JOURNAL OF APPLIED MEASUREMENT, *24*(3/4), 88-101


# Using Explanatory Item Response Models to Evaluate Surveys

Jing Li
*University of Georgia*

George Engelhard
*University of Georgia*

This study evaluates the psychometric quality of surveys by using explanatory item response models. The specific focus is on how item properties can be used to improve the meaning and usefulness of survey results. The study uses a food insecurity survey (HFSSM) as a case study. The case study examines 500 households with data collected between 2012 and 2014 in the United States. Eleven items from the HFSSM are classified in terms of two item properties: referent (household, adult, and child) and content (worry, ate less, cut meal size, hungry, and not eat for the whole day). A set of explanatory linear logistic Rasch models is used to explore the relationships between these item properties and their locations on the food insecurity scale. The results suggest that both the referent and item content are significant predictors of item location on the food insecurity scale. It is demonstrated that the explanatory item response models are a potential method for examining the psychometric quality of surveys. Explanatory item response models can be used to enhance the meaning and usefulness of survey results by providing insights into the relationship between item properties and survey responses. This approach can help researchers improve the psychometric quality of surveys and ensure that they are measuring what they intend to measure. It can lead to better-informed policy decisions and interventions aimed at tackling social issues such as food insecurity, poverty, and inequality.

*Keywords:* Explanatory item response models, linear logistic Rasch model, household food insecurity, surveys

Requests for reprints should be sent to (1) Jing Li, 160 Scandia Cir. Apt 4. Athens, GA 30605, USA; email: jl46730@uga.edu, and/or (2) George Engelhard Jr., 311 Connecticut Ave. NE, Atlanta 30307, USA; email: gengelh@uga.edu

Surveys are a commonly used research method in a variety of fields within social sciences (Somekh & Lewin, 2005), and they can be used to collect data on a wide range of topics such as attitudes, behaviors, demographics, opinions, and preferences (Cook et al., 2002). When conducting a study, researchers gather self-reported data/responses from participants who voluntarily agree to take part in the study. Later on, they conduct analysis and draw conclusions based on these data/responses. The problem with this procedure lies in the extent to which the answers obtained can accurately represent the intended measure by the researchers. This is linked to the validity and reliability of the survey response, which affects the trustworthiness of the survey data/responses and the interpretation of results (Dillman et al., 2014).

Item response models, such as the Rasch model (Rasch, 1960/1980), are being increasingly applied to the development of survey items. Surveys developed based on Rasch measurement theory promote the idea of using survey data/responses to achieve invariant measurement when good model-data fit is obtained, which includes both item-invariant of person measures and person-invariant of item measures. Item-invariant of person measurement indicates that a person's location on a latent variable is not dependent on different items being used, and person-invariant of item measurement suggests the calibrated items developed using Rasch models can be used for measurement that is not dependent on different groups of people. Good model-data fit also supports the inference that item locations on a scale are conditionally independent of persons (Engelhard, 2013; Hambleton, 2000).

A broader approach to measurement can be based on explanatory item response models (EIRMs). Under the explanatory approach, item properties and person properties can be used to explain a person's responses to items (De Boeck & Wilson, 2004). From the statistical approach, *generalized linear mixed models* (GLMM) fall into this framework by using covariates to predict item/person responses (either item properties, person properties, the interaction of item and person properties, or neither of them). This study uses one of the generalized linear mixed models, specifically, an item explanatory model—the linear logistic Rasch model (LLRM). The LLRM has also been referred to as the linear logistic test model (LLTM). The LLRM is an item-explanatory model that uses item properties to explain item responses while treating person as a random effect (Wilson et al., 2008).

Food insecurity is an important component of quality of life (Campbell, 1991). In the United States, the Household Food Security Survey Module (HFSSM) is commonly used to track food insecurity. It is part of the yearly Current Population Survey Food Security Supplement (CPS-FSS) conducted by the United States Department of Agriculture (USDA; Economic Research Service, n.d.). This survey module was developed to capture and distinguish the severity levels over the entire severity spectrum of the issue of food insecurity (Bickel et al., 2000). We use an 11-item scale based on the HFSSM to demonstrate the framework proposed in this study. The original HFSSM was developed based on Rasch measurement theory (Marques et al., 2015), and it is viewed as producing reliable and valid score inferences for measuring food insecurity examined by multiple researchers (Derrickson et al., 2000; Wilde, 2004).

Previous studies have examined the HFSSM from multiple psychometric perspectives, including model-data fit, differential item functioning, bifactor analysis, and so on (Engelhard et al., 2018; Tanaka et al., 2019, 2020). This survey has not been examined using LLRM before. In this study, we examine the measurement quality of 11 items selected from HFSSM, and we explore how item characteristics influence item responses using linear logistic Rasch models (LLRM). Our approach also examines validity evidence based on test content as advised by the *Standards for Educational and Psychological Testing*

(American Educational Research Association [AERA] et al., 2014).

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) suggest that the content of survey items, as they relate to the subject matter being tested, can impact the way test scores are interpreted. Additionally, evidence of validity based on the content of the survey can be utilized to address any concerns about differences in the interpretation of scores among different subgroups of respondents (AERA et al., 2014). Therefore, our approach can provide validity evidence based on item content that promotes people's understanding and interpretation of survey responses.

### Explanatory Item Response Models

In explanatory item response modeling (EIRM), item responses are seen as dependent variables, and the properties of person, items, or interaction of both serve as independent variables. When dealing with survey data, generalized linear mixed modeling (GLMM) is applied to convert usable dependent variables and connect independent variables to dependent variables. In this explanatory approach, the data/responses are something researchers tried to explain, and creating a valid measurement is still the goal of researchers. What's more, the EIRM lets the measurement be explanatory by linking responses to known properties of items (De Boeck & Wilson, 2016).

### Linear Logistic Rasch Model

In this study, the linear logistic Rasch model (LLRM) is the EIRM being used. The simple dichotomous Rasch model is a Rasch model for dichotomous response (0 and 1) created by Georg Rasch (Rasch, 1960/1980). The model can be written in general as (Engelhard, 2013):

$$\phi_{nik} = \frac{P_{nik}}{P_{nik-1} + P_{nik}} = \frac{\exp\left(\theta_n - \delta_i\right)}{1 + \exp\left(\theta_n - \delta_i\right)}.$$

When the response equals 1 (for correct or higher ranking), the conditional probability $\Pr(x = 1)$ is:

$$P_{ni1} = \frac{\exp\left(\theta_n - \delta_i\right)}{1 + \exp\left(\theta_n - \delta_i\right)}.$$

When the response equals 0 (for incorrect or lower ranking), the conditional probability $\Pr(x = 0)$ is:

$$P_{ni0} = \frac{1}{1 + \exp\left(\theta_n - \delta_i\right)},$$

in which $\theta_n$ is the location of person $n$ on the latent variable, and $\delta_i$ is the location of item $i$ on the latent variable, and for dichotomous case, $\delta_i$ is defined as the difficulty of item $i$.

Therefore, the odds ratio between correct response (x = 1) and incorrect response (x = 0) can be written as:

$$\frac{P_{ni1}}{P_{ni0}} = \exp\left(\theta_n - \delta_i\right),$$

and the log-odds of operating characteristic function can be written as:

$$\ln\frac{P_{ni1}}{P_{ni0}} = \theta_n - \delta_i.$$

As an extension of the Rasch model (RM), the linear logistic Rasch model (LLRM) assumes that the item difficulty $\delta_i$ can be described as an additive function of basic parameters; therefore, the log-odds of the operating characteristic function can be extended to (De Boeck & Wilson, 2016):

$$\ln\frac{P_{ni1}}{P_{ni0}} = \theta_n - \sum_{k=1}^{k} \delta_i X_{ik},$$

in which $\sum_{k=1}^{k} \delta_i X_{ik}$ are parameters for additional facets that could influence response probability from the item level. The extension is motivated by the idea that item difficulty can be conceived as a function of certain cognitive operations involved in the solution process, each of which has a difficulty parameter (Fischer, 1995).

### Purpose

The purpose of this study is to examine the use of explanatory item response models to evaluate surveys. Specifically, we describe the

use of linear logistic Rasch models (LLRM) to explore the psychometric quality of a survey designed to create a scale to measure food insecurity. The influences of two item properties are examined in detail using several LLRM models. The main research questions are as follows:

1. Can we use item properties to examine survey data using the EIRM framework?
2. What do item properties contribute to our understanding of household food insecurity as measured by the HFSSM?

## Methodology

### Participants

This study uses a random sample of 500 households who participated in the HFSSM embedded in the Current Population Survey Food Security Supplement (CPS-FSS) from 2012 to 2014. The original survey included polytomous items, but they were dichotomized by USDA for their policy efforts related to food insecurity. Nord (2012) evaluated the consequences of dichotomizing the responses, and he suggested that dichotomization did not result in any significant impact on the U.S. Food Security Scale. It needs to be noted that only households with incomes below 185 percent of the Federal poverty threshold and households above the range who indicated some difficulty in meeting their food needs were administrated the survey items included in the HFSSM.

### Procedure

First, a Rasch analysis was conducted with the *Facets* computer program (Linacre, 2019). The Facets analyses provide the typical results reported in Rasch analyses. Next, LLRMs were fit using the *eirm* package in R (Bulut et al., 2021). EIRM analysis provides estimates of the LLRM, including model-data fit analyses for various extended Rasch models. Participant responses regarding their perspectives about whether they feel secure or insecure about their food conditions define the dependent variable, while the two item covariates are item referent (household items, adult items, or child items), and item content (worried, ate less, cut meal size, hungry, not eat for the whole day).

**Table 1**

*Classification of Items on Food Insecurity Scale (11 Items)*

| Household | Referent | Content |
|---|---|---|
| 1. Worried food would run out | H | Worry |
| 2. Food bought would not last | H | Worry |
| 3. Could not afford to eat balanced meal | H | Worry |
| **Adult items** | | |
| 4. Respondent ate less than should have | A | Ate less |
| 5. Adult(s) cut size or skipped meals | A | Cut meal size |
| 6. Respondent hungry but did not eat | A | Hungry |
| 7. Adult(s) not eat for the whole day | A | Not eat for the whole day |
| **Child items** | | |
| 8. Child(ren) not eating enough | C | Ate less |
| 9. Cut size of child(ren)s meals | C | Cut meal size |
| 10. Child(ren) hungry | C | Hungry |
| 11. Child(ren) not eat for the whole day | C | Not eat for the whole day |

*Note*. H = Household, A = Adult, and C = Child. Also, items 1, 2, 3, and 8 had polytomous responses in the original form of the survey. These items were re-scored to be dichotomous by the USDA for their policy purposes.

## Survey

The instrument used in this study is based on the Household Food Security Survey Module (HFSSM). The full survey consists of 18 items with ten questions (Items 1–10) for all households and eight (Items 11–18) for households with children. Sample items like "(I/we) couldn't afford to eat balanced meals" and "(My/Our child was/The children were) not eating enough because (I/we) just couldn't afford enough food." Eleven items were selected from HFSSM (three household items, four adult items, and four child items) for the purposes of this study. The household, adult, and child items were chosen to be parallel in terms of content related to the level of worry, behaviors (ate less, cut meal size, not eat for the whole day), and feelings of hunger. Table 1 shows the item classifications by referent and content for the 11 items. More information regarding the combined HFSSM can be found on this website: https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-u-s/measurement/.

## Results

The analyses were conducted with both the *Facets* program and R (*eirm* package). Table 1 shows the classification of the 11 items of the HFSSM by referent and content. The referent grouping of the items indicates who is being referred to in the item (either household, adult, or child), and the content grouping of the items shows the alignment among items based on the following: worry, ate less, cut meal size, hungry, and not eat for the whole day.

Next, a summary statistics table was generated based on the Facets program, and these statistics are shown in Table 2. Overall, 66.07% variance was explained by the Rasch measures, and this indicates that the data fit the Rasch model well (Linacre, 2006). The reliability of separation for items is .99, while the reliability of person separation is .80. The reliability of person separation can be interpreted in a similar way to the traditional coefficient alpha, and the two reliability of separation indices indicate good reliability for both person and item measures (Smith, 2001). Additionally, the mean (*M*) and the standard deviation (*SD*) of person measures are −2.45 and 2.65 with logits as units. The mean (*M*) and the standard deviation (*SD*) of item measures are 0 and 3.58. The standard deviation statistics are relatively high, and this suggests that the item measures and person measures are spread out along the latent construct.

**Table 2**

*Summary Statistics for Rasch Model Based on Facets*

| Measures | Person | Item |
|---|---|---|
| *M* | −2.45 | .00 |
| *SD* | 2.65 | 3.56 |
| *N* | 500 | 11 |
| **Infit MSE** | | |
| *M* | .97 | .94 |
| *SD* | .77 | .17 |
| **Outfit MSE** | | |
| *M* | .70 | .84 |
| *SD* | 1.30 | .57 |
| Reliability of separation | .80 | .99 |
| $\chi^2$ statistic | 2213.1* | 2496.9* |
| Degrees of freedom | 499 | 10 |
| Variance explained by Rasch measures | 66.07% | |

*Note*. MSE is the mean square error.
* $p < .05$.

The HFSSM was analyzed using the dichotomous Rasch model in both Facets program and R. Table 3 summarizes the calibration of items using Rasch analysis in Facets. The item location (δ) values vary, and Table 3 shows the item fit information (both Infit MSE and Outfit MSE). The Infit MSE and Outfit MSE values are close to the expected mean value of 1.00 overall, except for the Outfit MSE of Item 2, Item 3, and Item 9. According to the guidelines for interpreting MSE statistics shown in Table 4, these three items have Outfit

## Table 3

*Rasch Calibration of Items*

| Item | Calibrations | | | Item Fit | | |
|---|---|---|---|---|---|---|
| | Location (δ) | SE | Infit MSQ | Category | Outfit MSQ | Category |
| 1 | −5.66 | 0.16 | 0.94 | A | 0.65 | A |
| 2 | −4.33 | 0.14 | 0.89 | A | 1.73 | C |
| 3 | −3.06 | 0.13 | 1.25 | A | 1.69 | C |
| 4 | −1.56 | 0.13 | 0.87 | A | 0.66 | A |
| 5 | −0.73 | 0.14 | 0.85 | A | 0.72 | A |
| 6 | 0.35 | 0.16 | 0.73 | A | 0.41 | B |
| 7 | 2.41 | 0.23 | 1.08 | A | 0.60 | A |
| 8 | 0.53 | 0.16 | 1.17 | A | 0.81 | A |
| 9 | 2.46 | 0.24 | 0.97 | A | 1.60 | C |
| 10 | 3.21 | 0.29 | 0.79 | A | 0.43 | B |
| 11 | 6.38 | 0.76 | 0.77 | A | 0.02 | B |

## Table 4

*Interpretive Framework for Model-Data Fit Statistics*

| Mean square error (MSE) | Interpretation | Category |
|---|---|---|
| 0.50 < MSE < 1.50 | Productive for measurement | A |
| MSE < 0.50 | Less productive for measurement, but not distorting of measures | B |
| 1.50 < MSE < 2.00 | Unproductive for measurement, but not distorting of measures | C |
| MSE > 2.00 | Unproductive for measurement, distorting of measures | D |

*Note.* The framework is adapted from Wright and Linacre (1994) and Engelhard and Wind (2017).

## Table 5

*Summary of Model-Data Fit*

| Models | AIC | BIC | logL | Deviance | Residual DF | Model |
|---|---|---|---|---|---|---|
| Baseline model 1 | 6515.3 | 6528.5 | −3255.6 | 6511.3 | 5498 | response ~ (1\|ID) |
| Baseline model 2 | 3932.8 | 3952.6 | −1963.4 | 3926.8 | 5497 | response ~ (1\|item) + (1\|ID) |
| Rasch model | 3864.6 | 3944.0 | −1920.3 | 3840.6 | 5488 | response ~ −1 + item + (1\|ID) |
| LLRM (referent) | 3919.3 | 3952.4 | −1954.7 | 3909.3 | 5495 | response ~ −1 + referent + (1\|item) + (1\|ID) |
| LLRM (content) | 3920.6 | 3966.9 | −1953.3 | 3906.6 | 5493 | response ~ −1 + content + (1\|item) + (1\|ID) |
| LLRM (Combined model) | 3903.1 | 3956.0 | −1943.6 | 3887.1 | 5492 | response ~ −1 + referent + content + (1\|item) + (1\|ID) |

MSE values belonging to category C. Category C suggests that these items are unproductive for measurement but not distorting of measures (Wright & Linacre, 1994).

The linear logistic Rasch models (LLRM) were estimated using the *eirm* package. Several models were estimated, and Table 5 presents model types, model-selected information criteria, model fit indices, and corresponding residual degrees of freedom. The last column is the R code for each model. As indicated in the model summary table, this study includes two baseline models (with one having person as a random effect and the other having both person and item as a random effect). The baseline model 2 can also be considered a Rasch model, and we found that baseline model 2 was helpful for interpreting our data. Baseline model 2 differs from the standard Rasch model in that it doesn't include the fixed item effect.

Model fit indices like Akaike's information criteria (AIC; Akaike, 1973), Bayesian information criteria (BIC; Schwarz, 1978), log-likelihood values, Deviance, and corresponding Residual DF values were included for all models. AIC can be calculated by using $-2\log L + 2q$, BIC can be calculated by using $-2\log L + qln(N)$, in which q is the number of parameters and N is the sample size. Log-likelihood values can be calculated by $\log L$, and Deviance (McCullagh, 1989/2019) can be calculated by using $-2\log lik(Fitted\ model)$. For this study, we used the *anova* function in R to compare generalized linear models and generate those model fit indices (Fox & Weisberg, 2011).

In terms of the interpretation of these information criteria, a smaller number suggests a better model fit. Comparing baseline model 1(with person as a random effect) and baseline model 2 (with both item and person as random effects), baseline model 2 has a better model fit. As expected, the Rasch model has the best model fit overall. Besides, comparing three linear logistic Rasch models, which have both item and person as random effects, LLRM (combined model) is relatively better than LLRM (referent) and LLRM (content), with a better fit index. The difference between LLRM (referent) and LLRM (content) is undecided at this stage of analysis.

A statistical comparison was conducted using Likelihood ratio tests in R, and the results are shown in Table 6. The first model comparison was between baseline model 1 and baseline model 2, and the result is consistent with comparison using information criteria: baseline model 2 has a better model fit. In the second series of model comparisons, when comparing the baseline model 2 to the Rasch model, the outcome aligns with the findings from the comparison using information criteria, indicating that the Rasch model exhibits superior model fit.

The third set of model comparisons is between baseline model 2 and three LLRMs since they all have person and item as random effects. The results stated that three LLRMs with item properties as predictors fit the data better than no item properties, with $X^2_{diff}(2) = 17.427$, $p < .001$, $X^2_{diff}(4) = 20.195$, $p < .001$, and $X^2_{diff}(5) = 39.659$, $p < .001$, respectively.

The fourth set of model comparisons compares LLRM (referent), LLRM (content), and LLRM (Combined model). For the comparison between the referent model versus the combined model, the combined model offers a better fit, and it's statistically significant, with $X^2_{diff}(3) = 22.232$, $p < .001$. For the comparison between the content model versus the combined model, the combined model offers a better fit, and it's statistically significant as well, with $X^2_{diff}(1) = 22.232$, $p < .001$. To compare the difference between these two models, we brought up the fit index for LLRM developed by Embretson (1997), $\triangle^2$. This goodness of fit index utilizes a comparison of a null model (denoted as $L_0$), a saturated model (denoted as $L_s$), and a model to be evaluated (denoted as $L_m$). The fit index, $\triangle^2$, can be calculated as follows:

$$\triangle^2 = (\ln L_0 - \ln L_m)/(\ln L_0 - \ln L_s)$$

## Table 6

*Model Comparisons*

| Model comparisons | | $\chi^2_{diff}$ | $df_{dif}$ | $p$ |
|---|---|---|---|---|
| Baseline model 2 (Npar = 3) | Baseline model 1 (Npar = 2) | 2584.5 | 1 | <.001* |
| Rasch model 1 (Npar = 12) | Baseline model 1 (Npar = 2) | 2670.7 | 10 | <.001* |
| LLRM- referent (Npar = 5) | Baseline model 2 (Npar = 3) | 17.427 | 2 | <.001* |
| LLRM- content (Npar = 7) | Baseline model 2 (Npar = 3) | 20.195 | 4 | <.001* |
| LLRM- combined model (Npar = 8) | Baseline model 2 (Npar = 3) | 39.659 | 5 | <.001* |
| LLRM- combined model (Npar = 8) | LLRM- referent (Npar = 5) | 22.232 | 3 | <.001* |
| LLRM- combined model (Npar = 8) | LLRM- content (Npar = 7) | 19.463 | 1 | <.001* |

*Note.* Npar = number of parameters, $\chi^2_{diff}$ = difference of Chi-square values, and $df_{dif}$ = degree of freedom difference. The combined model includes both referent and content as item predictors.
* $p < .05$.

## Table 7

*Summary of LLRMs by Item Characteristics (Referent and Content)*

| Item covariates | Facets | EIRM LLRM (referent) | | Facets | EIRM LLRM (referent) | |
|---|---|---|---|---|---|---|
| | $\eta_A$ | $\eta_B$ | $SE$ | $\eta_A$ | $\eta_B$ | $SE$ |
| **Referent** | | | | | | |
| Household (H) | −4.350 | −1.536* | 0.697 | | | |
| Adult (A) | 0.115 | 2.168* | 0.607 | | | |
| Child (C) | 3.145 | 4.469* | 0.620 | | | |
| **Content** | | | | | | |
| Worry | | | | −4.350 | −1.536* | 0.619 |
| Ate less | | | | −0.515 | 1.642* | 0.756 |
| Cut meal size | | | | 0.865 | 2.790* | 0.760 |
| Hungry | | | | 1.780 | 3.537* | 0.764 |
| Not eat for the whole day | | | | 4.390 | 5.323* | 0.792 |

*Note.* $\eta$ is the average item location value for each item characteristic. $\eta_A$ is in the logit metric based on the Facets model, while $\eta_B$ is in logit metric based on the EIRM. These two metrics order the items in the same way (r = .99).
* $p < .05$.

The denominator indicates the total amount of information that could be modeled since the saturated model is compared with the null model. The numerator compares the information in the estimated model with the null model, which provides information regarding the amount of information that the estimated model could explain. Based on the calculation of $\triangle^2$, LLRM referent model has a $\triangle^2$ value of 0.44, which indicates 44% of the information that can be modeled using item referent as a predictor in explaining item responses. On the other hand, the LLRM content model has a $\triangle^2$ value of 0.51, which indicates 51% of the information we can model using item content as a predictor in explaining item responses.

The difference of the $\triangle^2$ between the two models is .07, which suggests there is a 7% difference in terms of how much the model can be used to predict item responses. This difference between these two LLRMs is statistically significant based on the chi-square difference test result. Additionally, the combined model with both predictors is significantly better than the reduced model, with $X^2_{diff}(1) = 19.464$, $p < .001$.

To further explain why referent and content can be selected as predictors for item response, a traditional two-way ANOVA was conducted to examine the influence of referent and content on item location. The different logit metrics generated by *eirm* and Facets programs generate different average item locations. We examined the item location values generated by *eirm* and Facets, and these items have an almost perfect correlation, as expected. Based on this high correlation, we have used the item locations interchangeably (after re-scaling). The results indicated that both referent and content significantly influence item locations, with $F(2, 5) = 53.120$, $p < .001$, and $F(3, 5) = 8.262$, $p = .022$. Therefore, a series of logistic linear Rasch model analyses were conducted in the next step.

Table 7 displays the average location of each condition generated by Facets and the *eirm* package in R, together with standard error values. The item calibration values (Table 1) are in agreement with the average item location levels for the three referent conditions. The results suggest that items classified as child-related have the highest item location value, indicating that they are less likely to give a positive response than items classified with other contents. Conversely, household items have the lowest item location values, indicating that participants are more likely to give a positive response than items classified with other contents. The adult items have a medium level of item location. In Table 7, the mean location values for the items based on content, arranged from lowest to highest, are as follows: "worrying," "ate less," "cut meal size," "hungry," and "not eat for the whole day."

A Wright map was developed using the Facets program (Figure 1) and it corresponds to the information provided in Table 2. Both item measures and person measures (households) were plotted based on the order of location and the person ability. Corresponding to the context of this study, households who reported more food insecurity were listed at the top of the line, and households who reported less food insecurity were listed close to the bottom of the line. As shown in Figure 1, more households feel more food secure based on the "Household/Count" column. For items, items that participants were least likely to give a positive response to were listed at the top (participants are least likely to give a positive response to item "Children not eat for the whole day"), and items that participants are most likely to give a positive response were listed at the bottom ("worried food would run out").

The effect of referent and content on item location is presented graphically in Figure 2. As demonstrated in Figure 2, when the item referent is a child, then the item is harder to endorse, and when the item referent is a household item, then the items are easier to endorse. Furthermore, the results suggest that when the item content is "not eat for the whole day," then the items are the hardest to endorse, and when the items have a content of "worry,"

## Figure 1

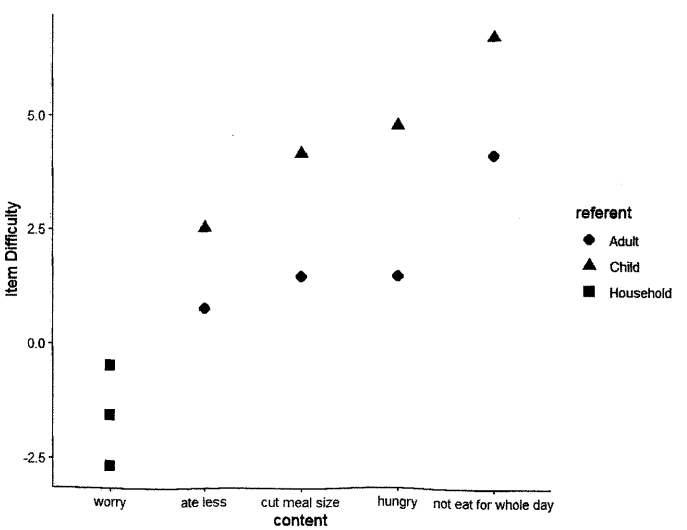
*Wright Map*

| Measr | +Household | -Items |
|---|---|---|
| 7 | 3 | |
| | | Child(ren) not eat for whole day |
| 6 | | |
| | 5 | |
| 5 | | |
| | | |
| 4 | | |
| | 4 | Child(ren) hungry |
| 3 | | |
| | 16 | Cut size of child(ren)s meals |
| 2 | | Adult(s) not eat for whole day |
| | 18 | |
| 1 | | |
| | 45 | C Child(ren) not eating enough |
| | | Adult(s) Respondent hungry but did not eat |
| 0 | 47 | |
| | | Adult(s) cut or skipped meals frequency follow up question |
| -1 | | |
| | 62 | |
| | | Adult(s) Respondent ate less than should have |
| -2 | | |
| | 84 | |
| -3 | | Household Could not afford to eat balanced meals |
| | 99 | |
| -4 | | |
| | | Household Food bought would not last |
| -5 | | |
| | 88 | |
| | | Household Worried food would run out |
| -6 | 29 | |
| Measr | Count | -Items |

*Note*. The Wright map was generated in Facets. The first column listed as "Measr" is measure, which is the scale in logits. The second column listed "Household" on the top and "Count" at the bottom is the distribution of households. It indicates the household location along the (how many households are at person's location on the latent variable). The third column listed as "Items" is the item locations on the scale.

## Figure 2

*Effect of Item Characteristics (Referent and Content) on Item Locations*



then the item is the easiest to endorse. The effects of classification and category together on item location are shown as the hardest items would be items with a content of "not eat for the whole day" and in the referent group "Child."

## Discussion

The study provides important insights into the psychometric properties of a survey designed to measure food insecurity, specifically the Household Food Security Survey Module (HFSSM). The use of multiple models, including the Rasch model and linear logistic Rasch models, enabled the researchers to explore the impact of item characteristics on response patterns.

The study found that item characteristics, such as referent (household, adult, and child) and content (worry, ate less, cut meal size, hungry, and not eat for the whole day), are important predictors that influence the likelihood of obtaining item responses in the HFSSM. This highlights the importance of carefully considering the wording and reference groups used in survey items to ensure that they are appropriate and do not introduce potential bias into the responses.

Furthermore, the study showed that the Rasch model yielded the best fit with the data. The Rasch models are a powerful tool for analyzing the psychometric properties of survey items, and their use in this study suggests that it may be a useful approach for analyzing other surveys as well. Incorporating item properties such as Referent and Content in the linear logistic Rasch models also increased the amount of variance accounted for by the model, indicating that these factors are important in predicting item responses. The finding that utilizing both item referent and content as predictors is superior to using either one of them individually to predict item responses highlights the importance of considering multiple item characteristics when developing and analyzing survey items.

Overall, the study provides important insights into the psychometric properties of the HFSSM and highlights the importance of considering item characteristics in the development and analysis of survey items.

### Implication and Future Research

Our study conducted a series of analyses using extended Rasch models to examine the psychometric quality of survey data and provided validity evidence (construct validity) from the substantive aspect of it (Messick, 1995). We conclude that the properties of items are useful for understanding item responses in surveys. We also demonstrated the usefulness of linear logistic Rasch models and discussed the issue of predicting item responses in surveys.

As an application of our analysis results, when adults were asked about their food conditions, parents and adults in the family prioritized feeding their children first before themselves when food was scarce. This finding has important implications for identifying households that are experiencing severe food insecurity and need food assistance. When households report that their children have gone without food for a whole day, this indicates that the household is unable to meet the basic food needs of all its members, including the most vulnerable ones. In this situation, it is critical to provide food assistance to the household to ensure that all members, including children, have access to enough food to meet their basic needs.

Future research should address several limitations of this study. First of all, the items being used in this study are relatively small (N = 11), although typical of many surveys. Future studies can use a larger number of items to see whether the results hold. Regarding the parallel content structure of the chosen items in this study, it is noteworthy that all household items have the content of "Worry" with no representation in the other content conditions. This may constrain the comparisons we can draw based on the content conditions. Future studies would benefit from including a greater number of items in each condition and predictor to facilitate more meaningful comparisons.

Additionally, future researchers should replicate this work with other instruments to extend the generalization of the results. This study provides a road map for conducting a series of analyses using explanatory Rasch models to examine the psychometric quality of survey data.

## Author's Note

An earlier version of the manuscript was presented at the International Meeting of Objective Measurement Workshop in Chicago (April 2023).

## References

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing: National Council on measurement in education*. American Educational Research Association.

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

Bulut, O., Gorgun, G., & Yildirim-Erbasli, S. N. (2021). Estimating explanatory extensions of dichotomous and polytomous Rasch models: The eirm package in R. *Psych*, *3*(3), 308–321. https://doi.org/10.3390/psych3030023

Campbell, C. C. (1991). Food insecurity: A nutritional outcome or a predictor variable? *The Journal of Nutrition*, *121*(3), 408–415. https://doi.org/10.1093/jn/121.3.408

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9

De Boeck, P., & Wilson, M. R. (2016). Explanatory response models. In W. J. van der Linden (Ed.), *Handbook of item response theory, Volume one: Models* (pp. 593–608). Chapman and Hall/CRC.

Derrickson, J. P., Fisher, A. G., & Anderson, J. E. (2000). The core food security module scale measure is valid and reliable when used with Asians and Pacific Islanders. *The Journal of Nutrition*, *130*(11), 2666–2674. https://doi.org/10.1093/jn/130.11.2666

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). John Wiley & Sons.

Economic Research Service. (n.d.). *Measurement*. United States Department of Agriculture. Retrieved 2018, from https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/measurement.aspx

Embretson, S. E. (1997). Multicomponent response models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 305–321). https://doi.org/10.1007/978-1-4757-2691-6_18

Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Engelhard, G., Jr., Rabbitt, M. P., & Engelhard, E. M. (2018). Using household fit indices to examine the psychometric quality of food insecurity measures. *Educational and Psychological Measurement*, *78*(6), 1089–1107. https://doi.org/10.1177/0013164417728317

Engelhard, G., & Wind, S. (2018). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge.

Fischer, G. H. (1995). The linear logistic test model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 131–

155). Springer. https://doi.org/10.1007/978-1-4612-4230-7_8

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). SAGE Publications.

Hambleton, R. K., Robin, F., & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. E. A. Tinsley & S. D. Brown, *Handbook of applied multivariate statistics and mathematical modeling* (pp. 553–581). Academic Press. https://doi.org/10.1016/B978-012691360-6/50020-3

Linacre, J. M. (2006). Data variance explained by Rasch measures. *Rasch Measurement Transactions*, *20*(1), 1045.

Linacre, J. M. (2019). *Facets computer program for many-facet Rasch measurement, version 3.81.2*. Winsteps.com.

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*(5), 877–892. https://doi.org/10.1017/S1368980014001050

McCullagh, P. (2019). *Generalized linear models* (2nd ed.). Routledge. (Original work published 1989)

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Nord, M. (2012). *Assessing potential technical enhancements to the U.S. household food security measures*. U.S. Department of Agriculture, Economic Research Service.

Poinstingl, H. (2009). The linear logistic test model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science*, *51*(2), 123–134.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. (Expanded ed.). University of Chicago Press. (Original work published 1960)

Schneider, B., González-Romá, V., Ostroff, C., & West, M. A. (2017). Organizational climate and culture: Reflections on the history of the constructs in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 468–482. https://doi.org/10.1037/apl0000090

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://www.jstor.org/stable/2958889

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, *2*(3), 281–311.

Somekh, B., & Lewin, C. (Eds.). (2005). *Research methods in the social sciences*. Sage.

Tanaka, V. T., Engelhard, G., Jr., & Rabbitt, M. P. (2019). Examining differential item functioning in the Household Food Insecurity Scale: Does participation in SNAP affect measurement invariance. *Journal of Applied Measurement*, *20*(1), 100–111.

Tanaka, V. T., Engelhard, G., Jr, & Rabbitt, M. P. (2020). Using a bifactor model to measure food insecurity in households with children. *Journal of Family and Economic Issues*, *41*(3), 492–504. https://doi.org/10.1007/s10834-020-09686-9

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.

Wilson, M., De Boeck, P., & Carstensen,

C. H. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 91–120). Hogrefe & Huber Publishers.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.