

Using the Rasch Modeling for psychometrics examination of food security and acculturation surveys



Jill F. Kilanowski, PhD, APRN, CPNP
Associate Professor

Alpha Zeta & Mu Chi



Acknowledgements

- **Dr. Li Lin, Cincinnati Children's Hospital Medical Center**
- **Dr. Hugo Melgar-Quinonez, The Ohio State University**
- **Funding:**
 - **P30NRO10676 NIH, National Institute for Nursing Research, Center of Excellence to Build the Science of Self-Management: A Systems Approach. Case Western Reserve University, Frances Payne Bolton School of Nursing, Dietary Intake and Nutrition Education (DINE)**
 - **KL2RR024990 Case/Cleveland Clinic Multidisciplinary Clinical Research Training Program, National Center for Research Resources & the National Center for Advancing Translational Sciences, National Institutes of Health**



The author has no conflicts of interest to report

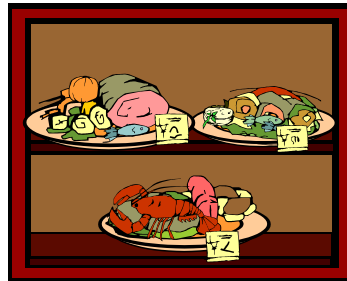
Dietary Intake & Nutrition Education

- **Three years, three phases**
 - **Phase One - Descriptive and correlation study**
 - **Phase Two - Pilot testing of technology and development of health education content**
 - **Phase Three - Quasi-experimental**
- **Total sample for analysis
N = 112**



Food Security

- ***Food security*** is defined as the ability of individuals to obtain sufficient food for an active healthy life in socially acceptable ways (Bickel & Nord, 2000)



How was food security measured?

- **The Household Food Security Survey Modules (HFSSM) is able to distinguish various levels of food insecurity**
 - **A 5-question short form is used in non-interview data collections, with alternative language formats decreasing test burden without sacrificing reliability**
 - **The short form for households with children is sensitive and specific to determine overall food security (85.9% and 99.5%)**
 - **Completed surveys are given scale scores and classified into food security levels based on standard values in total number of affirmatives: 0 - 1 = High or marginal food secure, 2 - 4 = Low food security, and 5 - 6 = very low food security**
 - **In this study for years 2008 and 2010, Cronbach's alpha for the HFSSM was .808 and .818.**

Acculturation

Acculturation is defined as the psychological and social changes occurring when individuals from different cultures come into continuous contact with each other. (Cabassa, 2003)



What acculturation survey was used?

■ The Short Acculturation Scale for Hispanics (SASH)

- Reliable method to identify Hispanics with low or high acculturation
- The SASH 12-item version responses are averaged, and an average score of 2.99 is used to differentiate the less acculturated from the more acculturated
- The Cronbach's alpha for the 12 items in this scale ranges from .78 to .92
- The instrument showed significant correlations with length of stay in the US, and US nativity generational level (.70 and .65, respectively)
- In this study for years 2008 and 2010, Cronbach's alpha for the SASH was .893 and .902



The value of reliability

- **It allows tests to be interpretable**
- **Gives confidence in determining relationships between variables**
- **If a decision made by the test is important, final, irreversible, unconfirmable, concerned with individuals, and/or has lasting consequences = high reliability is desired.**
- **If a decision is of minor importance: made early, is reversible, confirmable by other data, concerns groups, and/or has temporary effects = low reliability is acceptable**

The value of validity

- Does the test measure what we think it measures?
- Content, criterion-related (predictive and concurrent), construct
 - *Construct- what factors account for variance in the test performance*
- What portion of the total variance is accounted for by each of the constructs
- Validates theory behind the test



Item analysis

- Gives us information on how good or how poor test items are within the measuring instrument

Item analysis

- Evaluates each item separately to determine if the item is good or poor.
- Two measures used

□ **Item Difficulty** = $\frac{\text{number of people answering item correctly}}{\text{total number of people taking test}}$

The larger the value = easier item

Best item have values of 0.5 - 0.7

□ **Index of endorsement** = $\frac{\text{number of people selecting answer}}{\text{total number of people taking test}}$

Best item have values of 0.5 - 0.7

Item discrimination index

- Tells the researcher how effectively the item was able to discriminate between high and low scores
 - A “good” item is where high scores get the item correct and low scores answer incorrectly
 - Best suited for cognitive tests (*have a right and wrong answer*)

Item discrimination index

- Discrimination index for item $i = \frac{P_T - P_B}{\text{\# of people in Top group}}$

P_T = # of people in top group that answered item correctly

P_B = # of people in bottom group that got the item correct

If value is negative (reverse discrimination) something is wrong with the item

Good items have positive values, higher values means greater the discrimination

Item response theory (IRT)

- Scales the difficulty or endorsement of the items
- Uses item-characteristic curve with latent-trait theory –
 - *Assumes that test performance can be accounted for by the test takers' position on a hypothetical and unobservable characteristics (trait)*
- Basic measurement is probability
 - *Better items have a pattern where high scorers get the item correct & lower scores get the item incorrect*
 - *Steeper the curve going from low to high scores, the better discrimination power of that item*
 - *Negative discrimination have a negative slope and the items have problems that need to be examined further*

Item response theory

- One-parameter (1 PL) model of IRT assumes that guessing is a part of the ability
- All items that fit the model have equivalent discrimination (*items are assumed to vary only with respect to their difficulty*)
- Items are only described by a single parameter
- Rank of the item difficulty:
 - *Same for all respondents, independent of difficulty, equivalent in terms of discrimination*
- Rank of the person ability:
 - *Same for items independently of difficulty*

What is the Rasch model?

■ Part of item response theory

- Examines the *fit* of questionnaire items measuring identical underlying constructs along a *logit* continuum
 - *Fit = estimate item characteristics and then identify individuals whose responses to items do not adhere to those parameters*
 - *Logit = unit of measurement to report relative differences between candidate ability estimates and item difficulties; are an equal interval level of measurement. It puts candidate ability and item difficulty on the same measurement scale*
- Probability of a specified response is modeled as a function of person and item parameters
- Indicates whether a total score to characterize a person is justified

Rasch model...*continued*

- Assumes that items are all equal in discrimination (*weight equally on a factor*) and chance factors (*guessing*) do not influence the response
- Forms a basis for maximum likelihood estimation of the locations of objects or persons on a continuum, based on collections of categorical data
- The theory is that the probability of endorsing an individual item is decided by the difference between the item severity (*difficulty*) and a person's position (*ability*)
 - *If item severity is lower than the person's position, then the item has more chance to be endorsed*
 - *It is a model that represents the structure which data should exhibit in order to obtain measurements from the data (criterion for successful measurement)*



Rasch Model is often considered to be the 1 PL IRT model

- **Concern with the measurement of individuals, rather distributions among populations**
- **A given trait is quantitative and measurable, as operationalized in a particular experimental context**

Advantages of Rasch Model

- Persons and items can be mapped onto the same invariant scale
- Provides an estimation of parameters is more straightforward:
 - *one-to-one mapping of raw number-correct scores*
- Provides diagnostic information on how well items or questions on assessments work to measure an ability or trait
- Makes it possible to test that a particular challenge represent the infinite population of all possible challenges in that domain
- A model in the sense of an *ideal*, even when it is never actually observed in practice

Infit and outfit, overfit & underfit

- **Infit** is a weighted sum which gives more value to on-target observation
 - When the responses fit the model perfectly, the *resulting infit score is 1.0, with a recommended range of 0.8 to 1.2 and a wider acceptable range of 0.7 to 1.3*
 - **Item fit** is an index of whether items function logically and provide a continuum useful for all respondents
 - **Item misfit** may result from items that are too complex and confusing to the respondent, or are measuring a different construct
- **Outfit** is the conventional averaged sum of squared standardized residuals
- **Overfit** too little variation in the response pattern, perhaps indicating the presence of redundant items
- **Underfit** suggests unusual or inappropriate response patterns

Person & item separation

- **Person and item separation assess instrument spread across the trait continuum in standard error units**
 - *For an instrument to be useful, the items and persons should be able to be separated, so the separation should **exceed 1.0**, with higher values of separation representing greater spread of items and persons along a continuum*
 - *Lower values of separation indicate redundancy among the items or less variability of persons on the trait*
- **Each item should contain a different amount of the trait**
- **Person reliability is conceptually equivalent to Cronbach's alpha with different formulas**



Relative item severities & person positions

- Relative item severities and person positions estimates are calculated by Rasch model
- These scores check whether the tool was valid and all the items were performed as expected severity
 - *Whether the specific sample is well targeted*
- Useful in determining the ability of respondents to distinguish between items in the food security questionnaire
- We can assess differences between groups of less or more acculturated households to evaluate the differences in response patterns to the HFSSM

DIF contrast

- Rasch model generates the **Differential item functioning (DIF) contrast**, which allows comparisons across groups while holding the level of psychological disturbances constant
- *DIF contrast* , represents the difference in **relative severity** scores between the groups being compared
 - *Above 0.5 indicates groups answered differently*
- A substantial DIF contrast demonstrates that response probabilities are not fully explained by the latent trait
 - *Other variables are influencing the response*
 - *Comparisons between groups are problematic*
- The statistical significance of the DIF contrast is assessed using the *Welch t-test*
 - DIF contrast scores larger than 1.0 logit unit call for attention
 - *Probably showing a difference in response patterns among groups being compared*
 - *Some researchers consider scores under 2.0 not substantial*

Data Analysis

- Rasch model - examines the fitness and internal validity of household food security surveys in this specific population
- Responses to the food security items applied were fit into the a Rasch model for partial credit scoring, **Winsteps software** (Rasch Measurement 3.72, Winsteps, Chicago, IL, 2011)
- Fit statistics were reported as mean-square residuals, which have approximate chi-square and *t*-standardized distributions
- HFSSM item responses were coded as four dichotomous variables and one ordinal variable (*for question 3*) according to the tool recommendation
- HFSSM response pattern differences among acculturation levels were checked by differential item functioning (DIF) analysis



Results

- The data showed a balanced spread and cohesive order
- Item infit statistics showed no substantial deviations from expectations for all items
 - *Infit mean 0.97 with a standard score of -0.2*
 - *If responses fit the model perfectly infit score is 1.0 close to a perfect fit of 1 and 0*
- Item separate score was 5.79
 - *Suggesting we measured items on a continuum*
- Person reliability was 0.65 and separation score was 1.37
 - *Acceptable but not very high*
 - *Due to few items, small categories in the items*

Item Fit Statistics in Misfit Order comparing HFSSM & SASH using Rasch Model *(worst to best fitting)*

Entry No.	Total score	Measure	Rasch S.E.	Infit MNSQ	Infit ZSTD	Correlation
1	71	-2.85	.40	1.23	2.2	.73
4	26	1.92	.35	.95	1.3	.68
3	69	.92	.23	.96	-.4	.81
5	25	2.04	.35	.93	-.3	.69
2	65	-2.04	.34	.77	-.6	.81
Mean	51.2	0	.33	.97	.4	
S.D.	21.1	2.05	.06	.15	1.1	

Very Good range

Note. Entry No.: the sequence number of the item in the data

Total score: the total raw score of an item by the persons

Measure: the item severity estimate in logit

Rasch S.E.: the standard error of the estimate

Infit MNSQ: the information-weighted mean square statistic with expectation 1

Infit ZSTD: the t standardized information-weighted mean square statistic with expectation 0

Correlation: the point-measure correlation between the item with the unidimensionality structure

S.D.: Standard deviation

Differential Item Analysis for Acculturation Levels Using Median of 1.42 as a Cut-Point



Person Class	DIF Measure	DIF S.E.	Person Class	DIF Measure	DIF S.E.	DIF Contrast	Joint S.E.	t	Welch d. f.	p-value	Item No.
Low	-2.77	.69	High	-2.91	.49	.14	.85	.17	63	.8674	1
Low	-2.77	.69	High	-1.73	.41	-1.03	.80	-1.28	60	.2040	2
Low	.95	.31	High	.89	.33	.06	.45	.12	67	.9019	3
Low	2.12	.46	High	1.67	.51	.44	.69	.64	67	.5239	4
Low	2.12	.46	High	1.95	.54	.16	.71	.23	67	.8165	5

Note. Person Class: Low /High acculturation

DIF Measure: the difficulty of this item for this class

DIF S.E.: the standard error of the DIF MEASURE


DIF Contrast: the difference between the DIF Measures

Joint S.E.: the standard error of the DIF Contrast

Welch t: *t* statistics for DIF Contrast

p-value: the *p*-value of the Welch *t*-test

Differential Item Analysis for Acculturation Levels Using 2.99 as a Cut-Point



Person Class	DIF Measure	DIF S.E.	Person Class	DIF Measure	DIF S.E.	DIF Contrast	Joint S.E.	t	Welch d.f.	p-value	Item No.
Low	-2.75	.41	High	-3.82	1.97	1.06	2.01	.53	2	.6509	1
Low	-2.04	.36	High	-2.13	1.34	.10	1.38	.0	3	.9486	2
Low	.90	.23	High	1.26	1.75	-.36	1.76	-.20	2	.8577	3
Low	1.92	.35	High	1.92	2.75	0	2.77	0	2	1	4
Low	2.04	.35	High	2.04	2.90	0	2.92	0	2	1	5

Note:

Person Class: Low /High acculturation

DIF Measure: difficulty of this item for this class

DIF S.E.: standard error of the DIF MEASURE



Discussion

- Adequate sample size-above recommendations
- Limitations- small number of items in HFSSM made person estimates less reliable
- Imbalance between low and high acculturation levels
- Highest relative severity score for both acculturation groups in last two HFSSM questions
- Agreement with theoretical framework of food insecurity as a managed process



Implications for Research & Practice

- **The presence of low levels of food security in Latino MFW represents significant challenges for health care professionals who strive to improve the diets of children and families**
- **This study demonstrated that the US Department of Agriculture Household Food Security Survey Module (HFSSM) performed well in a unique MFW population and that levels of acculturation did not impact its performance**
- **The HFSSM can be used in research and in practice with confidence in this vulnerable population**



Questions ?

Kilanowski, J.F. & Lin, Li (2012). Rasch analysis of US Household Food Security Survey Module in Latino migrant farmworkers, *Journal of Hunger & Environmental Nutrition*.7 (2-3), 178-191.