

Comparison between Experience-based Food Insecurity scales

Confronto tra scale di insicurezza alimentare basate sull'esperienza

Federica Onori, Sara Viviani and Pierpaolo Brutti

Abstract In order to face food insecurity as a global phenomenon, it is essential to rely on measurement tools that guarantee comparability across countries. Although the official indicator adopted by the United Nations in the context of the Sustainable Development Goals (SDGs) and based on the Food Insecurity Experience Scale (FIES) already embeds cross-country comparability, other experience-based scales currently employ national thresholds. In this paper we address the issue of comparability by presenting two different studies. The first one between FIES and three national scales (ELCSA, EMSA and EBIA) included in national surveys in Guatemala, Ecuador, Mexico and Brazil. The second one between the adult and children versions of these national scales. Different methods from the equating practice of educational testing are explored: parametric, nonparametric, classical and based on the Item Response Theory (IRT).

Abstract *Al fine di affrontare il problema dell'insicurezza alimentare come un fenomeno globale, è essenziale poter contare su strumenti di misurazione che garantiscano comparabilità tra Paesi. Nonostante l'indicatore ufficialmente adottato dalle Nazioni Unite nel contesto dei Sustainable Development Goals e basato sulla scala FIES, formalmente assicuri questa possibilità, altre scale di insicurezza alimentare utilizzano soglie nazionali. Questo lavoro propone due studi di comparabilità. Il primo riguarda la scala FIES e le scale nazionali ELCSA, EMSA ed EBIA, mentre il secondo confronta le scale nazionali riferite a famiglie con e senza minori. Vengono implementati diversi metodi di confronto utilizzati nell'educational testing: parametrici, nonparametrici, classici e basati sull'Item Response Theory (IRT).*

Federica Onori
Sapienza University of Rome, e-mail: federica.onori@uniroma1.it

Sara Viviani
Food and Agriculture Organization of the United Nations, e-mail: sara.viviani@fao.org

Pierpaolo Brutti
Sapienza University of Rome, e-mail: pierpaolo.brutti@uniroma1.it

Key words: Food insecurity, Test equating, Item Response Theory, SDGs

1 Introduction

Food insecurity is formally defined as *the state of being without reliable access to a sufficient quantity of affordable and nutritious food* and food security is one of the target of the Sustainable Development Goals (SDGs) of the 2030 Agenda adopted by the United Nations. A number of indicators have been proposed to measure food insecurity and among all, *experience-based food insecurity scales* have proved to be valid and reliable tools to this aim [1]. These scales address the *access dimension* of food insecurity from the point of view of individual behaviours by directly asking people about the four aspects of psychological concern, food quality, food quantity and hunger. The first experience-based food insecurity scale was formulated in the United States where since 1995, the Household Food Security Survey Module (HFSSM) has been applied annually to monitor the phenomenon. Countries in Latin America, inspired by the HFSSM, developed their own national scales and, in this study we consider the *Brazilian Scale of Food Insecurity* (EBIA), the *Latin American and Caribbean Food Security Scale* (ELCSA) and the *Mexican Food Security Scale* (EMSA) [8], all included in national surveys for periodical monitoring. In order to provide a *global* measurement tool for food insecurity, in 2013 FAO launched the Voices of the Hungry project and introduced the Food Insecurity Experience Scale (FIES) [3]. Conceived as a global adaptation of the previous experience-based scales, the FIES was designed to produce formally comparable prevalences of food insecurity across countries and an index based on FIES was adopted as one of the official indicators for tracking progresses toward target 2.1 of the SDGs. Despite a common evolution, each national scale use specific thresholds to measure prevalences of food insecurity for *nominally* the same level of severity, and the problem of comparability arises. This paper aims at addressing this issue by proposing two comparability studies employing both classical and IRT-based methods from the educational testing field. A first study is proposed that compares FIES and the national scales ELCSA, EMSA and EBIA, while the second study computes, for each national scale, the corresponding raw scores between the adult and the children-referenced versions of the scale. In section 2 we present our data and the methods applied. In section 3 we focus on the main results and conclusions.

2 Materials and Methods

Equating is a statistical method that is used to adjust for differences in difficulty between tests' forms built to be similar in content and difficulty, so that scores can be used interchangeably [5]. The common evolution of EBIA, EMSA, ELCSA and FIES makes their Survey Modules similar to each other, allowing for statistical

equating procedure. Nonetheless, the way they "build" measures of food insecurity from observations differ under three main aspects: methodology, reference period and reporting unit. In fact, survey Modules used for ELCSA, EMSA and EBIA measure food insecurity at the *household* level and with a reference period of 3 months, while in this paper we consider for FIES a recalling period of 12 months referring to the *adult individual* (people age 15 or more). As far as for the methodology, national statistical offices using either ELCSA, EMSA or EBIA adopt a **deterministic** approach: a *raw score* is computed for each household by counting the number of items affirmed by that household. Prevalences of food insecurity at different levels of severity are then calculated as percentages of households scoring within a certain range expressed in terms of raw scores (Table 1). Conversely, FIES methodology is **probabilistic** in nature relying on the Item Response Theory (IRT) and, more specifically, on the Rasch model as the main tool for data validation and scale building [4]. Following this methodology a common metric called Global Standard is used as the reference metric to adjust model parameters estimates at each application and each respondent is assigned a distribution of his/her food insecurity along the latent trait used to compute percentages of the population whose severity is beyond global thresholds. Two indicators are then computed: the Prevalence of Food Insecurity at moderate or severe levels ($FI_{Mod+Sev}$, threshold -0.25 on the Global Standard) and the Prevalence of Food Insecurity at severe level (FI_{Sev} , threshold 1.81 on the Global Standard) [3]. Finally, FIES consists of one single scale based on 8 items referred to the adults, while EBIA, EMSA and ELCSA consist of two distinct scales, one for households without children and one for households with children, each one with specific thresholds for the different levels of food insecurity (Table 1).

Scale	Food insecurity Level	Households without children	Households with children
ELCSA	mild	1 to 3	1 to 5
	moderate	4 to 6	6 to 10
	severe	7 to 8	11 to 15
EMSA	mild	1 to 2	1 to 3
	moderate	3 to 4	4 to 7
	severe	5 to 6	8 to 12
EBIA	mild	1 to 3	1 to 5
	moderate	4 to 6	6 to 10
	severe	7 to 8	11 to 15

Table 1 National classification of food insecurity using ELCSA, EMSA and EBIA.

2.1 First Study: Equating FIES and National Scales

The aim of this comparability study is to find raw scores on the national scales EBIA, EMSA and ELCSA that are *equivalent* to the continuous FIES global thresh-

olds used to compute the two indicators $FI_{Mod+Sev}$ and FI_{Sev} (i.e. -0.25 and 1.81). Data come from the administration of EBIA in Brazil in 2013, EMSA in Mexico in 2014 and ELCSA in Ecuador and Guatemala in 2016 and 2014 respectively and, in order to perform equating with FIES, only adult questions of the national scales have been considered. Equating between FIES and national scales was carried out by means of three equating methods for investigation purposes:

1. **IRT True Score** (IRT-TS) equating
2. **Linking** via a linear transformation applied to ability parameters
3. **Minimization** of the difference between prevalences of food insecurity

The *IRT-True Score equating* (IRT-TS) method is an IRT-based technique that consists of three steps [2]: at first, an IRT-model is fitted to the data (Estimation), then the parameters' estimates are put on a common metric through a linear transformation based on a set A of common items (Linking) and finally, equivalent expected Raw Scores are computed through the Test Characteristic Curves (TCC) of the two tests (Equating). In this study, two IRT models have been fitted to the data, namely the Rasch model and the nonparametric Mokken Scale [7] with a Kernel smoothing estimation of the Item Characteristic Curves (ICCs) [6]. When the nonparametric IRT model is considered, a Kernel smoothing estimation of the ICCs is computed and the points on the latent trait for which the corresponding estimated ICCs equals 0.5 are taken as the item severities used to estimate the linear transformation. The Standard Error of Equating (SEE) [5] for the IRT-TS method was estimated using 1000 bootstrap replications and, due to the computational costs of the procedure, is only provided together with the fitting of the Rasch model.

The second method (Linking) consists in considering the linear transformation obtained at the second step of the IRT-TS method and applying it to the estimated ability parameters of the Rasch model. Once ability parameters are adjusted to the Global standard metric, raw scores corresponding to the ability parameters that are the closest to the two global thresholds are considered as the *equivalent* raw score.

Finally, the third method (Minimizing) consists in computing prevalences of food insecurity at the household level applying the FIES methodology to the data used for the national scales and comparing the prevalences so obtained with the percentages of population scoring from a certain raw score on. The two raw scores that realize the minimum distance with the two global thresholds are considered the *corresponding* raw scores in accordance to this method.

2.2 Second Study: Comparing Adult- and Children-referenced item scales

This second analysis aims at comparing the Adult and the Children scales within each national scale. To this aim, we consider the scores obtained by the households with children on the two Module Surveys following the approach also known as the Single Group (SG) data collection design [5]. Equating of the Adult and Children

scales in the four countries was carried out through implementation of five equating methods for investigation purpose: IRT True Score equating with the Rasch model, Mean, Linear, Equipercentile and Kernel Equipercentile equating methods [5].

3 Results

Outcomes from the first comparability study are summarized in Table 2, reporting the raw scores equivalent to the global thresholds used for $FI_{Mod+Sev}$ and FI_{Sev} . Results show that the global threshold used for $FI_{Mod+Sev}$ sometimes reflects a *less severe* condition of food insecurity compared to the one measured by national scales for the moderate category of food insecurity, all equated raw scores being either equal to or around one point less than the thresholds currently used by ELCSA, EMSA and EBIA (compare Table 1). On the contrary, the global threshold used for FI_{Sev} generally reflects a *more severe* condition of the food insecurity than the one captured by the national scales for the severe level of food insecurity, equated raw scores being either equal to or one point higher than the national thresholds currently in use.

FIES	Food Insecurity Scales	Internal Monitoring	IRT-TS Rasch (SEE)	IRT-TS NP	Linking	Min. Diff.
$FI_{Mod+Sev}$	ELCSA (Guatemala)	4	3.3 (0.19)	3.4	3	4
	ELCSA (Ecuador)	4	4.2 (0.14)	4.1	4	4
	EMSA (Mexico)	3	2.0 (0.23)	2.0	2	2
	EBIA (Brazil)	4	4.0 (0.09)	4.0	4	5
FI_{Sev}	ELCSA (Guatemala)	7	7.8 (0.18)	8.0	8	8
	ELCSA (Ecuador)	7	7.1 (0.18)	7.7	7	8
	EMSA (Mexico)	5	6.0 (0.26)	6.0	6	6
	EBIA (Brazil)	6	7.9 (0.07)	8.0	8	8

Table 2 Raw Scores on the national scales equivalent to the thresholds for $FI_{Mod+Sev}$ and FI_{Sev} .

Regarding the second comparability study, the equated raw scores on the Children scale shown in Table 3 seem to suggest that sometimes the current thresholds reflect a different severity of the condition measured by the two scales referred to households with and without children, respectively. This is mainly evident for the most severe category of food insecurity, for which the corresponding raw scores on the Children scale for ELCSA in Guatemala and EMSA in Mexico are generally around one point *higher* than the thresholds currently in use and between one and two points *lower* for EBIA (column "Severe"). On the other hand, we see that the corresponding raw scores for the moderate food insecurity substantially align with the thresholds currently in use for this category (column "Moderate"). Moreover, minor differences emerge between the behaviour of ELCSA in Guatemala and Ecuador (analysis not shown), possibly due to the specific features of the phe-

nomenon in the two countries, confirming the importance of an equating analysis even between applications of the same scale. Finally, among all methods implemented, the Equipercentile equating method is the one whose results generally resemble the current thresholds the most.

We believe that these comparability studies can contribute in creating a more homogeneous and consistent picture of the phenomenon of food insecurity by allowing the utilization of results from application of different scales. As a consequence, this would enable a more reliable monitoring of the progress toward the goal of a global food security, as expressed in target 2.1 of the Sustainable Development Goals.

Scale	Equating Method	Moderate Raw score (SEE)	Severe Raw score (SEE)
ELCSA Guatemala	IRT-TS	6.2 (0.09)	12.1 (0.1)
	Mean	6.6 (0.07)	12.2 (0.07)
	Linear	6.5 (0.07)	11.7 (0.11)
	Equip	6.3 (0.09)	11.3 (0.15)
	Kernel Equip	6.1 (0.02)	12.0 (0.04)
EMSA Mexico	IRT-TS	4.8 (0.12)	8.7 (0.13)
	Mean	5.5 (0.05)	9.5 (0.05)
	Linear	5.1 (0.07)	8.6 (0.10)
	Equip	4.8 (0.13)	8.1 (0.14)
	Kernel Equip	4.5 (0.03)	8.8 (0.04)
EBIA Brazil	IRT-TS	4.8 (0.12)	8.7 (0.13)
	Mean	5.5 (0.05)	9.5 (0.05)
	Linear	5.1 (0.07)	8.6 (0.10)
	Equip	4.8 (0.13)	8.1 (0.14)
	Kernel Equip	4.5 (0.03)	8.8 (0.04)

Table 3 Raw scores on the Children scale respectively corresponding to 4 and 7 on the Adult scale (Guatemala), 3 and 5 on the Adult scale (Mexico) and 4 and 7 on the Adult scale (Brazil).

References

1. Cafiero, C., Melgar-Quiñonez, H.R., Ballard, T.J., Kepple, A.W.: Validity and reliability of food security measures. *Annals of the New York Academy of Sciences* **1331**(1)
2. Cook, L.L., Eignor, D.R.: Irt equating methods. *Educational measurement: Issues and practice*
3. FAO.2016: Methods for estimating comparable rates of food insecurity experienced by adults throughout the world. Rome, Italy. FAO
4. Fischer, G.H., Molenaar, I.W.: Rasch models: Foundations, recent developments, and applications. Springer Science & Business Media (2012)
5. Kolen, M.J., Brennan, R.L.: Test equating, scaling, and linking: Methods and practices. Springer Science & Business Media (2014)
6. Ramsay, J.O.: Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika* **56**(4), 611–630 (1991)
7. Sijtsma, K., Molenaar, I.W.: Introduction to nonparametric item response theory. Sage (2002)
8. Urquía-Fernández, N.: La seguridad alimentaria en México. *Salud Pública de México* **56** (2014)