HUNGRY CITIES PARTNERSHIP

Discussion Papers

OCTOBER 2019

NO. 37
VALIDATION OF THE HCP SURVEY TOOL FOR MEASURING URBAN FOOD INSECURITY: AN ITEM RESPONSE THEORY APPROACH

VONAI CHARAMBA,[1] NDEYAPO NICKANOR[2] AND LAWRENCE N. KAZEMBE[3]

SERIES EDITOR: JONATHAN CRUSH[4]

[1] University of Namibia, Windhoek, Namibia (2018 QES-AS doctoral scholar at the Balsillie School of International Affairs) vcharamba@unam.na
[2] University of Namibia, nnickanor@unam.na
[3] University of Namibia, lkazembe@unam.na
[4] Balsillie School of International Affairs, Waterloo, jcrush@balsillieschool.ca

Abstract

There is some controversy on the applicability of the summand-based Household Food Security Assessment Score (HFIAS) and Household Dietary Diversity Score (HDDS) as measures of food insecurity in urban areas in the Global South. These measures were primarily designed for measurement in rural communities where food insecurity itself was first identified and is still predominantly conceptualized. The objective of the research reported in this paper is to assess the internal and external validity of the HFIAS and the HDDS in determining food insecurity levels in an urban context, using HCP data for urban households in Windhoek, Namibia. Validation was performed on data collected through a city-wide survey of 890 randomly-selected households in Windhoek. Internal validation of sum-based measures is dependent on them meeting the assumptions of item response theory (IRT) Rasch models. For external validation, the HFIAS and the HDDS scores were correlated with variables including household income, the Lived Poverty Index (LPI), access to water, access to medical facilities, months of adequate food supply, and dwelling type. The Rasch IRT model and the area under the curve (AUC) of the receiver operating curve (ROC) were applied to validate the HDDS with binary response items, while the Partial Credit Model validated the HFIAS with polytomous response items. Principal Components Analysis (PCA) and Mokken analysis were applied to determine their dimensionality and monotonicity. The results from the ROC show that the HDDS is effective in separating Windhoek urban households according to their dietary quality and quantity. The results from the Mokken analysis, PCA, and PCM led to the conclusion that the HFIAS is monotonically non-decreasing, unidimensional, and no differential-item-functioning; all item-fit statistics were "very good" justifying the use of the HFIAS as a measure of food insecurity in Windhoek. The correlations indicated that low-income households have poor diets and higher levels of food insecurity, as expected. The paper concludes that the HDDS and the HFIAS indices are internally and externally valid in measuring the food insecurity status of Windhoek urban households.

Keywords

food insecurity measurements, internal validation, external validity, urban households, Windhoek

Suggested Citation

Charamba, V., Nickanor, N. and Kazembe, L. (2019). *Validation of the HCP Survey Tool for Measuring Urban Food Insecurity: An Item Response Theory Approach* HCP Discussion Paper No. 37, Waterloo and Cape Town.

This is the 37th discussion paper in a series published by the Hungry Cities Partnership (HCP), an international research project examining food security and inclusive growth in cities in the Global South. The five-year collaborative project aims to understand how cities in the Global South will manage the food security challenges arising from rapid urbanization and the transformation of urban food systems. The Partnership is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) and the International Development Research Centre (IDRC) through the International Partnerships for Sustainable Societies (IPaSS) Program. Vonai Chiramba received additional support from the Queen Elizabeth Diamond Jubilee Advanced Scholars Program (QE-AS).

HUNGRY CITIES PARTNERSHIP

QUEEN ELIZABETH Scholars

Introduction

Food security exists when all people always have physical and economic access to sufficient, safe, and nutritious food that meets their dietary needs and preferences for an active healthy life (FAO 1996). Based on this definition, food security involves the intersection of four dimensions: availability, access, utilization and stability (FAO 2008, Haysom 2017). It is important that tools and scales for food security measurement produce reliable results, particularly if they provide the evidence base for policy interventions (Deitchler et al 2010, Jones et al 2013). Furthermore, the ability to measure hunger and food insecurity accurately is crucial in monitoring progress towards the attainment of SDG 2, which calls for ending hunger, achieving food security, improving nutrition, and promoting sustainable agriculture.

The challenge of urban food security in the Global South remains under-researched as food insecurity is still seen as a primarily rural problem (Crush and Frayne 2010, Crush and Riley 2019). Haysom and Tawodzera (2018) argue that increasing evidence of food insecurity at the urban scale means that the development of appropriate metrics requires urgent attention. According to Deitchler et al (2010), the measurement instrument used to collect food insecurity data across different urban sites requires a common instrument and standard metrics for assessment, monitoring, and evaluation.

A number of different methods for measuring the various dimensions of food security have been suggested. The FAO advocates a suite of methods including (a) estimating calories available per capita at the national level; (b) household income and expenditure surveys; (c) individual dietary intake; (d) anthropometric measures; and (d) experience-based food insecurity measurement scales. The USDA FANTA Project developed four widely-used experiential measures: (i) the Household Food Insecurity Access Scale (HFIAS); (ii) the Household Dietary Diversity Score (HDDS); (iii) the Household Food Insecurity Access Prevalence scale (HFIAP); and (iv) the Months of Adequate Household Food Provisioning (MAHFP) measure (Coates et al 2007). These measures were developed primarily for use in rural contexts. Haysom and Tawodzera (2018) have questioned their applicability as measures of food insecurity in cities of the Global South. In its efforts to develop a standardized instrument for its baseline household food security surveys, the Hungry Cities Partnership (HCP) has primarily relied on these self-reporting access indicators developed and refined by FANTA (Wilde 2011).

The HFIAS is based on the idea that the experience of food insecurity causes predictable reactions and responses at the household level that can be quantified through a survey and summary score for each household (Coates et al 2007, Frongillo 1999, Jones et al 2013). The HFIAP categorizes households into four levels of food insecurity: food secure, and mild, moderately, and severely food insecure using an algorithm based on responses to the HFIAS items. Households are categorized as increasingly food insecure as they respond affirmatively to more severe conditions and/or experience those conditions more frequently. The HDDS, defined as the number of unique food groups consumed within the household in a given time period, is seen as a proxy for both the quality and quantity of food consumption (Hodinott and Yohannes 2002, Jones et al 2013, Ruel 2003). According to Bilinsky and Swindale (2010), the MAHFP is a food security stability measure based on the total number of months in the previous year in which the households had adequate food provision. The measurement scales of the HFIAS and HDDS are based on summations of the household item category responses (Tables 1 and 2).

TABLE 1: HFIAS Items and Response Codes with Four Week Recall Period

No.	Items	Response codes
1	Worry that the household would not have enough food	Never, Rarely, Sometimes, Often
2	Not able to eat the kinds of foods preferred	Never, Rarely, Sometimes, Often
3	Eat a limited variety of foods	Never, Rarely, Sometimes, Often
4	Eat some foods that you really did not want to eat	Never, Rarely, Sometimes, Often
5	Eat a smaller meal than you felt you needed	Never, Rarely, Sometimes, Often
6	Eat fewer meals in a day	Never, Rarely, Sometimes, Often
7	No food to eat of any kind in your household	Never, Rarely, Sometimes, Often
8	Go to sleep at night hungry	Never, Rarely, Sometimes, Often
9	Go a whole day and night without eating	Never, Rarely, Sometimes, Often
Source: Deitchler et al (2010: 5)		

TABLE 2: HDDS Items and Response Codes with 24 Hour Recall Period

No.	Item	Response codes
A	Any [insert local foods], bread, rice noodles, biscuits, or any other foods made from millet, sorghum, maize, wheat, rice [or any other local cereal]	0=No, 1=Yes
B	Any potatoes, yams, manioc, cassava or any other foods made from roots or tubers	0=No, 1=Yes
C	Any vegetables	0=No, 1=Yes
D	Any fruits	0=No, 1=Yes
E	Any beef, pork, lamb, goat, rabbit, wild game, chicken, duck, or other birds, liver, kidney, heart, or other organ meats	0=No, 1=Yes
F	Any eggs	0=No, 1=Yes
G	Any fresh or dried fish or shellfish	0=No, 1=Yes
H	Any foods made from beans, peas, lentils or nuts	0=No, 1=Yes
I	Any cheese, milk, yogurt or other milk products	0=No, 1=Yes
J	Any foods made with oil, fat or butter	0=No, 1=Yes
K	Any sugar or honey	0=No, 1=Yes
L	Any other foods, such as condiments, coffee, tea	0=No, 1=Yes
Source: Swindale and Bilinsky (2006: 4)		

Validating Food Insecurity Metrics

Item response theory (IRT) models are reliant on the notion that the probability of endorsing a response category is modelled as a linear function of the overall household food insecurity implied by the item (Na at al 2015). This probability is shown on an item characteristic curve (ICC), which measures the difficulty of affirming an item response. IRT models, specifically Rasch models, provide a theoretical base and set of statistical tools to assess the suitability of a set of survey items for scale construction. IRT models create a scale from the items and compare performance of the scale in various populations and survey contexts (Nord 2014). They also formalize the concept of severity, ordering items by providing standard statistical methods to estimate the severity of each item.

Whenever a statistical model is utilized for data analysis, it is important to ascertain that the assumptions underlying the model have not been violated before making any inferences from the results. The assumptions of Rasch models include the following:

- *Unidimensionality,* which suggests that the correlation among the items can be explained by a single factor (such as food insecurity) and that all items from the same instrument either test or subscale measure that common trait.

- *Local independence of items and persons,* meaning that an individual's response to an item is not influenced by his or her response to other items in the test.

- *Monotonicity,* which implies that the probability of endorsing response categories is non-decreasing; that is, households with increasing levels of food insecurity have higher chances of affirming more severe item categories than households with less severe food insecurity. This allows the researcher to order households on the latent continuum with respect to the sum of the scores of the items belonging to the same scale. The ordering of questions on the HFIAS tool (as well as the computation of the four associated HFIAP categories) is based on the monotonicity assumption.

- *Differential item functioning (or DIF)* which implies that the score on any given item is the same for respondents in different subgroups within the population. If the response behaviours for subgroups within the population are not the same, then estimates of the parameters are biased.

- *Measurement invariance.* Invariant item ordering assumes that items measure the same level of food insecurity severity regardless of the food insecurity level of the households; that is, an item has the same difficulty level for both food secure and food insecure households. The probability of a household affirming an item is then determined by its food insecurity severity level. This assumption enables the researcher to rank items according to difficulty levels or prevalence, thereby allowing hierarchical ordering of the scale. For example, if a person answers six out of ten answers correctly, it is assumed the six items answered correctly were the first six easiest questions in the ordering (Stochl et al 2012).

Mokken scaling techniques are another useful tool for validating of unidimensional tests or questionnaires that comprise binary or polytomous items. They are suitable when the intention is to score an underlying latent trait by simple addition of item responses. The Mokken's stochastic cumulative scaling model can assist in the determination of the dimensionality of the measure and enables consideration of the measure. It can be applied as a secondary approach to scrutinize the appropriateness or performance of well-established IRT functions such as the Rasch family of models (Stochl et al 2012). The Mokken approach entails two probability models: (a) the monotone homogeneity model (MHM) which assumes unidimensionality, monotonicity, and local item independence; and (b) the double monotonicity model (DMM) which, in addition to the MHM assumptions, assumes non-intersection of items. If all the DMM assumptions are met, then the questionnaire/measure has invariant item ordering.

Validation is the process of determining whether a method is suitable for providing useful analytical measurement for a given purpose and context (Frongillo 1999, Leroy et al 2015) by ensuring that method construction is well grounded in an understanding of the phenomenon. Consistent, precise, reliable, and accurate method construction within the specified assumptions, performance standards, and accuracy is attributable to a well-grounded understanding (Nord 2014). If all of these criteria are fulfilled, then the method is valid for that purpose and context (Koch 1987).

Internal validity is a way of gauging the strength of the research design and methods used, and involves ensuring that the underlying assumptions have been met. External validity evaluates the suitability of a research or measurement tool in real life situations. External validity implies that the data collected using the measurement instruments are measurable against a gold standard of the construct being measured by the tool. If a measure is not coherent with other standard measures then its external validity is low. If the assumptions are not met, both the ordinal raw scores based on the number of items affirmed by households in answering questions, and the interval latent traits derived from IRT, are uncertain.

Several studies have been conducted to validate experience-based measures of food insecurity, such as the HFIAS and HDDS. These include HFIAS empirical validation studies on Burkina Faso (Frongillo and Nanama 2006), Ethiopia (Gebreyesus et al 2015), Tanzania (Knueppel et al 2009), Iran (Mohammed et al 2012), and a multi-country study on Mozambique, Malawi, West Bank/Gaza Strip, Kenya, Zimbabwe, and South Africa (Deitchler et al 2010). Most of these studies deploy statistical methods based on the Rasch measurement model with varying results. Empirical validation of the HDDS is less common, although Vellema et al (2016) applied the Rasch model to data collected in Colombia and Ecuador. Other validation studies include Abuelhaj (2007) and IFPRI (2006). The HCP has done no validation analysis to date of its use of the FANTA metrics for measuring food insecurity among urban households in the Global South. This paper therefore sets out to validate the nine-item HFIAS and the 12-item HDDS food group questions using household survey data from Windhoek, Namibia. Tool validation using Rasch models should provide evidence that the HFIAS and the HDDS are sufficient proxies of food insecurity (Deitchler et al 2010).

Methodology

This paper uses household data collected in Windhoek in 2017 using the HCP Household Food Security Baseline Survey. The survey tool collected a wide range of demographic, economic, food consumption, and food sourcing data at the household level. It also includes questions on household experience of food insecurity and dietary consumption. Respondent households were selected in the 10 constituencies of Windhoek using a two-stage sampling design procedure. Firstly, primary sampling units (PSUs) from a master frame developed and demarcated for the 2011 Population and Housing Census were randomly selected within the 10 constituencies of Windhoek. A total of 35 PSUs were selected covering the whole of Windhoek. Twenty-five households were systematically selected in each PSU, making a total sample size of 875 households. The sampled PSUs and households were located on maps, which were used to target households for interviews.

For this study, both the Rasch models and Mokken scale were applied to validate the internal and external construct validity of the HDDS and the HFIAS. The Rasch model and the 2-Parameter Logistic Model were applied to the HDDS binary (yes/no) response questions, while the Partial Credit Model (PCM) was used for the HFIAS items where the responses have four category options. The use of the PCM for validating the HFIAS scale follows Owino et al (2014) who, after validating the HFIAS with the Rasch Model, recommended further research through application of polytomous response IRT.

Sensitivity and specificity values were computed to assess the HDDS scale's ability to separate food secure households from food insecure households. The area under the receiver operating curve (ROC) was determined to assess the overall diagnostic accuracy of the measure. Tripepi et al (2009) argue that an area under the curve (AUC) between 0.7 and 0.8 is acceptable, 0.8 to 0.9 is excellent, 0.9 to 1 is outstanding. Further, factor analysis using principal components analysis was employed to group test items that measure the same food security dimension by looking at the polychloric correlation matrix to detect the number of factors (dimensions) measured by the HFIAS and the HDDS. The dimension was determined by the number of resultant factor components, the variances explained by the components as well as the number of eigen values greater than unit.

Cronbach's alpha was used to assess the construct validity of the data collection tool, that is, to ensure that the items measure the same latent trait, in this case food insecurity. A Cronbach's alpha value of at least 0.7 guarantees unidimensionality and consistency of the measurement scale. The reliability rho correlation gives an estimate of the reliability of the total score (Sitjima and Moleenar 1995). The monotonicity assumption was checked using Mokken scale analysis. This is based on the assumption that the items are arranged hierarchically by degree of difficulty, where a respondent who answered a difficult question correctly is assumed to answer an easy question correctly. Violations of this assumption are called Guttman errors. The Loevinger's coefficient H assesses the Guttman errors for possible violations of monotonicity and unidimensionality, while H_i looks at the homogeneity of the individual items and assesses whether they are coherent enough to be included in the scale. An H<0.3 indicates the test items are not unidimensional. An H greater than 0.3 but less than 0.4 indicates a weak scale (unidimensional but not strong in a scaling sense); between 0.4 and 0.5 indicates median strength; and greater than 0.5 implies unidimensionality and sufficient ordering of items (Ligvoet et al 2010). All H_i values in a unidimensional scale should be larger than 0.3 (Stochl et al 2012).

Chi-square-based item infit and outfit statistics measuring the difference between the expected and observed performance were computed to check if the test items fit the models. Infit and outfit statistics are normally expressed as the mean of the summed squared standardized residuals (MNSQ). Infit and outfit statistic values of 1 suggest a perfect fit, which rarely occurs. However, infit and outfit values between 0.8 and 1.2 are recommended (Na et al 2015) and between 0.7 and 1.3 are acceptable (Na et al 2015, Nord 2014). Bond and Fox (2001) and Linacre (2006) consider ranges between 0.6-1.4 and 0.5-1.5 as reasonable for assessing rating scales. The HFIAS items that played a significant role in disaggregating Windhoek households according to their food insecurity status (that is, separating food insecure from food secure households) were determined based on the item characteristic curves (ICC) and the item discriminant parameter (slope). The ability of a question in a test/survey to separate respondents according to the attribute being measured is one of the important attributes of the survey questions; otherwise the survey may include some redundant questions that do not contribute much to the measurement intended to be captured.

For external validity, the HFIAS and the HDDS scores were correlated against proxy-independent variables such as household income categories, the Lived Poverty Index (LPI), access to water, access to medical facilities, months of adequate food supply, and dwelling type using Spearman's rho correlation coefficient and graphical presentations. Nickanor et al (2017) conducted a primary data analysis on the Windhoek HFIAP and HDDS with the lived poverty index and income levels and found that households in informal settlements, households with low incomes and households with higher levels of lived poverty indices had higher food insecurity levels and poor diets.

Results

Columns 1-4 of Table 3 give the percentage of respondents responding affirmatively to each of the HFIAS items. The general response pattern was that the majority sometimes experienced the condition followed by those who never experienced the condition. Fewer said they often experience the condition and the lowest number said they rarely experience the condition. The response behaviour is further supported by the item-step difficulty, where it was difficult to move from "never" to "rarely", easier to move from "rarely" to "sometimes", and fairly difficult to move from "sometimes" to "often". The response behaviour is unexpected according to the desired monotonicity of response categories, suggesting bimodality in the data. The ordering of items according to their difficulty level (column b) is as expected; i.e. the items at the beginning of the test have lower difficulty levels when compared to the items at the end of the test. In addition, all the discrimination parameters are close to each other, suggesting equal discrimination parameters of items (column a), implying that all the questions in the scale are influential in determining the food insecurity levels and, hence, there are no redundant questions in the survey.

From the results in Table 3, the infit statistics for all HFIAS items are within the range 0.8-1.2; a range considered "very good" by Linacre (2006) and Nord

(2014). This indicates that there is no wide discrepancy between the observed deviations of responses and the deviations of expected responses, leading to the conclusion that the PCM model fits the data well. Also, the item oufit statistics are within the 0.5-1.5 acceptable range (Linacre 2006). In addition, none of the questions appears to be redundant, which suggests that there is statistical evidence in the Windhoek data that the HFIAS adequately explains the food security status of households.

Table 4 gives the proportion of respondents responding to the two HDDS item categories. Cereals were consumed by most of the respondents, followed by meat then food with sugar. According to the difficulty parameter, "cereals" was the easiest/least severe items consumed by both food secure and food insecure households while pulses were rarely consumed. However, although "pulses/legumes" were consumed by very few, it was not necessarily by those who are highly food secure – hence its lower discrimination ability. Furthermore, although the outfit statistic for the food group "fish and sea food" (1.64) is outside the 0.5-1.5 range considered acceptable for measurement purposes by Linacre (2006), its infit statistic is still in the 0.7-1.3 scale recommended by Na (2015). This implies that all the HDDS items are suitable for the Rasch measurement models. With the exception of "fish and sea food" and "legumes/pulses", all other items in the HDDS have discrimination parameters greater than 0.5. They are therefore productive in explaining the latent construct being measured (food insecurity).

TABLE 3: Item Statistics for the HFIAS Scale

Responses (%) (N=861)											
Item	Never	Rarely	Sometimes	Often	Outfit	Infit	b	Catgr.1	Catgr.2	Catgr.3	a
Worry about food	24.7	10.8	43.9	20.2	1.06	1.06	-0.71	-0.22	-0.71	1.20	1.71
Unable to eat preferred food	28.6	10.5	43.3	17.4	0.86	0.92	-0.56	-0.11	-0.46	1.25	2.12
Eat just a few kinds of food	27.1	10.7	36.4	25.6	0.90	1.00	-0.62	-0.19	-0.47	0.89	2.01
Eat unwanted food	27.2	11.5	29.0	22.0	0.78	0.83	-0.61	-0.27	-0.34	1.02	2.43
Eat a smaller meal	30 .1	13.6	37.7	18.4	0.78	0.85	-0.50	-0.21	-0.14	1.13	2.55
Eat fewer meals	28.5	4.3	36.2	21.0	0.77	0.87	-0.56	-0.28	-0.16	1.03	2.41
No food of any kind	39.5	16.3	30.3	13.8	1.08	1.10	-0.56	0.27	0.05	1.40	1.64
Go to sleep hungry	49.7	15.2	26.2	8.8	0.81	0.98	-0.19	0.55	0.33	1.68	1.95
Entire day and night without food	52.5	15.4	25.2	6.9	0.81	1.02	0.09	0.64	0.40	1.86	1.91

TABLE 4: Item Statistics for the HDDS Scale

Item	Consumed	Not consumed	Outfit MSQ	Infit MSQ	Discrimination	Difficulty
Cereals	94.2	5.8	1.14	0.96	0.80	-3.30
Roots and tubers	11.2	88.7	1.00	0.97	0.89	2.48
Vegetables	20.8	79.2	1.37	1.12	0.21	1.64
Fruits	5.8	94.2	0.53	0.87	1.93	3.33
Meat	49.7	50.3	1.07	1.07	0.91	0.01
Eggs	5.3	94.7	0.60	0.89	1.60	3.43
Fish and sea food	20.9	79.1	1.64	1.29	-0.41	1.63
Pulses/legumes	5.1	94.9	1.47	1.02	0.40	3.31
Milk and milk products	14.6	85.4	1.08	0.98	1.06	2.15
Oil/fats	29.5	70.5	0.96	0.98	0.88	1.07
Sugar/honey	33.4	66.6	0.81	0.85	8.92	0.82
Miscellaneous	26.4	73.6	0.76	0.84	7.35	1.26

The results of the analysis assessing the internal validity of the HFIAS and HDDS scales were as follows:

Unidimensionality

The Cronbach's alpha for the HFIAS was 0.942 with a Loevinger's scalability coefficient of 0.685. The two values were way above their acceptance cut-off points of 0.7 and 0.5 respectively, and hence strong, implying that the scale is homogeneous enough to be considered unidimensional. In addition, the PCA analysis resulted in the following eigen values: 6.60; 0.92; 0.53; 0.41; 0.41, 0.32, 0.26; 0.20, and 0.17. The first component explains 66% of the total variance, significantly higher than the rest and it is the only component with an eigen value greater than the unit. The high explanatory value of the first component suggests that the HFIAS is unidimensional; that is, it captures only one of the food insecurity dimensions, as expected. The factor analysis only generated one component; an indication that all the 10 HFIAS measure one latent trait (i.e. food (in)security). However, the Cronbach's alpha and the Loevinger's scalability coefficient for the HDDS were 0.594 and 0.208 respectively, and are therefore lower than the cut-offs and can be considered weak. The following eigen values were generated for the PCA for the HDDS: 2.84, 1.41, 1.34, 1.02, 1.01, 0.92, 0.88, 0.76, 0.71, 0.57, 0.38 and 0.19. In addition, five of the eigen values are greater than one (Figure 1) and hence the PCA generated five component dimensions for this measure.

Table 5 provides the five factor components for the HDDS as determined by the PCA. Factor 1 is strongly correlated to "sugar/honey" and "miscellaneous" and correlated to "milk and milk products" and "oil/fats", while Factor 2 is correlated to "fruits" and "pulses/legumes". Factor 3 is correlated to "meat" and "fish", and Factors 4 and 5 are correlated to "vegetables" and "cereals" respectively. However, both the scree plot and the rotated component loadings for the HDDS do not provide evidence to suggest that the HDDS is unidimensional as the principal components analysis resulted in five factors, implying that the 12 unique foods measured by the HDDS can be compressed into five dietary components by aggregating related foods. The first component is associated with milk and milk products, oil and fats, sugar and honey and miscellaneous foods, while the second component is associated with fruits and eggs. The third component captures meat, poultry and offal and is negatively correlated with fish and seafood, implying that these are lacking in that component. The fourth component captures the consumption of vegetables and the final component captures the consumption of cereals.

FIGURE 1: Scree Plots for the HFIAS Scale and the HDDS Scale

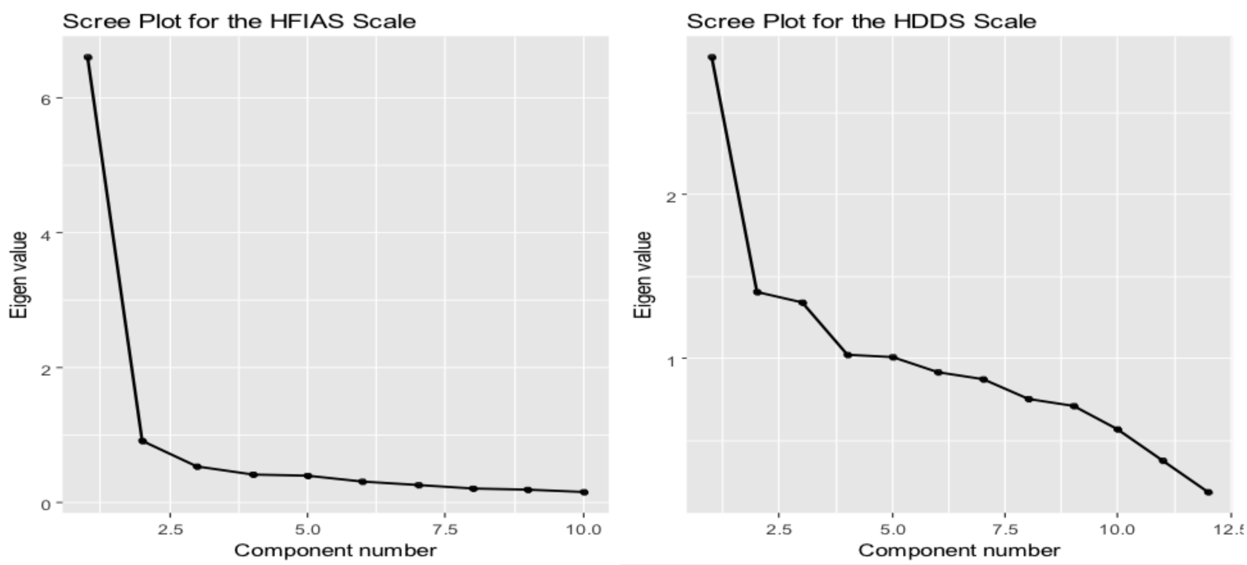


TABLE 5: Rotated Component Loadings for HDDS

Food group	Component 1	Component 2	Component 3	Component 4	Component 5
Cereals	0.175	0.074	-0.036	0.091	0.866
Roots and tubers	0.155	0.380	0.355	0.157	-0.303
Vegetables	0.035	0.102	0.043	0.947	0.05
Fruits	0.319	0.642	0.031	0.115	-0.216
Meat, poultry, offal	0.237	0.156	0.783	-0.207	0.134
Eggs	0.209	0.716	-0.004	0.105	0.016
Fish and sea food	0.059	0.103	-0.826	-0.184	0.12
Pulses/legumes	-0.188	0.713	0.03	-0.092	0.17
Milk and milk products	0.433	0.267	-0.059	0.099	-0.343
Oil/fats	0.540	0.037	-0.043	0.163	0.046
Sugar/honey	0.863	0.068	0.16	-0.072	0.074
Miscellaneous	0.846	0.108	0.159	-0.101	0.047

Monotonicity

The Loevinger coefficients for the HFIAS scale are all greater than 0.5, implying that it is homogeneous (unidimensional) and monotonically non-decreasing. Table 6 confirms that there are no significant violations of monotonicity for the HFIAS scale items as all H_i values are above 0.5, indicating that the items disaggregate households according to their food insecurity levels properly and are coherent enough to be included in the scale. These results are coherent with the results from the PCA where only one latent dimension was detected. The rho value of 0.949 indicates the reliability of the total HFIAS score. These results validate the ordering of responses on a latent continuum based on the sum of HFIAS scores.

The monotonicity assumption asserts that items in a measurement instrument are monotonically arranged if they are arranged according to their level of difficulty. For most of the items in the HDDS scale the monotonicity assumption has been violated except for "fish and sea food", "vegetables" and "meat", implying that the order in which they appear on the measurement instrument is in line with the frequency of consumption by the households. The H_i coefficients for "cereals", "sugar / honey", "fruits" and "eggs" are all above the cut point of 0.3 recommended by Stochl et al (2012), implying that these items are homogeneous and have been fairly positioned on the measurement scale according to their difficulty level. However, the remaining items have lower H_i values than the 0.3 cutoff (Table 7). A specificity value of 0.945 and a sensitivity value of 0.59 resulted from the analysis while the area under the receiver operating characteristic (ROC) curve was 0.89 indicating good performance of the HDDS in separating food insecure from food secure households. In addition, an accuracy value of 0.840 was recorded from the ROC analysis, implying that the estimates from the HDDS are close to their true (unknown) values. However, a Gini coefficient of 0.78 indicated inequalities within the food security status of the Windhoek population.

TABLE 6: Monotonicity Assessment for HFIAS Scale

Item	Monotonicity	Invariance	H_i
Worry about food	0	1	0.682
Unable to eat preferred food	0	0	0.704
Eat just a few kinds of food	0	0	0.696
Eat food they don't want	0	1	0.720
Eat a smaller meal	0	0	0.713
Eat fewer meals in a day	0	0	0.707
No food of any kind in the household	0	0	0.658
Go to sleep hungry	0	0	0.694
Go a whole day and night without food	0	0	0.689

TABLE 7: Monotonicity Assessment for HDDS Scale

Item	Monotonicity	Invariance	H_i
Cereals	0	0	0.474
Roots and tubers	1	1	0.231
Vegetables	4	4	0.048
Fruits	0	1	0.417
Meat, poultry, offal	3	2	0.185
Eggs	0	2	0.389
Fish and sea food	4	8	-0.123
Pulses/legumes	3	3	0.108
Milk and milk products	3	3	0.200
Oil/fats	0	6	0.222
Sugar/honey	1	4	0.372
Miscellaneous	0	5	0.359

Differential item functioning and invariant item ordering

The differential item functioning (DIF) analysis did not detect any violations for HFIAS items for different income quantiles, type of dwelling and family structure, implying that the scale discriminates equally among respondents in different subgroups within the population. In addition, no DIF was detected for HDDS for income quantiles and family structure. However, DIF was detected for dwelling type for item numbers 3, 5 and 10 measuring vegetables; meat, offal and poultry; and oil and fats consumption respectively. This suggests that households on the same food insecurity levels residing in different dwelling types may have different consumption patterns on the three food types. As is evident in the results in Table 6, there are substantially fewer violations of the monotonicity and invariance assumptions to compromise the double monotonicity assumption for the HFIAS. This implies that the items have been ordered according to the level of food insecurity they measure, and the measurement scale remains constant across different respondents and subgroups within the population. The HFIAS is thus a reliable measure of household food insecurity levels for the Windhoek urban households.

Figures 2a and 2b show the item characteristic curves (ICC) plots for the HFIAS and HDDS scales. The HFIAS has also been binned to binary responses by considering a never response as a negative response (0) and combining "rarely", "sometimes" and "often" to mean a positive (1) response for ease of interpretation. The method is one of the binary categorization methods employed by Deitchler et al (2010) in an HFIAS validation study conducted on households in Mozambique, Zimbabwe, Kenya, West Bank and Gaza Strip. The item plots on the extreme right correspond to higher levels of food insecurity while those on the extreme left show lower levels. All items in the HFIAS exhibit monotonicity as the probability of endorsing items increased with increasing food insecurity. Items 2 and 5 violate monotonic ordering according to the item ordering on the questionnaire, although the others do not. For the HDDS scale, the easiest outcome was item 1 "cereals" followed by item 5 "meat", and item 11 "sugar/honey". This outcome may be because the staples *mahangu* and maize meal, together with bread and fat cakes, are common in Namibian diets. Meat and tea are also very common foods for the entire Namibian populace.

External validity

Figure 3 shows the relationship between the HDDS and the four food security categories of the HFIAP. Food security increases with a rise in dietary diversity, and the severity of food insecurity decreases with an increase in dietary diversity. Figure 4 shows the performance of the HFIAP against type of dwelling and household income quintiles. Food insecurity (as measured by the HFIAP) decreases as income increases and is more intense in households residing in shacks in informal settlements than those in houses and flats. In addition, HFIAS was higher for households with poor access to clean water and medical facilities. The HDDS increased with income, decreased with lack of access to water and medical facilities and was poorest in households in informal settlements.

Correlation analysis was performed between the HDDS and the HFIAS scales to assess the external validity of both measures against one another (Table 8). As expected, there was a significant negative correlation (-0.525) between the scales, meaning that dietary diversity decreases with an increase in food insecurity levels. The internal validity of each scale was then tested against household income, months of adequate food supply, access to medical facilities, access to water, type of dwelling, and type of household. There was a statistically significant negative correlation (-0.449) between HFIAS and household income, meaning that food insecurity decreases as household income increases. On the other hand, there was a significant positive correlation (0.343) between the HDDS and income, indicating that household diets become more diverse as income increases. A negative correlation between HFIAS and months of adequate food supply (MAHFP) was also detected, indicating that

FIGURE 2A: Item Characteristic Curves for HFIAS items

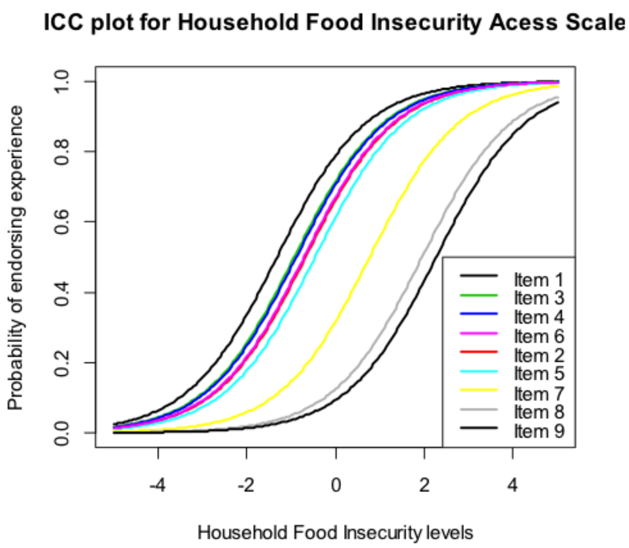
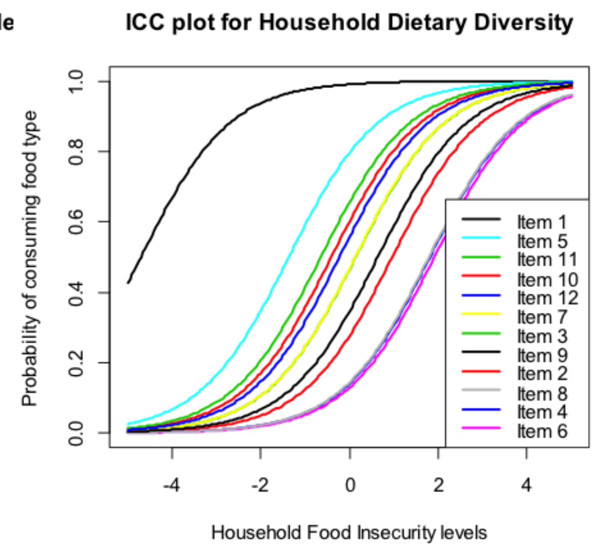


FIGURE 2b: Item Characteristic Curves for HDDS items



the more food insecure the household is, the greater the number of months they did not have enough food in the previous year. Furthermore, there was a significant positive correlation between HDDS and months of adequate food supply, suggesting that households with a more diverse diet also have more consistent access to food.

FIGURE 3: Proportion of Households in each HFIAP Category by HDDS

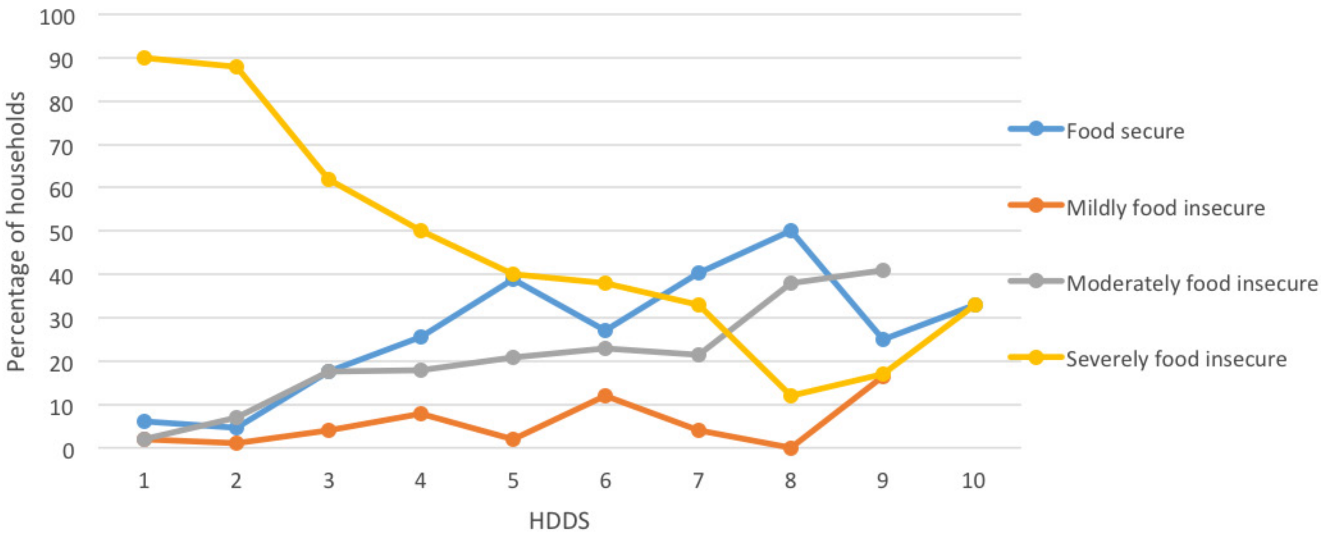


FIGURE 4: Association between Food Insecurity Levels, Income and Dwelling Type

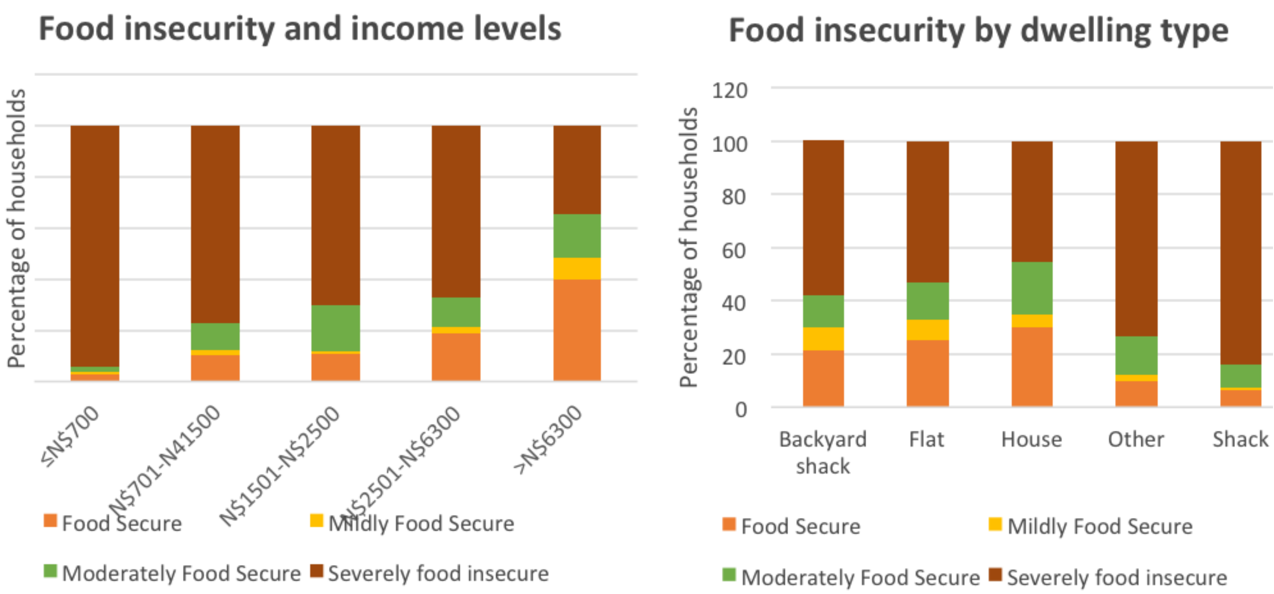


TABLE 8: Spearman's Rank Correlation Coefficients for Assessing External Validity

	HDDS	HDDS (IRT)	HFIAS	HFIAS (IRT)	Household income	MAHFP	LPI
HDDS	1.000	1.000	-0.558**	-0.540**	0.387**	0.351**	-0.584**
HDDS (IRT)		1.000	-0.550**	-0.525**	0.393**	0.343**	-0.579**
HFIAS			1.000	0.938**	-0.481**	-0.615**	0.661**
HFIAS (IRT)				1.000	-0.499**	-0.557**	0.610**
HH income					1.000	0.367**	-0.515**
MAHFP						1.000	-0.617**
LPI							1.000

**Correlation is significant at the 0.01 level (2-sided)

The LPI is an experiential measure based on a series of questions about how frequently households go without various basic necessities during the course of a year (Mattes et al 2016). Figure 5 shows that most food secure households have an LPI of 1 or less, while severely food insecure households have higher LPI values greater than 1. This means that the HFIAP categories are externally valid as they are able to match the food insecurity status of households with their levels of lived poverty. Similarly, the HDDS is higher for households with lower LPIs than households with higher LPIs. Thus, the higher the level of lived poverty, the lower the dietary diversity; again suggesting the HDDS is externally valid.

Table 8 also shows a strong positive correlation between the HFIAS and HDDS measures (a) computed from summing responses and (b) computed from IRT methods, despite the fact that the HDDS scale failed to meet some of the assumptions of the Rasch model. Although the magnitude of the measures differ, the two measures still rank the respondents on the same dietary diversity ranks.

The HFIAS scoring system in the HFIAS guide (Swindale et al 2007) is based on the assumption that the "worry" and "lack of resources" items indicate "mild food insecurity" status, while the "house empty", "sleep hungry" and "whole day without food" items purportedly measure "severe food insecurity" status. While the remaining items are aligned to "moderate food insecurity", the response items are theoretically supposed to be increasing monotonically in that order. In the Windhoek response behaviour (Figure 2), it is notable that the order is slightly different: item number 2 is more difficult, implying that it was only affirmed by households that are more food insecure when compared to most of the items presumed to measure higher levels of food insecurity. The differences in the distribution of food insecurity according to the HFIAS and its IRT counterpart may be because the IRT measures are based on actual probabilistic-based difficulty parameters and not the ordering on the questionnaire.

Figure 6 shows the dietary diversity categories using Swindale and Bilinsky's (2005) classification where households consuming 1-3 food groups have low diversity diets, those consuming 4-5 food groups have moderate diversity, and those consuming 6 or more food categories have high diversity. The majority of households have low dietary diversity according to both the HDDS computed by summing food groups consumed (66%) and by using the Rasch model (73%).

FIGURE 5: Relationship Between HFIAP Categories and Lived Poverty Index

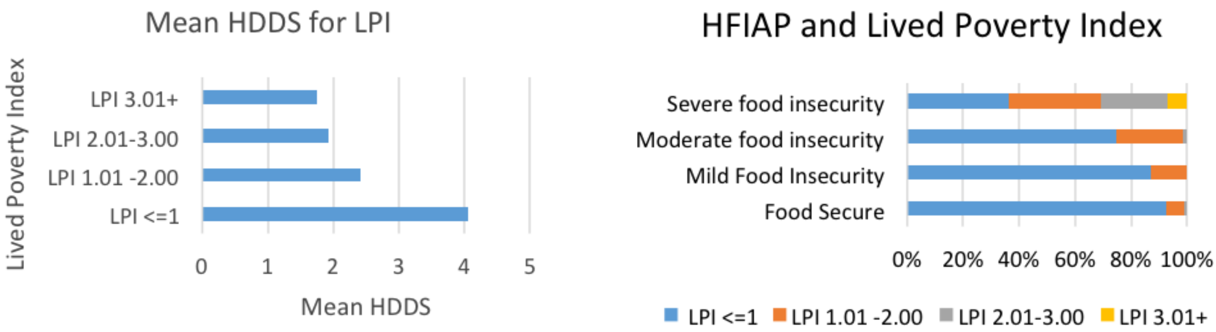
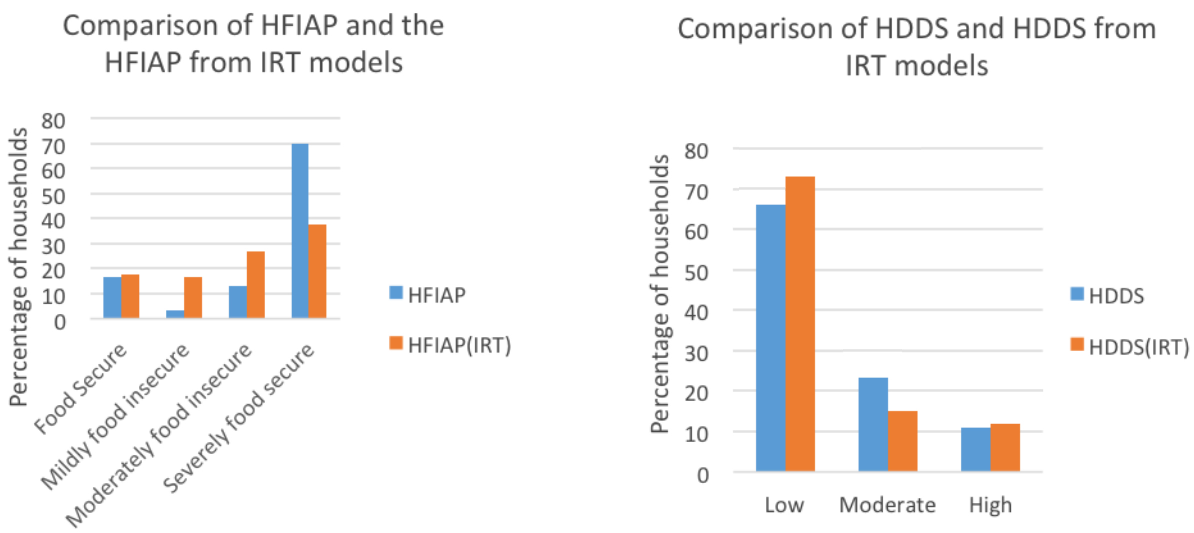


FIGURE 6: Comparison of HFIAP and HDDS Measures



Discussion

Internal Validity

According to the research results, all HDDS and HFIAS items demonstrate good model fit with acceptable infit statistics between 0.5-1.5 (a range deemed acceptable by Linacre (2006)). None of the items showed any sign of discrepancy, confirming that they are constructive in the measurement of food insecurity. In addition, PCA, Cronbach's alpha and the Loevinger's scalability coefficient provide evidence that the ordered responses of the 9 items of the HFIAS scale display a unidimensional, monotonically non-decreasing response pattern consistent with expectations. The accuracy of the measure has also been qualified because of the absence of DIF for all subgroups within the population on which the analysis was performed (type of dwelling, family structure, and household income). In addition, the scale has negligible invariance. These findings validate the criteria for calculation of cut-off points in FANTA's HFIAS Indicator Guide (Swindale et al 2007). The HFIAS scale exhibits unidimensionality and clearly increases with heightened food insecurity.

The HFIAS is based on the assumption of monotonicity; that is, households are more likely to answer "yes" to less severe items than to more severe items and that they are likely to answer "yes" to more severe food insecurity items than households with less severe food insecurity (Deitchler et al 2010). The computation of the HFIAP food insecurity category is based on this assumption, where food insecurity levels increase with the increase in the position of items and the classes are non-overlapping. It is expected that households that do not endorse lower category items will not endorse higher level categories, and hence the food insecurity category levels do not overlap and the summation of category responses has been justified (see Swindale et al 2006). The results from the Loevinger coefficients for testing for monotonicity (Table 6) indicate that the HFIAS is monotonically non-decreasing. The violation of the monotonocity of response categories was noted in other validation studies and researchers attributed it to problems with the format or wording of questions and vague response categories (Teh et al 2017, Abuelhaj 2007, Deitchler et al 2010).

Although the assumption of monotonicity was not violated for the HFIAS, some households gave negative responses for item 1, "In the past four weeks, did you worry that your household would not have enough food?", but then went on to affirm some higher order items such as those on dietary quality. This implies that (1) the question may not have been properly understood by some respondents, (2) some households may have had so much food in their households that they did not worry about their food supplies depleting even if their dietary quality may be very limited, making them eat the same foods repeatedly, and hence endorsing items measuring the food quality, or (3) anxiety and uncertainty about household food supply has become the norm and is no longer perceived as anxiety. The ordering from the HFIAS recorded as binary response (1,3,4,6,2,5,7,8,9) may imply that these items were difficult to differentiate in Namibian vernacular languages. Problems on differentiating meanings of similar items were noted in Mozambique where items 2, 3 and 4 were considered repetition, and in Zimbabwe and Malawi there were problems in distinguishing "fewer meals" and "smaller meals" reported in a validation study conducted by Deitchler et al (2010). Teh et al (2017) treated the disordering of items as potential problems in their wording and format.

Unlike the HFIAS, the HDDS does not weigh the food types in any particular order and hence its computation is not based on unidimensionality and monotonicity assumptions. The failure of the HDDS to meet the differential item functioning for questions 3, 5 and 10 across different dwelling types is probably due to the fact that diets in Windhoek are culturally diverse. As a result, the consumption of a foodstuff might fail to discriminate households according to their food insecurity levels if the IRT methods were employed to determine the dietary diversity levels of the households. However, the HDDS score does not assign weights to the food types and hence might not be compromised. In the current study, the ROC analysis for the HDDS exhibited an excellent ability to classify households (Tripepi et al 2009). In contrast, Abuelhaj (2007) found the scale had a poor ability to classify undernourished households with the area under the curve estimated to be 0.62 and concluded their results were analogous to a similar study by IFPRI (2006).

The HDDS has limitations on its internal validity as some items failed to meet the Rasch assumptions (although most have good discriminatory ability and item-fit statistics). The study findings concur with Vellema et al (2016) who conclude that the HDDS indicator in its current form is not internally valid. However, although some of the assumptions of the Rasch model have been compromised, the objective of measurement is to reflect the variety of food consumed by households (Jones et al 2003, Ruel 2003). Unlike the HFIAS which was developed as a latent trait measurement model, the responses to the HDDS are taken at face value and summed to indicate the diversity of sources of nutrition. It is not necessarily a latent trait measurement model as a number of different yet unobserved forces such as economic status, culture, and personal preferences may influence the diversity of food intake (Abuelhaj 2007). However, it is not very clear whether an HDDS of 5 in one population is a 5 in another population, as the measurement equivalence has not been established and is unlikely to be (Leroy at al 2015).

External validity

The HDDS and the HFIAS behaved as expected when correlated: an increase in dietary diversity resulted in reduced food insecurity as measured by the HFIAS. In addition, the HFIAS has a significant negative correlation while the HDDS has a significant positive correlation with the MAHFP. Gandure et al (2010) similarly found a significant inverse relationship between the HFIAS and HDDS (r=-0.425) in Zimbabwe and that households reporting food shortages in the previous 12 months (using the MAHFP) had worse HDDS and HFIAS scores than those that did not. The Windhoek analysis also found that both the HFIAS and HDDS measures are externally valid when tested against known determinants of food insecurity such as income, access to water (a proxy for sanitation), and medical facilities.

Performance of the IRT models

There was a very high correlation between the FANTA HFIAS and HDDS computed by (a) summing up individual responses and (b) their counterparts computed using the PCM and Rasch IRT model. The IRT method thus did not result in much change in food insecurity rankings when compared to the FANTA measures. This implies that the rank ordering of respondents by summing up individual responses did not differ much from the more complex IRT measures. The use of the summand (FANTA HFIAS and HDDS) scales justifies Deitchler et al's (2010) argument that complex approaches to the measurement of food insecurity are highly quantitative and at times highly computationally intensive, and hence may not be appropriate in the field during programme implementation, beneficiary targeting, and impact assessment.

With the description of the HFIAP categories, Coates et al (2007) imply that the HFIAS items are ranked by their severity on the questionnaire and households are to be classified according to their highest severity level. However, the IRT analysis has shown that the ordering of questions on the questionnaire does not tally with empirical severity based on the observed data. The IRT scales are based on the empirical difficulty levels from the responses of respondents. This is probably why the categories of the IRT-based scales and the HDDS scales differ.

The HDDS and HFIAS based on raw scores are only valid proxies of food insecurity if the assumptions of the Rasch model and PCM are met, and when then there are no missing responses in the data set. However, IRT responses can be adjusted to estimate food insecurity even when the assumptions have been violated and when some responses are missing. Rasch models enable the estimation of food insecurity levels even when there are non-responses or when there are different but partially overlapping items. In addition, the computation of the HFIAS guide is based on the increasing severity of food security following the order of the items. In this study, the second item and fifth item violated the ordering, making the IRT scales superior as they are reliant on empirical ordering of items. When rapid measures are required, it is desirable to use the HDDS and HFIAS scores as they are easier to compute. However, for more important decisions and policy formulation, IRT measures are more desirable. Further research on the use of IRT techniques to set cut-off points and evaluate the distribution of the households according to their food insecurity status is recommended.

Conclusion

The validation of the HFIAS and HDDS measures in Windhoek implies their ability to measure the extent or severity of food insecurity. In turn, these verified measurements make it possible to come up with more realistic, robust, and adequate measures of food insecurity useful for identifying food insecure households, assessing the severity of food insecurity for informed programme interventions, and formulating and implementing policy. The validity of the PCM statistical properties demonstrated in this paper justifies the computation of the HFIAS score and the HFIAP scales based on summands of category responses for the items. However, continual modification and validation to ensure that all questions are understood in the way intended by the measurement tool would ensure that food security is assessed using the best possible metrics.

References

1. Abuelhaj, T. (2007). "Methodological Concerns in the Measurement of Undernourishment, Dietary Diversity and Household Food Insecurity" At: www.ipcinfo.org
2. Bilinsky, P. and Swindale, A. (2010). *Months of Adequate Household Food Provisioning (MAHFP) for Measurement of Household Food Access: Indicator Guide* (Version 4) (Washington, DC: FANTA).
3. Bond, T. and Fox, C. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences* (Mahwah, NJ: Lawrence Erlbaum Associations, Inc.).
4. Coates, J., Swindale, A. and Bilinsky, P. (2007). *Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator Guide* (Version 3) (Washington, DC: FANTA).
5. Coates, J., Webb, P. and Houser, R. (2003). *Measuring Food Insecurity: Going Beyond Indicators of Income and Anthropometry* (Washington, DC: FANTA).
6. Crush, J. and Frayne, B. (2011). "Urban Food Insecurity and the New International Food Security Agenda" *Development Southern Africa* 28: 527-544.
7. Deitchler, M., Ballard, T., Swindale, A. and Coates, J. (2010). *Validation of a Measure of Household Hunger for Cross-Cultural Use* (Washington, DC: FANTA).
8. FAO (1996). *Rome Declaration on World Food Security and World Food Summit Plan of Action.* (Rome: FAO).
9. FAO (2008). "An Introduction to the Basic Concepts of Food Security" At: www.foodsec.org
10. Frongillo, A. (1999). "Validation of Measures of Food Insecurity and Hunger" *Journal of Nutrition* 129: 506S-509S.
11. Frongillo, A. and S. Nanama. (2006). "Development and Validation of an Experience based Tool to Directly Measure Household Food Insecurity Within and Across Seasons in Northern Burkina Faso" *Journal of Nutrition* 136: 1409S-1419S.
12. Gebreyesus, S., Lunde, T., Mariam, D. Woldehanna, T. and Lindtjørn, B. (2015). "Is the Adapted Household Food Insecurity Access Scale (HFIAS) Developed Internationally to Measure Food Insecurity Valid in Urban and Rural Households of Ethiopia?" *BMC Nutrition* 1: 2.
13. Haysom, G. and Tawodzera, G. (2018). "Measurement Drives Diagnosis and Response: Gaps in Transferring Food Security Assessment to the Urban Scale" *Food Policy* 74: 117-125.
14. Haysom, G. (2017). "Climate Change, Food and the City: Agency and Urban Scale Food System Networks" In E. Thomas-Hope (ed), *Climate Change and Food Security: Africa and the Caribbean* (London: Routledge), pp. 145-155.
15. Hoddinott, J. and Yohannes, Y. (2002). "Dietary Diversity as a Food Security Indicator" FCND Discussion Paper No. 136, International Food Policy Research Institute, Washington DC.
16. IFPRI (2006). "Review and Validation of Dietary Diversity, Food Frequency and Other Proxy Indicators of Household Food Security" International Food Policy Research Institute. Washington, D.C.
17. Jones, A., Ngure, F., Pelto, G. and Young, S. (2013). "What are We Assessing When We Measure Food Security? A Compendium and Review of Current Metrics" *Advances in Nutrition* 4: 481-506.
18. Knueppel, D., Demment, M. and Kaiser L. (2009). "Validation of the Household Food Insecurity Access Scale in Rural Tanzania" *Public Health Nutrition* 13: 360-367
19. Leroy, J., Ruel, M., Frongillo, A., Harris, J. and Ballard, T. (2015). "Measuring the Food Access Dimension of Food Security: A Critical Review and Mapping of Indicators" *Food and Nutrition Bulletin* 36: 167-195.
20. Linacre, J. (2006). *A User's Guide to WINSTEPS, MINISTEP Rasch-Model Computer Programs* (Chicago: John M. Linacre).
21. Mattes, R., Dulani, B. and Gyimah-Boadi, E. (2016). "Africa's Growth Dividend? Lived Poverty Drops Across Much of the Continent" At: https://afrobarometer.org/publications/pp29-africas-growth-dividend-lived-poverty-drops-across-the-continent
22. Mohammadi, F., Omidvar, N., Houshiar-Rad, A., Khoshfetrat, M., Abdollahi, M. and Mehrabi, Y. (2012). "Validity of an Adapted Household Food Insecurity Access Scale in Urban Households in Iran" *Public Health Nutrition* 15: 149-157.
23. Na, M., Gross, A. and West, K. (2015). "Validation of the Food Access Survey Tool to Assess Household Food Insecurity in Rural Bangladesh" *BMC Public Health* 15: 863.
24. NSA (2016). *Namibia Household Income and Expenditure Survey (NHIES): 2015/2016 Report* (Windhoek: Namibia Statistical Agency).
25. Nickanor, L., Kazembe, L., Crush, J. and Wagner, J. (2016). *The Supermarket Revolution and Food Security in Namibia* AFSUN Urban Food Security Series No. 26, Cape Town.
26. Nord, M. (2014). "Introduction to Item Response Theory Applied to Food Security Measurement: Basic Concepts, Parameters, and Statistics" At: http://www.fao.org/economic/ess/ess-fs/voices/en

27. Owino, A., Wesonga, R. and Nabugoomu, F. (2014). "Determining Food Insecurity: An Application of the Rasch Model with Household Survey Data in Uganda" *International Journal of Food Science* Article ID 121269.

28. Ruel, M. (2003). "Operationalizing Dietary Diversity: A Review of Measurement Issues and Research Priorities" *Journal of Nutrition* 133: 3911-3926.

29. Stochl, J., Jones, P., and Croudace, T. (2012). "Mokken Scale Analysis of Mental Health and Well-being Questionnaire Item Responses: A Non-parametric IRT Method in Empirical Research for Applied Health Researchers" *BMC Medical Research Methodology* 12: 74.

30. Swindle, A. and Bilinsky, P. (2005). *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide* (Washington, DC: FANTA).

31. Teh, L., Pirkle, C., Furgal, C., Fillion, M. and Lucas, M. (2017). "Psychometric Validation of the Household Food Insecurity Access Scale Among Inuit Pregnant Women from Northern Quebec" *PLoS ONE* 12(6): e0178708.

32. Vellema, W., Desiere, S. and Haese, M. (2016). "Verifying Validity of the Household Dietary Diversity Score: An Application of Rasch Modelling" *Food and Nutrition Bulletin* 37: 27-41.

33. Webb, P., Coates, J., Frongillo, A., Rogers, B., Swindale, A. and Bilinsky, P. (2006). "Measuring Household Food Insecurity: Why it's So Important and Yet So Difficult to Do" *Journal of Nutrition* 136: 1404S-1408S.

34. Wilde, P. (2011). "Food Security Policy Document in Developed Countries" In L. Lusk, J. Roosen and J. Shogren (eds), *Oxford Handbook of the Economics of Food Consumption and Policy* (Oxford: OUP).