38	68.37	67.87	60.79*	60.45*
Male	31.62	31.63	32.13	39.21*	39.55*
High school degree/GED or lower education attainment	60.28	61.97	65.81*	64.32*	64.71*
Some college or higher educ. attainment	39.72	38.03	34.19*	35.68*	35.29*
Employed full-time or part-time	56.40	53.33*	52.70	53.31	52.74
Unemployed or not in the labor force	43.60	46.67*	47.30	46.69	47.26
Household characteristics					
Adult members only speak Spanish	5.94	6.27	9.25*	8.32*	7.91
Adult members speak Spanish or another language	94.06	93.73	90.75*	91.68*	92.09
At least one child younger than 6 years	49.90	48.78	46.53	50.21	50.71
No children younger than 6 years	50.10	51.22	53.47	49.79	49.29
Metropolitan household	74.75	75.10	75.84	75.74	77.28
Nonmetropolitan household	25.25	24.90	24.16	24.26	22.72
SNAP receipt, past 12 months	56.84	57.68	52.70	54.30	52.33*
No SNAP receipt, past 12 months	43.16	42.32	47.30	45.70	47.67*
Home owned by a household member	33.56	33.29	29.56	36.25	34.48
Home rented by a household member	66.44	66.71	70.44	63.75	65.52
Income below poverty line	57.76	58.71	57.33	57.12	57.81
Income above poverty line	42.24	41.29	42.67	42.88	42.19
Number of households (N)	7,324	1,562	493	709	389
Percentage of total		21.33	6.73	9.68	5.31

Note. MSE = mean square error; SNAP = Supplemental Nutrition Assistance Program. Characteristics are coded dichotomously in this table. Unweighted household data from the 2012-2014 Current Population Survey Food Security Supplement are used in this table. MSE values ≥ 1.5 and ≥ 2.0 are defined as misfitting households for the purposes of this study. Statistical analyses test the null hypothesis that the mean of the characteristic being considered is not equal to the mean for households who do fit based on the fit statistic being considered.

* $p \leq .05$.

Method

Participants

This study includes all the households surveyed in the 2012, 2013, and 2014 Current Population Survey Food Security Supplement (CPS-FSS) with income less than 185% of the federal poverty line, and who indicated having at least one child younger than 18 years in the household. The income threshold of 185% of the federal poverty line was used because it represents the income screening threshold for a household to be administered the HFSSM. Households above the income threshold are also administered the HFSSM if they showed any signs of food stress; however, they represented a small proportion of the households administered the HFSSM during our analysis period. Omitting higher income households from the analyses mitigates bias associated with income that could result from the screen (Nord, 2012). The total number of households included in this study is 7,324. Column 2 in Table 1 (All) indicates the percentage of households by selected respondent and household characteristics.

Instrument

Each year, the HFSSM is administered to households as part of the CPS-FSS, a supplement to the monthly Current Population Survey. The HFSSM consists of 18 items that are designed to capture the conditions and behaviors characterizing households having difficulty meeting their food needs. The survey module is described in detail on the USDA website (<https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/measurement/#measurement>)

It should be noted that some of the items include multiple categories, and that USDA recodes all items dichotomously for their policy work on food insecurity. Nord (2012) assessed the impact of dichotomizing responses to the HFSSM, and he found that it had no significant effect on the U.S. Food Security Scale. We follow the current item scoring practices recommended by the USDA in this study (Bickel et al., 2000). The first 10 items are related to the household and adult food insecurity. The last eight items relate to issues of food insecurity for children in the household. The plots of the official item calibrations by the USDA (Bickel et al., 2000) with current calibrations from this study are shown in Figure 1 (Engelhard et al., 2016). The item difficulties on the HFSSM have remained remarkably invariant over time.

Procedures

We used a three-component conceptual framework for examining model–data fit proposed by Engelhard (2009). The three components are (a) identification of a theoretical definition of measurement, (b) selection of a measurement model to reflect this philosophical view of measurement, and (c) conduct detailed numerical and graphical analyses of residuals. Our theoretical view of measurement is based on the ideal of invariant measurement (Engelhard, 2013) with the use of the Rasch model to

Figure 1. Plot of item calibrations from current study and official USDA values for the U.S. Food Security Scale.

Note. USDA = U.S. Department of Agriculture. Item calibrations used by USDA are scaled to range between 0 and 14 logits with a mean of 7. See Bickel, Nord, Price, Hamilton, and Cook (2000) for more details on USDA calibrations. For the current study, data from the 2012-2014 Current Population Survey Food Security Supplement were used.

represent the measurement model for evaluating invariance. For the third component, we used dichotomous indices of household misfit based on the MSE fit statistics. Logistic regression analyses were conducted with dichotomous misfit indices of household fit as dependent variables with selected respondent and household characteristics as predictors of household misfit. See Table 1 for a listing of the variables used in these analyses. We also used graphical displays to illustrate the use of residual analyses and HRFs for selected households.

Results

The Wright Map for the Household Food Insecurity Scale is shown in Figure 2. The first column is the underlying logit scale, the second column shows the distribution of the households on the food insecurity scale (counts), the third column indicates the locations of the items, and the last column includes a descriptive label for the items. The item calibrations range in severity from Item 18 (Child[ren] not eat for whole day) that is the hardest to endorse to Item 1 (Worried food would run out) that is the easiest to endorse.

The summary statistics from the Rasch analyses are shown in Table 2. The mean location of the items is centered at zero logits ($SD = 2.99$), while the mean location of the households is -2.61 logits ($SD = 2.24$). Overall, model–data fit is quite good with means of both the Infit and Outfit statistics close to the expected values of one. The reliability of household separation is 0.84, and this value is comparable in meaning to

Figure 2. Wright Map of the U.S. Food Security Scale for households with children.

Table 2. Rasch Summary Statistics for Household Food Insecurity Scale for Households With Children.

	Households	Items
Measure		
Mean	-2.61	0.00
Standard deviation	2.24	2.99
Infit mean square error		
Mean	0.98	0.97
Standard deviation	0.55	0.15
Outfit mean square error		
Mean	0.72	0.89
Standard deviation	1.19	0.61
Reliability of separation	0.84	> 0.99
Number of observations	7,324	18

coefficient alpha in classical test theory. Table 3 shows the details of the item analyses, and the item-fit indices are quite good. It is interesting to note that Item 1 (the first item administered) has the highest Outfit misfit statistic of 2.28 (“worried food would run out”). It is also interesting to note that all the item Infit statistics are within the range for acceptable item fit ($0.50 \leq \text{MSE} < 1.50$). As pointed out earlier, previous research on evaluating and maintaining the psychometric quality of the HFSSM based on model–data fit have focused primarily on the fit of the items to the Rasch model. As can be seen in Tables 2 and 3, item fit is quite good. Figure 1 illustrates that the item calibrations have been remarkably stable and robust.

Turning now to the dichotomous indicators of household misfit, 21.33 % of the households had Infit $\text{MSE} \geq 1.5$, and 6.73% of the households had Infit $\text{MSE} \geq 2.0$. Estimates of misfitting households ranged from 9.68% based on Outfit $\text{MSE} \geq 1.5$ to 5.31% as indicated by Infit $\text{MSE} \geq 2.0$. It is interesting to note that the number of misfitting households is estimated to be somewhat higher based on the Infit MSE as compared with the Outfit MSE indices.

Table 1 presents a summary of household misfit based on selected respondent and household characteristics for Infit and Outfit statistics. Educational attainment of the household’s most highly educated adult, employment status, and households with members who only speak Spanish are the only characteristics identified as influencing misfit for the Infit MSE, although the characteristics vary depending on the criteria used for classifying misfit based on Infit MSE. There were statistically significant differences in household misfit based on the Outfit MSE with four characteristics related to household misfit: (a) gender of the respondent, (b) educational attainment of the household’s most highly educated adult, (c) Spanish as the only language adult household members speak (only when using ≥ 2.0 as the threshold), and (d) participation in SNAP within the past 12 months (only when using ≥ 2.0 as the threshold).

Logistic regression models were used to estimate the average partial effects (APEs) for the selected independent variables. These results are shown in Table 4. APEs are particularly useful in this context because they describe how changes in the probability of a household exhibiting misfit is associated with changes in the predictor variable of interest. Conceptually, APEs from nonlinear models, such as the logistic model, are similar to coefficients obtained from an ordinary linear regression model. A useful introductory discussion of APEs and how they are calculated is provided by Wooldridge (2015).

The analyses based on the logistic regression models suggest that a respondent’s employment status and homeownership are statistically significant predictors of household misfit after controlling for other characteristics in the model when misfit is identified according to the household Outfit MSE index. Overall, four respondent and household characteristics appear to be statistically significant predictors of household misfit based on the logistic regression models; however, the effects of these characteristics appear to vary across the Infit MSE and Outfit MSE models. The only characteristic that consistently predicts household misfit, regardless of the index used, is

Table 3. Summary Statistics for Items From the Household Food Insecurity Scale.

Label	Item number	Item location	Standard error	Infit MSE	Category	Outfit MSE	Category
Household and adult items							
Worried food would run out	1	-5.45	0.04	1.02	A	2.28	D
Food bought would not last	2	-4.11	0.03	0.99	A	1.92	C
Could not afford to eat balanced meals	3	-3.28	0.03	1.11	A	1.47	A
Adult(s) cut size or skipped meals	4	-1.62	0.03	0.77	A	0.57	A
Respondent ate less than should have	5	-1.78	0.03	0.81	A	0.63	A
Adult(s) cut or skipped meals, frequency follow up	6	-0.95	0.04	0.80	A	0.57	A
Respondent hungry but did not eat	7	-0.14	0.04	0.86	A	0.57	A
Respondent lost weight	8	1.08	0.05	1.06	A	0.69	A
Adult(s) not eat for whole day	9	1.91	0.06	0.96	A	0.45	B
Adult(s) not eat for whole day, frequency follow-up	10	2.39	0.07	0.95	A	0.30	B
Child items							
Relied on low-cost foods for child(ren)	11	-3.21	0.03	1.25	A	1.90	C
Could not feed child(ren) balanced meals	12	-1.69	0.03	1.13	A	1.12	A
Child(ren) not eating enough	13	0.33	0.04	1.10	A	1.03	A
Cut size of child(ren)s meals	14	1.57	0.05	1.04	A	0.78	A
Child(ren) hungry	15	2.31	0.06	0.84	A	0.36	B
Child(ren) skipped meals	16	3.28	0.08	0.82	A	0.36	B
Child(ren) skipped meals, frequency follow-up	17	3.67	0.09	0.77	A	0.22	B
Child(ren) not eat for whole day	18	5.67	0.19	1.13	A	0.73	A
Mean		0.00	0.06	0.97		0.89	
Standard deviation		2.99	0.04	0.15		0.61	

Note. MSE = mean square error. The category labels reflect groupings of the MSE fit statistics: A is ($0.50 \leq \text{MSE} < 1.50$), B is ($\text{MSE} < 0.50$), C is ($1.50 \leq \text{MSE} < 2.00$), and D is ($\text{MSE} \geq 2.00$). MSE values ≥ 1.5 indicate that measures may be unproductive for measurement, whereas MSE values ≥ 2.0 indicate that measures may be unproductive for measurement, and possibly distorting of measures.

Table 4. Average Partial Effects From Logistic Models on Household Misfit Indicators From the Household Food Insecurity Scale for Households With Children.

	Infit MSE		Outfit MSE	
	≥ 1.5	≥ 2.0	≥ 1.5	≥ 2.0
Respondent characteristics				
Female	−0.004	−0.001	−0.033*	−0.025*
No college education	0.012	0.010	0.011	0.009
Employed full-time or part-time	−0.026*	−0.010	−0.019*	−0.016*
Household characteristics				
Adult members only speak Spanish	0.010	0.025	0.034*	0.015
At least one child younger than 6 years	−0.011	−0.008	0.001	0.002
Metropolitan household	0.004	0.000	0.006	0.009
SNAP receipt, past 12 months	0.003	−0.011	−0.008	−0.014*
Home owned by household member	−0.002	−0.011*	0.009	0.000
Income below poverty line	0.000	−0.002	−0.002	0.003
Log-likelihood	−3,790.159	−1,508.942	−2,310.064	−1,790.641
Somers' D	7,580.318	3,017.885	4,620.128	3,581.282
Number of households	7,324	7,324	7,324	7,324

Note. MSE = mean square error; SNAP = Supplemental Nutrition Assistance Program. Characteristics are coded dichotomously in this table. Unweighted data from the 2012-2014 Current Population Survey Food Security Supplement.
* $p \leq .05$.

the respondent's employment status. The APEs for the respondent's employment status are consistently negatively signed, indicating households with an employed respondent are less likely to exhibit household misfit. Similarly, households with a female respondent, self-reported SNAP receipt within the past 12 months, and home-ownership are associated with a reduction in the probability of a household exhibiting misfit. The only statistically significant and positive APE in Table 4 occurs for households with adult members who only speak Spanish when household misfit is indicated using the household Infit MSE misfit index. This is a surprising finding since the USDA implemented a standardized translation of the HFSSM in 2012 (Rabbitt & Coleman-Jensen, 2015). However, it may suggest that households with different cultural backgrounds experience food insecurity differently.

In addition to numerical indices of household misfit based on MSE statistics, it is important to recognize that households can misfit with different item response patterns. To illustrate the substantive interpretation of household fit indices, several graphical approaches are shown in this study.

First, we selected Household 12 to represent a household with acceptable model-data fit (Outfit = 1.008). Panel A in Figure 3 shows observed and expected values for Household 12 based on the Rasch model, while Panel B shows the standardized residuals. The residual analyses show that the majority of the responses are between

Figure 3. Household fit (Panel A) and residual plot (Panel B) with x-axis defined by item administration order: Acceptable model–data fit.

–2.00 and 2.00 suggesting good fit. It is also important to note that in Figure 3 the x-axes are defined by order of item administration. It is interesting to note that Panel A in Figure 3 suggests that there may be two underlying dimensions related to content of the items. The reader will recall that Items 1 to 10 reflect household and adult issues, while Items 11 to 18 reflect child-related issues of food insecurity.

Next, we selected Household 537 to reflect a household with comparable food insecurity with Household 12, while exhibiting unacceptable model–data fit (Outfit = 4.553). The observed and expected values are shown in Figure 4 (Panel A), and the standardized residual plot in Figure 4 (Panel B). Panel A in Figure 4 shows a similar pattern to Panel A in Figure 3. The standardized residuals in Figure 4 (Panel B) also are similar to Figure 3 (Panel B) with fewer standardized

Figure 4. Household fit (Panel A) and residual plot (Panel B) with x-axis defined by item administration order: Unacceptable model–data fit.

residuals between -2.00 and 2.00 suggesting poor fit. Future research should explore the possibility of within-household multidimensionality in responding to adult and child items on HFSSM.

Figure 5 presents the expected HRFs based on the Rasch model, as well as empirically smoothed functions based on the Tukey–Hann procedure. Panel A in Figure 5 shows Household 12 (acceptable model–data fit), while Panel B shows Household 537 (unacceptable model–data fit). In contrast to Figures 3 and 4, the x-axes in Figure 5 are defined by item difficulty order. Figure 5 provides another way to visualize how response probabilities are related to item calibrations for two households that vary in terms of the level of household fit. For example, Panels A and B in Figure 5 show differences in amplitude that may aid researchers in conceptualizing household misfit. The distinctive pattern of misfit in Figure 5 (Panel B) may also be indicative of multidimensionality within a household related to household food insecurity.

Figure 5. Tukey–Hann household (Panel A, Household 12; Panel B Household 537) response functions with x-axis defined in item difficulty order.

Discussion

It is essential to develop valid measures of food insecurity to inform research, theory, and policy. The psychometric quality of measures of food insecurity have been widely evaluated (Marques et al., 2015; Nord, 2012), but food insecurity measures have not been evaluated based on household model–data fit. Model–data fit has historically focused on item fit. In this study, we examine this gap in the research by providing a perspective on evaluating household fit. Household fit can be viewed as conceptually equivalent to person fit (Ferrando, 2015). It is important to stress that

the desirable properties of invariant measurement, such as item-invariant measurement of food insecurity, depend on good model–data fit (Engelhard, 2013).

Our findings suggest that there are detectable levels of misfit with Infit MSE statistics ranging from 6.73 % to 21.33% and Outfit MSE statistics ranging from 5.31% to 9.68%. Furthermore, the data suggest for Outfit MSE statistics that (a) male respondents, (b) respondents with lower levels of education, and (c) respondents who did not report participating in SNAP (Food Stamp Program) tend to have more misfit. For Infit MSE statistics, lack of homeownership appears to be a predictor of misfit.

Future research should explore other potential indicators of household fit based on research conducted on person fit (Ferrando, 2015; Karabatsos, 2000). Different indices are sensitive to different facets of household misfit, and this line of research has the potential to yield better estimates of household food insecurity that may be less biased. In addition to exploring other indices of household fit, additional research should focus on a variety of explanatory variables that may lead to identifying the household conditions that may affect misfitting responses from respondents. Mixed methods research using household fit indices and various qualitative methods may contribute to the development of a deeper understanding of household food insecurity in the United States.

In summary, this study provides a framework for examining aspects of household fit related to the measurement of food insecurity. Food insecurity is an important construct that reflects quality of life, and it is essential to develop valid, reliable, and fair indicators of household food insecurity.

Authors' Note

Authors' Note An earlier version of this article was presented at the International Meeting of the Objective Measurement Society, Washington, DC, April 2016. The views expressed in this article are those of the authors and do not necessarily reflect those of the USDA.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Bickel, G., Nord, M., Price, C., Hamilton, W., & Cook, J. (2000). *Guide to measuring household food security: Revised 2000*. Washington, DC: U.S. Department of Agriculture, Food and Nutrition Service.

- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Coleman-Jensen, A. (2015). *Commemorating 20 Years of U.S. Food Security Measurement* Amber Waves (feature). Washington, DC: U.S. Department of Agriculture, Economic Research Service.
- Coleman-Jensen, A., Rabbitt, M. P., Gregory, C. A., & Singh, A. (2016). *Household Food Security in the United States in 2015*. Washington, DC: U.S. Department of Agriculture, Economic Research Service.
- Engelhard, G. (2009). Using IRT and Model-Data fit to Conceptualize Differential Item and Person Functioning for Students with Disabilities. *Educational and Psychological Measurement*, 69(4), 585-602.
- Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge.
- Engelhard, G., Jr., Engelhard, E., & Rabbitt, M. P. (2016). Measurement of household food insecurity: Two decades of invariant measurement. *Rasch Measurement Transactions*, 30, 1598-1599.
- Ferrando, P. J. (2015). Assessing person fit in typical-response measures. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 128-155). New York, NY: Routledge.
- Gregory, C. A., Rabbitt, M. P., & Ribar, D. C. (2015). The Supplemental Nutrition Assistance Program and food insecurity. In J. P. Ziliak, J. Bartfield, T. Smeeding, & C. Gundersen (Eds.), *SNAP matters: How food stamps affect health and well-being* (pp. 74-106). Stanford, CA: Stanford University Press.
- Gundersen, C., Kreider, B., & Pepper, J. (2011). The economics of food insecurity in the United States. *Applied Economic Perspectives and Policy*, 33, 281-303.
- Gundersen, C., & Ziliak, J. P. (2015). Food insecurity and health outcomes. *Health Affairs*, 34, 1830-1839.
- Jennings, J. K., & Engelhard, G. (2017). Evaluation of model-data fit by comparing parametric and nonparametric item response functions: Application of a Tukey-Hann Procedure. *Journal of Applied Measurement*, 18, 1-13.
- Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement*, 1, 152-176.
- Marques, E. S., Reichenheim, M. E., de Moraes, C. L., Antunes, M. M., & Salles-Costa, R. (2015). Household food insecurity: A systematic review of the measuring instruments used in epidemiological studies. *Public Health Nutrition*, 18, 877-892.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Nord, M. (2012). *Assessing potential technical enhancements to the US Household Food Security Measures*. Washington, DC: U.S. Department of Agriculture, Economic Research Service.
- Opsomer, J. D., Jensen, H. H., & Pan, S. (2003). An evaluation of the US Department of Agriculture food security measure with generalized linear mixed models. *Journal of Nutrition*, 133, 421-427.
- Rabbitt, M., & Coleman-Jensen, A. (2015). *Does a Spanish translation of the U.S. household food security survey module affect differential item functioning between Hispanic and*

- White non-Hispanic households?* Washington, DC: Federal Committee on Statistical Methodology Research Conference.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Copenhagen, Denmark: Danish Institute for Educational Research. (Original work published in 1960).
- Smith, M. D., Rabbitt, M. P., & Coleman-Jensen, A. (2017). Who are the world's food insecure? New evidence from the food and agriculture organization's food insecurity experience scale. *World Development*, 93, 402-412.
- Smith, R. M. (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*, 51, 541-565.
- Smith, R. M. (2004). Fit analysis in latent trait measurement models. In E. Smith, Jr., and R. Smith (Eds.), *Introduction to Rasch Measurement* (pp. 73-92). Maple Grove, MN: JAM Press.
- Smith, R. M., & Hedges, L. V. (1982). Comparison of likelihood ratio $\div^2$ and Pearsonian χ^2 tests of fit in the Rasch model. *Education Research & Perspectives*, 9(1), 44-54.
- Smith, R. M., Schumacker, R. E., & Busch, M. J. (1982). Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement*, 2, 66-78.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Boston, MA: Duxbury Press.
- von Hann, J. (1903). *Handbook of climatology*. New York, NY: Macmillan.
- Walker, A. A., & Engelhard, G. (2015). Exploring person fit with an approach based on multilevel logistic regression. *Applied Measurement in Education*, 28, 274-291.
- Walker, A. A., Engelhard, G., Royal, K., & Hedgpeth, M. (2016). Exploring aberrant responses using person fit and person response functions. *Journal of Applied Measurement*, 17, 194-208.
- Wells, C. S., & Hambleton, R. K. (2016). Model fit with residual analyses. In W. J. van der Linden (Ed.), *Handbook of item response theory: Vol. 2. Statistical tools* (pp. 395-413). Boca Raton, FL: CRC Press.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Scarborough, Ontario, Canada: Nelson Education.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.