Universidad Nacional Autónoma de México

Programa Universitario de Estudios del Desarrollo UNAM

Documento de trabajo

Changes in reliability over time of Mexico's official poverty measure: 2008-2018

Héctor Nájera

21

5 de diciembre

2019

Changes in reliability over time of Mexico´s official poverty measure: 2008-2018

Dr. Héctor Nájera.
Programa Universitario de Estudios del Desarrollo (PUED). UNAM
Honorary Research Fellow. Univeristy of Bristol.

1 Abstract

Townsend (1979)'s theory of relative deprivation predicts that indicators of material and social deprivation become less suitable to identify poverty over time and, consequently, less reliable and valid to measure poverty. CONEVAL's measure of poverty was put forward in 2008, and it is vital to know if all or some of its indicators are still suitable to measure poverty in Mexico. Townsend's prediction can be framed in terms of measurement theory and assessed using both validity and reliability principles. Reliability is a property that tells us the amount of signal we are capturing with a set of indicators and is a necessary condition for valid deprivation scores. This paper assesses the amount and sources of measurement error of CONEVAL's index for the period 2008-2018 using latent variable methods. The findings suggest that the overall reliability of CONEVAL's index has fallen below reasonable standards and that some dimensions (and its indicators) require updating. The paper reflects on the implication of these findings and puts forward some recommendations to improve the official Mexican measure.

2 Introduction

The concept of poverty has several meanings (e.g. basic needs, exclusion, social position) whose significance is given by different theories (Relative Deprivation (Townsend 1979); Capability framework (Sen 1981); Unsatisfied Basic Needs (Max-Neef, Elizalde, and Hopenhayn 1992)) (Spicker, Alvarez, and Gordon 2006). Concepts in poverty research are many, but some are more useful for poverty research than others. A minimum criterion for an acceptable and helpful definition in poverty research is the *scientific* character of the concept, i.e. clear-cut constructs that can be tractable and falsified via observation (Pantazis, Gordon, and Levitas 2006; Gordon 2006). One definition that fulfils this standard is Townsend (1979)'s definition: poverty is the lack of command of resources over time and material and social deprivations its main (observable) consequences (Townsend 1993; Gordon 2006). Within this framework, resources are not merely income but comprise a range of inputs that households use to fulfil their necessities of life. Thus, in his theory, there is a clear relationship between resources (the cause) and deprivation (the consequences).

As Gordon (2006) points out, there is barely a disagreement about the relations established between resources and deprivation by this definition of poverty. Of course, there are differences in the terms used but not necessarily concerning the specification about inputs and outputs in the form of deprivation. For example, Sen (1981)'s capability theory states a

similar connection between the source of poverty and the consequences of it. Sen proposes a relationship between capabilities, functionings and achievements (the things and activities we can observe). Similarly, social rights unfulfillment relates to some deprivation (Gordon et al. 2003).

The disagreements in the conceptualization of poverty revolves around the space to identify deprivation/achievements/unsatisfied needs (Boltvinik and Hernández-Láos 2001). Once poverty is clearly defined, there is a question about how to *identify* the poor via data (Sen 1981). The identification problem in poverty research refers to the selection of the resources/capabilities and deprivations/achievements/needs/social rights deprivation capture via data the concept poverty (Boltvinik and Hernández-Láos 2001; Gordon 2006; Alkire et al. 2015). Many of the debates about poverty measurement revolve around how to select a suitable set of deprivations and resources that will help to identify the poor from the not poor. There are different theories of human needs, deprivation, capabilities and social rights that propose different sets and interpretation to deprivations (Townsend1993a, Townsend 1987). Therefore, the number and contents of dimensions, the thresholds for the indicators and weights vary from one framework to another.

In the early 2000s, the National Council for the Evaluation of the Social Development Policy (CONEVAL) faced the task of addressing these two central aspects in poverty measurement (definition and identification operation) (CONEVAL 2011). CONEVAL had to address both elements following the Mexican legislation. In particular, the General Social Development Law (2004), which draws upon the Mexican Constitution to enlist a series of rights and aspects that are regarded to be constituent of poverty. The Law provided CONEVAL *some* guidelines for the structure of the MPM but insufficient information about the contents of each dimension. The Law stated the set rights but no details about the deprivation indicators. To put differently, CONEVAL did not have too much clarity for full identification of the poor based on a rights perspective.

CONEVAL thus undertook a complex, lengthy and robust process of consultation to give content to the measure and to decide how to aggregate the indicators (CONEVAL 2011). Human rights are ill-defined [Donnelly (2013); Fukuda-Parr2011; Sen2004]; and rarely data is collected to explicitly measure human rights or, for instance, poverty (Beccaria and Fernández 2020). Therefore, having human rights as a reference framework compromised the depth and scope of the dimensions and indicators. CONEVAL drew upon the Latin American tradition of multidimensional poverty measurement and took some aspects from unsatisfied basic needs approach. In particular, the dimensions of housing, education and services. CONEVAL also relied upon relative deprivation theory and upon Mack and Lansley (1985)'s consensual method to design the national survey about the socially perceived standards (CONEVAL 2007)[1]. It is unclear, for instance, if CONEVAL could have fully used methods like the consensual deprivation method to set the thresholds for some variables. The final section discusses this issue].

The design and production of the official Mexican measure, like any other measurement process, involved making several assumptions and compromises and this invariably results in some degree of *measurement error*. Noisy measurement is inevitable because poverty is

[1]It is not clear for the readers of the Law if CONEVAL had the mandate to fully use the human rights perspective to define the relevant space to construct the deprivation indicators

a concept, and it is not directly measurable. Thus, it is reasonable to expect indexes that capture variability that is not of interest. Spearman (1904)'s seminal paper set the bases to frame measurement error, and it is at the core of measurement theory and methods (Loken and Gelman 2017; Bandalos 2018).

Drawing upon measurement theory, Gordon (2010) provided an idea of the potential of the official Mexican measure using validity and reliability analyses using the 2005 household income and expenditure survey. His results suggested that almost all indicators lead to reliable and valid scores. Nonetheless, Gordon (2010) study illustrated that CONEVAL's index could be highly reliable without dimensions, but his analysis did not provide an idea of what would happen after aggregation. Hence, he suggested that CONEVAL should strive for better data and be careful with dimensions with few indicators and validity problems like social security and health. Gordon (2010)'s empirical work gave CONEVAL an idea of the lower bound reliability for the official measure and the importance of new and better data for increasing it.

Measurement error changes over time. In poverty research, there is a clear theory that proposes that we should expect a drop in reliability. Townsend (1979) theory states that deprivation is relative and predicts that some indicators become less suitable over time. From the measurement point of view, this loss should have an impact upon the extent of measurement error. There have been some exercises that have looked at some aspects of the reliability of the measure with the most recent data. Guillen (2017) showed that some dimensions lead to unreliable scores and that many of the indicators required updating. Similarly, Nájera (2016) showed that for young people, the thresholds of the measure are too severe and that the index needed some updating. Nonetheless, there has not been a comprehensive decomposition and examination of the reliability of the official scale. This paper aims to assess and decompose the reliability for the biannual series 2008-2018. The questions are: How does the overall reliability of the official measure has changed from 2008-2018? What is the effect of the aggregation across dimensions upon reliability? Which indicators need updating?

The paper is organized as follows. The first section presents some considerations, drawn from relative deprivation theory, for the official measure. The second section offers an overview of the methods. The third section presents the findings, and the last one concludes the article.

3 Theoretical considerations for changes in reliability overtime of poverty measures

Townsend (1979) observed that living standards changed over time and that society is the relevant *space* to define the activities and things to cover the necessities of life. Townsend construes poverty as follows (Townsend 1979, 31):

> *Individuals, families and groups in the population can be said to be in poverty when they lack the resources to obtain the types of diet, participate in the activities and have the living conditions and amenities which are customary, or are at least widely encouraged or approved, in the societies to which they belong. Their resources are so seriously below those commanded by the average individual or*

family that they are, in effect, excluded from ordinary living patterns, customs and activities.

This definition of poverty states a relationship between resources and *objective* deprivation. It is important to underline some of the connections between relative deprivation theory and CONEVAL's approach to define and identify poverty to frame the potential loss in *suitability* of some indicators to measure poverty in Mexico, i.e. how adequate are some indicators to reflect deprivation of social rights in this case. CONEVAL did not use relative deprivation theory to define the dimensions and the indicators of the Multidimensional Poverty Measure (MPM). Instead, CONEVAL's set of dimensions reflected the Mexican legislation and the human rights perspective was used in the main framework. The Law proposes a series of rights/dimensions that constitute poverty. CONEVAL index is multidimensional in that the rights capture different aspects that are necessary to live with dignity and freedom. The index consists of a hierarchical structure where multidimensional poverty is at the top with two main domains: income and social rights (where six dimensions are nested). The human rights literature propose the idea of *minimum core* and this seems to relate to a notion of absolute minimum fulfilment of social rights (Sengupta 2008). What is somewhat clear is that realisation of social rights leads to a living a decent life is the basis of the connection between human rights and the freedom from poverty (Pogge 2005). Furthermore, such a minimum core is ill-defined, and human rights fulfilment is a continuum, and the standards (cut-offs) are relative to society (Pemberton 2007). For instance, the set of rights established in the legislation to define poverty are *relative* in that are meant to represent the will of the people at the time of writing the Law.

When CONEVAL was designing the MPM, the human rights literature agreed that it was essential to measure human rights, but it was very unclear how to do so [Pemberton (2007); UN (2012); Fukuda-Parr2011]. There is no consensus about how human rights should be measured, and the applied literature uses available data and attaches a human rights interpretation of the existent indicators. Very rarely, survey questionnaires are explicitly designed to measure human rights. Instead, CONEVAL uses the Law as a reference to delimit the dimensions of the multidimensional measure. In measurement, this means that CONEVAL used the legislation as means for the content validity of the dimensions of poverty. The second challenge was to find a strategy to provide content to these dimensions. It is unclear from the Law whether CONEVAL had to use human rights-indicators to measure poverty. This grey area is not an irrelevant point as CONEVAL could have used another theory or framework to provide poverty-relevant content to each dimension. CONEVAL followed a hybrid approach that aimed to make the most of the different available frameworks. CONEVAL used three primary sources to decide the thresholds for some indicators: The Law, experts' recommendation and the socially perceived standards of the Mexican population (CONEVAL 2011). Yet, there are some discrepancies between the last two. Overall, the thresholds chosen by the experts were more severe (i.e. capture acute deprivation) than those that emerged from the population (i.e. capture less acute deprivation) (Guillen 2017).

Under closer inspection, the indicators of the dimensions of the official measure relate can relate to different traditions (Boltvinik 2014). For example, the indicators of adequacy of dwelling and basic services can be traced back to the unsatisfied basic needs (UBN) approach that has been widely used in the region since the late 1970s in Latin America (Beccaria and Minujín 1985; Boltvinik 2014). Guillen (2017) shows that most of the thresholds for

these indicators mismatch the socially perceived standards of the Mexican population. Other indicators, like those used for education, health and social security, draw upon some standards of the Mexican legislation and mainly capture institutional access to these social rights. The food security indicator draws upon FAO's food security scale and the National Income and Expenditure Household survey was modified to include it in the module. Yet, other modules like the consensual deprivation module include questions on safe and regular access to food.

CONEVAL's measure is a product of a great effort than involved several actors and aimed to have suitable indicators and dimensions to measure poverty in 2008. How much is enough to live with dignity? Townsend (1979) pointed out that measuring poverty invariably involves making value judgements, but that does not mean that poverty cannot be objectively measured. From empirical evidence, it is clear that the standards of the population captured via the consensual method (Mack and Lansley 1985), tend to lead to suitable, reliable and valid indicators of deprivation (Guio, Gordon, and Marlier 2012; Guio et al. 2017; Nandy and Pomati 2015; Lau et al. 2015; Estadística-Censos 2019). The existent gap between the chosen thresholds of CONEVAL and those that emerged from the population could compromise the quality of the MPM. Townsend's prediction suggests then that the components will no longer be suitable overtime due to changes in living standard in society. Furthermore, the gap between the MPM's cut-offs and the socially perceived standards has very likely increased over time. From the point of view of measurement, this would mean a fall in both reliability and validity of the scores. The deprivation count will not reliably rank the population from very low to higher living standards.

The lack of suitability and the fall in reliability and validity is, of course, an empirical question. Townsend's hypothesis demands to put forward empirical evidence. The following section describes the procedure to undertake a statistical study of such a nature. The prediction would mean the resulting (and inevitable) random noise of aggregating indicators to measure a construct will increase over time. The leading cause of this loss is that some indicators will be too severe (i.e. too extreme). Furthermore, the indicators will have low discriminatory power to distinguish the poor from the not poor in its own.

3.1 Measurement theory and updating

In poverty research, not often there is a clear distinction between the different theories to propose, produce and empirically examine a poverty index. That is, a theory to frame the meaning of poverty (i.e. the significance of the concept), a theory to define the properties of space to measure it (the dimensions and aspects of identifying of poverty), and a theory to produce an empirically examine a measure (aggregate and measure poverty via instruments and models). This paper draws upon relative deprivation to frame the relevant space to *identify* the suitability of substantive deprivations and on measurement theory to put under scrutiny CONEVAL's measure (Bandalos 2018).

It is vital to understand the governing principles of measurement theory, as these have wide-ranging implications for poverty indices. It helps to frame problems like aggregation, weighting, cut-offs, updating, comparing groups or times, and error in measurement. The history of measurement theory is not linear and uncontested. More than 100 years ago, Spearman (1904) put forward a series of ideas that are now the basis of contemporary measurement theory (Loken and Gelman 2017). He makes an explicit acknowledgement that we measure concepts, that these are latent (are not directly observable), and that

error attenuates or observations and indices. However, years later, Campbell (1920) and others like Stevens and others (1946) conceptualised measurement as the process of assigning numerals to objects. For many years this view led to the production of axioms as a means to set a series of rules to some counting operations. However, the measurement of concepts cannot be credible should the gap between theory and empirical data remain open. Hence, this formulation about the meaning of the process of measuring has been superseded by a framework that proposes the use of empirical evidence: measurement theory (Bandalos 2018). The key idea is akin to Spearman (1904)'s approach and recognises that measurement is an indirect operation that requires both some desirable properties and empirical evidence (Kyburg 1984; Hanson 2018).

In poverty research, however, this consensus has yet to influence measurement approaches (Gordon and Catalán 2019) universally. Some schools continue conflating measuring with the process of assigning numerals to objects and focus on axioms and seem to treat poverty as a directly observable phenomenon (Santos and Villatoro 2019; Alkire et al. 2015; Alkire and Foster 2011). Instead, contemporary measurement theory makes a clear distinction between index construction and aggregation, i.e. between the reliability and validity of the scores and how to aggregate them (Kyburg 1984). Under this theory, constructs are not directly observable so that researchers make assumptions to connect data and definitions, impose some properties that should be assessed by empirical evidence. That is, its starting point both random and systematic errors contaminate measurement (Bandalos 2018; Loken and Gelman 2017).

The statistical term for a construct/concept like poverty is *a latent variable*. There are, however, two ways to think about latent variables: formative and reflective measurement models (Coltman et al. 2008). Understanding the difference is crucial to understand some current misconceptions about the meaning of measuring constructs in poverty research (Nájera and Gordon 2019, Santos2019, @Gordon2019). But it is also useful for addressing some of the challenges involved in updating and changing an existent measure. The first type suggests that the indicators cause/shape poverty. For example, a decrease in the prevalence of dirt flooring causes a reduction in poverty. That would mean that if all the indicators of the official measure are zero, it follows that the number of poor people would be zero. It is easy to see the problems with this approach as it leads to the paradox of resolving poverty by targeting some indicators and not its roots. In this approach indicators *form* the construct. That is, the observed outcomes do not necessarily share a common theme and thus are not interchangeable. Instead, the items are a full and perfect set of the phenomenon of interest. Empirical assessments of reliability are unfeasible, and it is unclear how to quantify the error term (Coltman et al. 2008). Understandably, this approach has been severely criticised as measurement seems to be more a process governed by opinion and value judgements (Edwards 2011, Wilcox2008). Furthermore, this approach is seldom used and inconsistent with most poverty theories. For example, Sen (1976) propose the monotonicity axiom as a means to say that if poverty decreases, income/deprivation should reflect such a change. That is, poverty is the main cause of the observed measures (Townsend 1979). Formative approaches are rare in most fields and yet they are still advocated in poverty research (Santos and Villatoro 2016, 2019; Alkire and Santos 2010; Alkire and Foster 2011).

Contemporary measurement theory mainly proposes working with *reflective* models, i.e. models where the indicators are manifests of a latent variable (Bollen 2002, @McDon-

ald1999). The basis of this approach can be traced back to Spearman (1904) with continuous development ever since (Cudeck and MacCallum 2012, @Loken2017). This is a more useful approach because indicators are exchangeable (i.e. what changes are the indicators but not the phenomenon of interest so solving poverty is not about solving its consequences but its manifestations), items share a common theme (i.e. multidimensional poverty), adding or dropping and item does not change the phenomenon of interest, and reliability can be empirically assessed. Theoretically speaking, this approach is consistent with existent frameworks of poverty (Sen 1976; Townsend 1979). That is, if a material or social deprivation is the consequences of the lack of command of resources over time (poverty), then these are the observed manifestations that we can use to measure poverty. Measurement theory is increasingly used in poverty research (Dotto et al. 2019; Guio et al. 2017, @Guio2016; Nandy and Pomati 2015; Nájera 2018; Moisio 2004; Whelan and Maitre 2005; Szeles and Fusco 2013).

Reflective measurement is very concerned with error as it recognises the inevitable resulting noise from imperfect identification (which involves indicators, dimensions and weights). This approach rests on two of the governing principles in science: *validity and reliability*. In measurement, *reliability* is all the variability we are interested in and noise is all the variability that is not part of our interest. That is, we would like to have indicators that are manifestations of poverty and not part of a different phenomenon. Reliability is formally defined as homogeneity in measurement -not to be confused with unidimensional- because we are aiming at having indicators that belong to the same construct: multidimensional poverty. Correlation is a necessary condition for causation. If the items have low or zero correlation, that would mean that there is no evidence to say that deprivation is a consequence of poverty. This relationship is akin to Sen (1976) monotonicity axiom in that changes in poverty should reflect changes in achievements. If a measure is lumpy (heterogeneous), such desirable behaviour of an index does not exist. If our indicators do not lead to reliable scores, then we cannot say with confidence that we are capturing the variability of interest. We are interested then if our indicators have the same source.

Reliability has another advantage that is relevant to the discussion of weighting schemes. Very reliable scores imply low error and no much information can be added to improve the scale (Nájera 2018). Weighting will always add error because the optimal set of weights is unknown. If reliability is high, differential weights will do nothing but worsen the scale. Differently put, a highly reliable index is self-weighing because there is little information one can add to enhance the predictive value of the observed scores. Thus, the non-differential weighting is the second-best alternative to aggregate the indicators. Reliability is crucial for other reasons. If the noise is low, our index is protected against information losses. Dropping an indicator will have little impact upon population orderings based on a deprivation score. This paper focuses on reliability because it is a necessary condition of sound measurement. A series of indicators could be highly homogeneous, but it could be measuring something else that is not poverty like the quality of life, well-being, etc. Validity is the amount of evidence we must interpret the scores in the way we want to do so. However, because reliability is a condition for validity, it is vital to first assess the former for evaluating the later. There are different ways to estimate reliability; the next section explains the estimators and the strategy followed in this paper.

4 Methods

4.1 Estimation strategy

There are different estimators of reliability (Revelle and Zinbarg 2009, @Zinbarg2005) — almost all estimates range from 0 to 1, where 1 is perfect reliability. The best estimates of reliability are ω and ω_h for multidimensional measures (McDonald 1999). The first estimate ω is useful to see the upper bound reliability (highest) of a scale. However, it has the limitation of not giving information about the influence of the multidimensional structure upon reliability, i.e. whether the dimensions are useful to produce deprivation scores. Both statistics need to be estimated from a Confirmatory Factor Model (CFA), i.e. a reflective model. Statistical identification is crucial for CFA modelling, i.e. the number of unknowns should be less than the number of knowns. Unfortunately, CONEVAL's measure is not statistically identified, i.e. some dimensions do not have enough information. For a dimension to be identified, it is necessary to have at least three indicators (Brown 2006). However, it is still possible to estimate especial cases of the general model. Three following CFA models were fitted:

1. Dimensional model with one-higher order factor:

CONEVAL produces a binary score for each of the six social rights. These, in turn, measure the social rights domain. Figure 1 shows the structure of Model 1. The direction of the arrows tells us that this is a reflective model in that right's deprivation are manifestations of overall (un)fulfilment of social rights. This model is restrictive because it collapses all the information of the observed indicators for each social right into one binary variable. From the perspective of measurement, this is not necessarily a bad practice; it depends on the cut-offs used to identify deprivation. However, this procedure often implies information loses, and the negative effect of this practice will depend on the internal reliability of each indicator -an aspect explored in Models 2 and 3-.

2. A partial multidimensional model with one-higher order factor:

It is possible to estimate an especial case of CONEVAL's multidimensional structure for the identified dimensions, i.e. those with a sufficient number of indicators. The three identified dimensions are dwelling's adequacy, basic services and food (in)security. Education has three items, but these are not different manifestations of the phenomenon of interest, i.e. these are the same manifestation disaggregated by age. Several entitlements are manifests of social security. However, legislation ties them to the household head's status and the indicators are redundant (yet it would be useful to pursue some alternative modelling strategy). Figure 2 shows a hierarchical model where the indicators are manifests of the three dimensions and these (partially) measure the higher-order factor. Because this is a multidimensional model, it is possible to estimate ω_h and ω. Hence, we will have an estimate of the reliability of the scores considering the three dimensions but also whether it makes sense to produce scores for these three dimensions.

3. A disaggregated model with one-higher order factor:

CONEVAL's measure has a hierarchical structure. That is, all the indicators are manifests of a higher-order phenomenon- multidimensional poverty. It should be expected

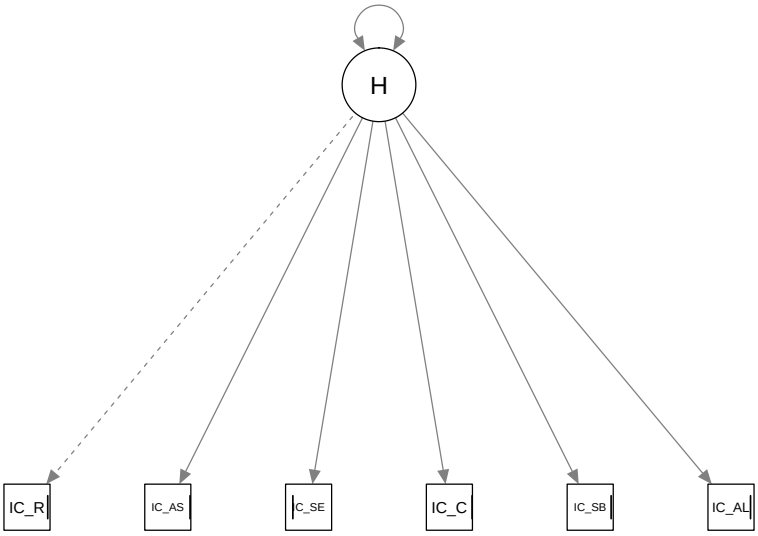

Figure 1: Higher-order factor only model. Full set of dimensions

that without dimensions, the items constitute a homogenous set. This specification enables us to include the indicators for the other three dimensions as a means to have an idea of its contribution to the reliability of the scores. Here we will use a CFA model but also Item Response Theory to assess some properties of the indicators (Reise 2014). In particular, we will examine two: severity and discrimination (Harris 1989). The first parameter tells us whether a given indicator of how acute the indicator is in terms of poverty. For example, whether lacking electricity is just too severe (i.e. too extreme) to provide any useful information to rank the population from the poor to the not so poor. Discrimination tells how well an indicator does in terms of splitting our population into meaningful groups: poor and not poor. For example, an item with high discrimination will tell us that if someone is deprived of those indicators, it is very likely that she will be classified as poor.

This paper estimates both ω and ω_h for the different models. Mplus 8.0 was used to compute ω and ω_h, The **R-Software** package `mplusAutomation()` helped to automate the estimations. Some functions were created to decompose by dimension ω and ω_h.

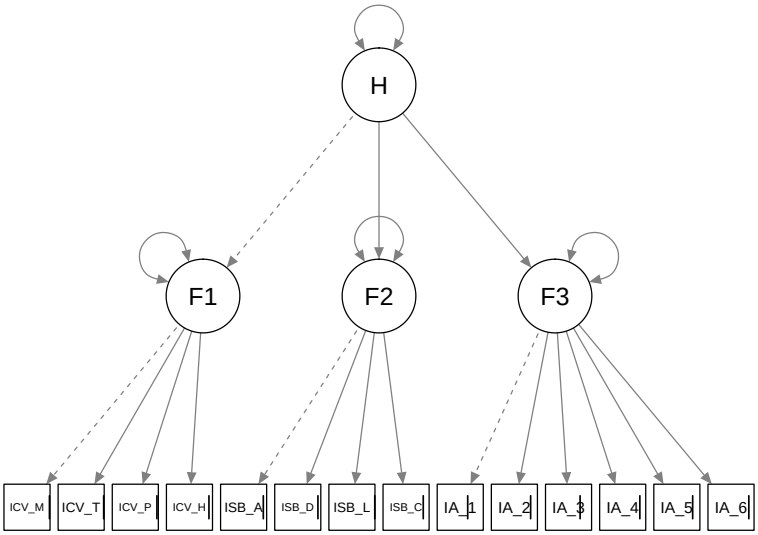

Figure 2: Higher-order factor hierarchical model. Identified dimensions

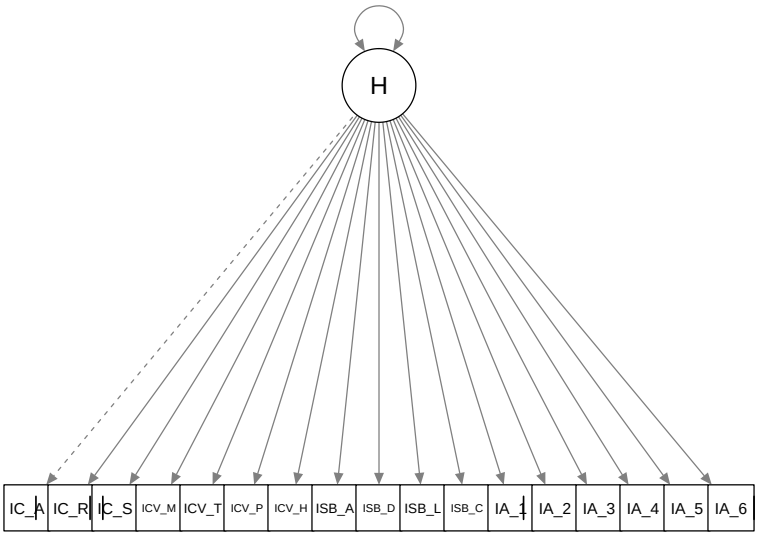

Figure 3: Higher-order factor only model. Full set of indicators

5 Results

This section presents the main findings of the reliability analyses for each one of the three models. The models are discussed in the following order: a) Unidimensional model with dimensions as observed variables; b) Partial multidimensional model with one-higher order factor and c) disaggregated model with one-higher order factor.

5.1 Unidimensional model with dimensions as observed variables

Table 1 shows the Model 1 omega total estimates for the biannual data of the period 2008 - 2018. It shows the upper bound reliability values for the model that considers a higher-order factor (social rights deprivation) measured through six indicators. The findings suggest that there is a definite drop in reliability. For the first round of the official multidimensional measure, reliability was near the acceptable standards ($\omega_t \geq .8$) (Nájera 2018). However, for the most recent years, the ω_t values are far below the recommended thresholds. This magnitude of omega mirrors the extent of the identification error of the poor relative to the not poor. Given a poverty line- is above 10%, i.e. at least 10% of the population that is misclassified given these six indicators and the poverty line (Nájera 2018). These results open the question about the effect of the aggregation upon overall reliability; the next section examines this problem from a different angle.

Table 1: Omega Estimates. First model (6 dimensions)

Omega	year
0.75	2008
0.72	2010
0.71	2012
0.73	2014
0.68	2016
0.68	2018

5.2 Partial multidimensional model with one-higher order factor

This section shows the reliability estimates of an especial case of CONEVAL's multidimensional model (Model 2). This hierarchical model is to model the dimensions with enough indicators per dimension. Table 2 shows the reliability estimates for ω_t and ω_h. The first column shows that ω_t is very high for this subset of indicators. That is, when put together, with no dimensional structure, measurement error is very small.

The values of ω_h tell us that classifying the indicators into dimensions is a sensible way to rank further the population. However, for the most recent rounds of the ENIGH, the values of ω_h are below the critical threshold ($\omega_h \geq .65$). These values suggest that although many indicators seem to belong to the same higher-order construct, some add unnecessary heterogeneity to the scale. In other words, some dimensions add more noise than a signal of interest to the index and are not that useful to classify the population using its deprivations scores. These results help to understand the low ω_t from Model 1. Nonetheless, it is possible to scrutinise further this loss in reliability by decomposing ω_t by dimension (See Table 3).

Table 2: Omega Total and Hierarchical Estimates. Second model (Hierarchical)

year	Omega Hierarchical	Omega Total
2008	0.66	0.96
2010	0.64	0.95
2012	0.63	0.95
2014	0.63	0.95
2016	0.57	0.94
2018	0.59	0.94

Table 3: Omega total per dimension

year	Omega_t Dwelling	Omega_t Services	Omega_t Food Security
2008	0.38	0.65	0.96
2010	0.34	0.60	0.96
2012	0.23	0.65	0.96
2014	0.47	0.59	0.96
2016	0.36	0.59	0.97
2018	0.52	0.57	0.96

Table 3 displays ω_t for each dimension for the period 2008-2018. The estimated values help us to understand the problem summed up by ω_h. The food security scale is highly reliable, and this is in sharp contrast with the other two dimensions. Both dwelling and essential services have very low-reliability values. That means that these two dimensions are contributing very little to the ranking of the population based on deprivation score. Being deprived in either of these two dimensions tells little about what the underlying level of deprivation is.

Table 4 shows the decomposition of the index's ω_t value. The food security scale accounts by for a lot of the information. These estimates help to see from a different angle the fall in ω_h and the very low ω_t of Model 1. It is essential to consider that a binary variable is produced from the scores of the food security scale. The aggregation results in a loss of information, and although this index is robust, it is insufficient to compensate for the shortcomings of the other dimensions.

Table 4: Omega decomposition. Contribution of each dimension

year	Omega Dwelling	Omega Services	Omega Food Security
2008	0.04	0.11	0.82
2010	0.03	0.09	0.84
2012	0.02	0.11	0.84
2014	0.06	0.08	0.82
2016	0.04	0.08	0.85
2018	0.07	0.07	0.83

5.3 Disaggregated model with one-higher order factor

Model 3 looks at the full disaggregated set of indicators of the six dimensions. When the items are considered without aggregation by social right, the values for the estimated omegas are: 0.76 and 0.6. This result is consistent with Model 2. However, we know that the food security scale account by for most of the variance. At this point, it is vital to assess which indicators are the primary source of the noise. This estimation can be done via Item Response Theory as we could inspect the severity and the discrimination properties of the indicators.

Figure 4 plots the estimated severity values for each indicator for 2008 (left) and 2018 (right). In 2008, many outcome measures showed low severity values, i.e. below the recommended severity threshold (< 3 sd) (Guio et al. 2017). Very high severity means that those indicators add valuable information and are not merely manifests of very extreme poverty. However, three items were too severe in 2008: roof materials, wall materials and electricity. That means that in 2008 lacking these indicators, although reflected very extreme poverty, added very little to the reliability of the ranking of the population based on the observed scores.

The plot with the 2018 severity estimates shows one of the reasons why there has been a loss in reliability over time. Nine indicators are above the recommended severity threshold. These means that these indicators add very little information in terms of classifying the population according to the underlying severity of poverty via the observed scores. This fall is one of the predictions of Townsend's theory of relative deprivation in the sense that some indicators become less suitable to identify poverty. These might be useful to diagnose extreme poverty in 2018 -given current living standards- but are not so convenient to identify those multidimensionally poor. This finding is one of the reasons for the low reliability of some of the dimensions.

Another useful property of a poverty indicator is its capacity to distinguish with high probability the poor from the not poor in its own. Figure 5 shows the estimated discrimination values for the year 2008 (left) and 2018 (right) for each indicator. As can be appreciated, only three indicators have discrimination values below the recommended threshold ($< .9$) (water, education and health care) (Nájera 2018; Guio et al. 2017). That lacking water (piped within the property) in Mexico helped little to distinguish the poor from the not poor, i.e. two people could have water piped within the property but very different living standards. Another way to interpret the impact of discrimination upon reliability is via its links with factor loadings (Nájera 2018). Low discrimination is equivalent to low factor loadings; this means that these three indicators are not sensitive to changes in latent poverty.

The 2018 discrimination estimates show that many of the indicators have very small values ($< .9$). For instance, only the six indicators of the food security scale have very high discrimination values. These estimates reflect the decrease in the suitability of the variables to measure poverty in Mexico for the period 2008 - 2018. The discrimination parameter is equivalent to the factor loadings of a unidimensional factor model. The IRT model thus shows from a different perspective the critical role that the EMSA indicators have for the reliability of the Mexican measure. However, when aggregating across dimensions, this contribution is outweighed by the information losses and the low reliability of the other five

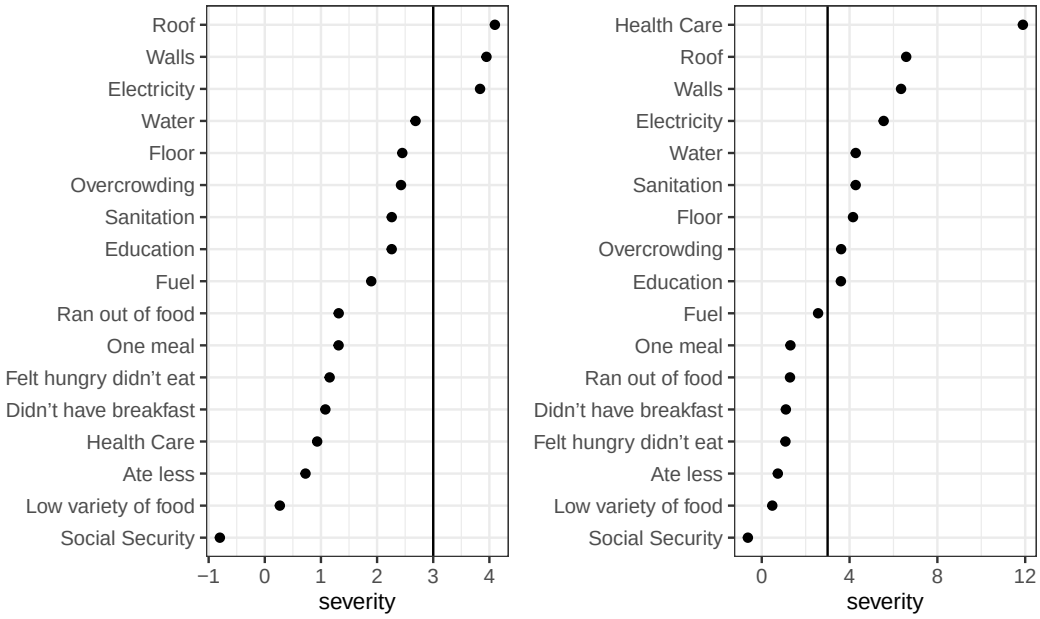

Figure 4: Severity estimates. 2008 (left) and 2018 (right)

dimensions.

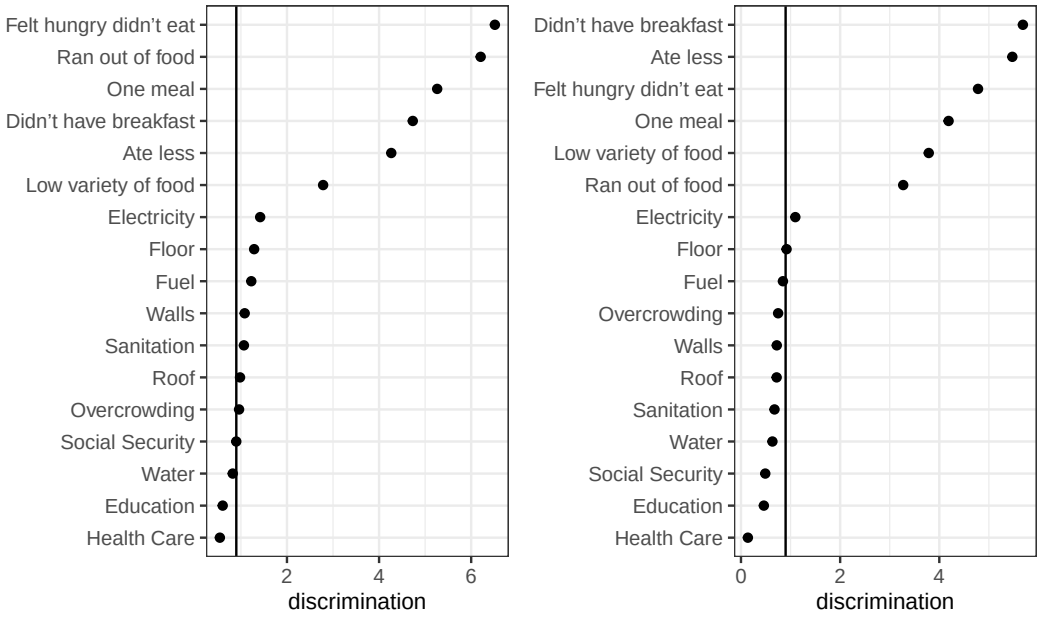

Figure 5: Discrimination estimates. 2008 (left) and 2018 (right)

6 Conclusion and discussion

Townsend (1979) theory of relative deprivation predicts that some deprivation measures become less suitable over time. This, of course, implies that the quality of measurement decreases invariably over time and we should be prepared to improve our multidimensional poverty measures periodically. Indeed, CONEVAL has been aware of this potential challenge and foresees an update.

This paper, drawing upon Townsend (1979)´s theory, examines whether reliability mirrors the loss in the suitability of some indicators for the period 2008-2018. The results show a sharp decrease in reliability, which seems to be caused mainly by the lack of depth of some items and the use of low thresholds for the current living standards.

Three kinds of models were put forward to estimate the overall reliability of deprivation scores of the social rights dimension and all three models lead to the same conclusions. First, the food deprivation dimension has highly reliable ratings, and these sub-scores explain a large part of the observed variance. Second, the other five dimensions require significant improvements to reduce measurement error (below is discussed how this could be achieved). Third, the aggregation of the available indicators into six dimensions leads to a substantial decrease in the reliability of the scores used to estimate multidimensional poverty, i.e. the sum scores that range between zero and six. Although this is an expected behaviour after aggregating the indicators, the sharp drop-in reliability leads to reliability values that are too low to be acceptable under current measurement standards.

The low reliability of the multidimensional poverty measure indicates that the indicators capture more noise (error) than signal (the phenomenon of interest). Policy-making demands sound indices, and this demands decreasing measurement error. Unfortunately, measurement error affects further statistical analyses of poverty. For example, hierarchical models use multidimensional poverty as a response variable. In the presence of noise, the estimates will be very biased and cannot be used to produce reliable small-area estimates. Furthermore, low reliability could produce biased results, or the evaluations of impact could also lead to misleading conclusions. Low reliability also adds problems to make conclusions over changes in poverty over time and across states or population groups.

Several questions arise from the findings of this paper: *Is it possible to increase the reliability of the official measure based on current data? If not, which kind of framework and data would be necessary? If there is an update, what are the implications in terms of comparability?*

With respect, the first question, several countries in Latin America are facing the same challenge. Relative deprivation theory suggests that the standards of the population, in that they lead to more suitable indicators, should lead to an increase in reliability. CONEVAL surveyed some of these standards, and the empirical literature suggests that higher thresholds (i.e. less severe) would increase reliability. It is unclear whether the 2007 standards will be enough and suitable in 2018, but this will likely help to produce a sharp increase in reliability.

If the increase is not enough, relative deprivation theory suggests updating the indicators and thresholds considering the socially perceived needs of the population. Several authors seem to agree on the need for improving survey modules. Estadística-Censos (2019) is a proposal that draws on a hybrid approach that combines the Unsatisfied Basic Needs (UBN) approach with the consensual deprivation method, which leads to very high reliability. From the existent literature, the recommendation is quite clear. Updating the thresholds of the existent indicators of both housing and services dimensions is likely to result in an enhancement of the reliability of the scores. For health, social security and education, it would be necessary to collect data about other aspects. For example, the consensual method asks about access to medicines, treatments, regular examinations (health); access to books, devices, desk/place to do the homework (education). These indicators are also useful in that it could be possible to produce child-specific measures of poverty (Gordon and Nandy 2012).

Concerning the third question, the reflective measurement approach has many useful advantages. Common sense tells us that we should keep the same indicators overtime to make valid and reliable comparisons across time points. This conclusion might be correct under formative models, but changing one indicator will imply changing the construct, and thus comparisons are not useful anymore. Furthermore, should the same items are preserved, the measurement error increases and make any comparison unreliable and invalid. Therefore, a formative approach is not useful for the task at hand. Reflective measurement is more helpful in that it allows changing the deprivation indicators, provided the sets of indicators that capture the latent construct equivalently. That is, changing indicators is necessary for high-quality measurement. This modification over time is akin to Townsend's prediction. Suitability of deprivation indicators is contingent to societies standards. The statistical term for this desirable behaviour is *measurement invariance* (MI) (Meredith 1993). MI can be achieved via a minimum common core of indicators (anchors) that measure the poverty equivalently across years. More outcome measures can be added to these anchors, and to

make 2018 scores comparable with scores of 2020, a procedure called scale equating and linking is required (González and Wiberg 2017; Kolen and Brennan 2004).

Measurement theory is not only about estimating error but also about fixing it. It is disappointing to see decreases in reliability for the official Mexican measure, but these are inevitable, expected and amendable. CONEVAL must take steps to enhance the quality of the MPM for the incoming years.

References

Alkire, Sabina, José Manuel Roche, Paola Ballon, James Foster, Maria Emma Santos, and Suman Seth. 2015. *Multidimensional Poverty Measurement and Analysis*. Oxford University Press, USA.

Alkire, S., and J. Foster. 2011. "Counting and Multidimensional Poverty Measurement." *Journal of Public Economics* 95 (7): 476–87.

Alkire, S., and M. Santos. 2010. "Acute Multidimensional Poverty: A New Index for Developing Countries." OPHI Working Paper No. 38.

Bandalos, Deborah L. 2018. *Measurement Theory and Applications for the Social Sciences*. Guilford Publications.

Beccaria, L., and A. Minujín. 1985. "Métodos Alternativos Para Medir La Evolución Del Tamaño de La Pobreza." Documento de Trabajo N. 6, INDEC, Buenos Aires.

Beccaria, Luis, and Ana Laura Fernández. 2020. "Midiendo La Pobreza Multidimensional Usando Encuestas de Hogares En Argentina." *Revista Latinoamericana de Economía* 51 (200).

Bollen, Kenneth A. 2002. "Latent Variables in Psychology and the Social Sciences." *Annual Review of Psychology* 53 (1): 605–34.

Boltvinik, J., and H. Hernández-Láos. 2001. *Pobreza Y Distribución Del Ingreso En México*. Edited by J. Boltvinik and H. Hernández-Láos. Siglo XXI Editores.

Boltvinik, Julio. 2014. "América Latina, de La Vanguardia Al Rezago En Medición Multidimensional de La Pobreza. La Experiencia Contrastante de México . Una Guía Para La Región." In *La Multidimensionalidad Como Un Desafío Para Los Métdos Y Técnicas de La Medición de La Pobreza*, edited by J et al. Boltviniki. CLACSO-CROP.

Brown, T. 2006. *Confirmatory Factor Analysis for Applied Research*. Edited by T Brown. The Guilford Press.

Campbell, Norman Robert. 1920. *Physics: The Elements*. CUP Archive.

Coltman, Tim, Timothy M Devinney, David F Midgley, and Sunil Venaik. 2008. "Formative Versus Reflective Measurement Models: Two Applications of Formative Measurement." *Journal of Business Research* 61 (12): 1250–62.

CONEVAL. 2007. "Encuesta Para La Determinación de Umbrales Mumultidimensional de Pobreza 2007." CONEVAL.

———. 2011. "Metodología Para La Medición Multidimensional de La Pobreza En México." *Realidad, Datos Y Espacio. Revista Internacional de Estadística Y Geografía* 2 (1): 36–63.

Cudeck, Robert, and Robert C. MacCallum. 2012. *Factor Analysis at 100: Historical Developments and Future Directions*. Routledge.

Donnelly, J. 2013. *Universal Human Rights in Theory and Practice*. Edited by J Donnelly. Cornell University Press.

Dotto, Francesco, Alessio Farcomeni, Maria Grazia Pittau, and Roberto Zelli. 2019. "A Dynamic Inhomogeneous Latent State Model for Measuring Material Deprivation." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182 (2): 495–516.

Edwards, Jeffrey R. 2011. "The Fallacy of Formative Measurement." *Organizational Research Methods* 14 (2): 370–88.

Estadística-Censos. 2019. "Prueba Piloto de La Medición de La Pobreza Multidimensional En La Ciudad de Buenos Aires." Dirección General de Estadistica y Censos. Ministerio de Economía y Finanzas. Buenos Aires Ciudad.

González, Jorge, and Marie Wiberg. 2017. "Applying Test Equating Methods." *New York: Springer. Doi* 10: 978–3.

Gordon, D. 2006. "The Concept and Measurement of Poverty." In *Poverty and Social Exclusion in Birtain: The Milenium Survey*, edited by C. Pantazis, D. Gordon, and R. Levitas, 29–69. Bristol Policy Press.

———. 2010. "Metodología de Medición Multidimensional de La Pobreza a Partir Del Concepto de Privación Relativa." In *La Medicion de La Pobreza Multidimensional En México*, edited by M. Mora, 401–98. El Colegio de México. CONEVAL.

Gordon, David, and Héctor E. Nájera Catalán. 2019. "Reply to Santos and Colleagues 'the Importance of Reliability in the Multidimensional Poverty Index for Latin America (Mpi-La)'." *The Journal of Development Studies* 0 (0): 1–5. https://doi.org/10.1080/00220388.2019.1663178.

Gordon, D., and S. Nandy. 2012. "Measuring Child Poverty and Deprivation: Measurement, Concepts, Policy and Action." In *Global Child Poverty and Well-Being*, 57–102. The Policy Press. University of Bristol.

Gordon, D, S. Nandy, C Pantazis, S Pemberton, and P. Townsend. 2003. *Child Poverty in the Developing World*. Edited by D. Gordon. The policy press. University of Bristol.

Guillen, Yedith. 2017. "Multidimensional Poverty Measurement from a Relative Deprivation Approach : A Comparative Study Between the United Kingdom and Mexico."

PhD thesis, University of Bristol.

Guio, A., D. Gordon, and E. Marlier. 2012. "MEASURING Material Deprivation in the Eu: Indicators for the Whole Population and Child-Specific Indicators." EUROSTAT.

Guio, Anne-Catherine, David Gordon, Eric Marlier, Hector Najera, and Marco Pomati. 2017. "Towards an Eu Measure of Child Deprivation." *Child Indicators Research*, October. https://doi.org/10.1007/s12187-017-9491-6.

Guio, Anne-Catherine, Eric Marlier, David Gordon, Eldin Fahmy, Shailen Nandy, and Marco Pomati. 2016. "Improving the Measurement of Material Deprivation at the European Union Level." *Journal of European Social Policy* 26 (3): 219–333. https://doi.org/10.1177/0958928716642947.

Hanson, Norwood Russell. 2018. *Perception and Discovery: An Introduction to Scientific Inquiry*. Vol. 389. Springer.

Harris, Deborah. 1989. "Comparison of 1, 2, and 3-Parameter Irt Models." *Educational Measurement: Issues and Practice* 8 (1): 35–41. https://doi.org/10.1111/j.1745-3992.1989.tb00313.x.

Kolen, M. J., and R. L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices*. Statistics for Social and Behavioral Sciences. Springer. https://books.google.com/books?id=tusOwjb7LWsC.

Kyburg, Henry E. 1984. *Theory and Measurement*. Cambridge University Press.

Lau, Maggie, Christina Pantazis, David Gordon, Lea Lai, and Eileen Sutton. 2015. "Poverty in Hong Kong." *China Review* 15 (2): 23–58.

LGDS. 2004. "Ley General de Desarrollo Social." Cámara de Diputados.

Loken, Eric, and Andrew Gelman. 2017. "Measurement Error and the Replication Crisis." *Science* 355 (6325): 584–85. https://doi.org/10.1126/science.aal3618.

Mack, J., and S. Lansley. 1985. *Poor Britain*. Edited by J. Mack and S. Lansley. London, George Allen & Unwin.

Max-Neef, Manfred, Antonio Elizalde, and Martin Hopenhayn. 1992. "Development and Human Needs." *Real-Life Economics: Understanding Wealth Creation*, 197–213.

McDonald, R. P. 1999. *Test Theory: A Unified Treatment*. Edited by R. P. McDonald. Mahwah, N.J. L. Erlbaum Associates.

Meredith, William. 1993. "Measurement Invariance, Factor Analysis and Factorial Invariance." *Psychometrika* 58 (4): 525–43. https://doi.org/10.1007/BF02294825.

Moisio, Pasi. 2004. "A Latent Class Application to the Multidimensional Measurement of Poverty." *Quality and Quantity* 38 (6): 703–17. https://doi.org/10.1007/s11135-004-5940-7.

Nandy, Shailen, and Marco Pomati. 2015. "Applying the Consensual Method of Estimating Poverty in a Low Income African Setting." *Social Indicators Research* 124 (3): 693–726. https://doi.org/10.1007/s11205-014-0819-z.

Nájera, Hector E. 2016. "Does Measurement Invariance Hold for the Official Mexican Multidimensional Poverty Measure? A State-Level Analysis 2012." *Quality & Quantity*, 1–25. https://doi.org/10.1007/s11135-016-0327-0.

Nájera, Héctor E. 2018. "Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study." *Social Indicators Research*, June. https://doi.org/10.1007/s11205-018-1950-z.

Nájera, Héctor E, and David Gordon. 2019. "The Importance of Reliability and Construct Validity in Multidimensional Poverty Measurement: An Illustration Using the Multidimensional Poverty Index for Latin America (Mpi-La)." *The Journal of Development Studies*, 1–21.

Pantazis, C., D. Gordon, and R. Levitas. 2006. *Poverty and Social Exclusion in Britain: The Millennium Survey*. Edited by C. Pantazis, D. Gordon, and R. Levitas. Studies in Poverty, Inequality, and Social Exclusion. Policy Press. https://books.google.com/books?id=o-H0J4BMWS8C.

Pemberton, David AND Nandy, Simon AND Gordon. 2007. "Child Rights and Child Poverty: Can the International Framework of Children's Rights Be Used to Improve Child Survival Rates?" *PLoS Med* 4 (10): e307. https://doi.org/10.1371/journal.pmed.0040307.

Pogge, Thomas. 2005. "Severe Poverty as a Violation of Negative Duties." *Ethics & International Affairs* 19 (1): 55–83. https://doi.org/10.1111/j.1747-7093.2005.tb00490.x.

Reise, Steven P. 2014. "Item Response Theory." In *The Encyclopedia of Clinical Psychology*, 1–10. American Cancer Society. https://doi.org/10.1002/9781118625392.wbecp357.

Revelle, William, and Richard. Zinbarg. 2009. "Coefficients Alpha, Beta, Omega, and the Glb: Comments on Sijtsma." *Psychometrika* 74 (1): 145–54. https://doi.org/10.1007/s11336-008-9102-z.

Santos, Maria Emma, and Pablo Villatoro. 2016. "A Multidimensional Poverty Index for Latin America." *Review of Income and Wealth*, n/a–n/a. https://doi.org/10.1111/roiw.12275.

———. 2019. "The Importance of Reliability in the Multidimensional Poverty Index for Latin America (Mpi-La)." *The Journal of Development Studies*, 1–6.

Sen, A. 1981. *Poverty and Famines: An Essay of Entitlement and Deprivation*. Edited by A. Sen. Oxford Clarendos Press.

Sen, Amartya. 1976. "Poverty: An Ordinal Approach to Measurement." *Econometrica* 44 (2): pp. 219–31. http://www.jstor.org/stable/1912718.

Sengupta, Arjun. 2008. "Poverty Erradication and Human Rights." 1. *Naciones*

Unidas. Vol. 1. United Nations.

Spearman, C. 1904. "The Proof and Measurement of Association Between Two Things." *Amérícan Journal of Psychology* 15 (1): 72–101.

Spicker, P., S. Alvarez, and D. Gordon. 2006. *Poverty and International Glossary*. Edited by P. Spicker, Alvarez S., and D. Gordon. International Studies in Poverty Research. International Social Science Council. Zen Books.

Stevens, Stanley Smith, and others. 1946. "On the Theory of Scales of Measurement."

Szeles, Monica, and Alessio Fusco. 2013. "Item Response Theory and the Measurement of Deprivation: Evidence from Luxembourg Data." *Quality & Quantity* 47 (3): 1545–60. https://doi.org/10.1007/s11135-011-9607-x.

Townsend, P. 1979. *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living*. Edited by P. Townsend. University of California.

Townsend, Peter. 1987. "Deprivation." *Journal of Social Policy* 16 (02): 125–46. https://doi.org/10.1017/S0047279400020341.

———. 1993. "The Politics of Poverty and Health." *BMJ: British Medical Journal* 306 (6873): p. 337. http://www.jstor.org/stable/29718417.

UN. 2012. "Human Rights Indicators. A Guide to Measurment and Implementation." United Nations. Office of the High Commisioner for Human Rights (OHCHR).

Whelan, Christopher T., and Bertrand Maitre. 2005. "Vulnerability and Multiple Deprivation Perspectives on Economic Exclusion in Europe: A Latent Class Analysis." *European Societies* 7 (3): 423–50. https://doi.org/10.1080/14616690500194050.

Zinbarg, RichardE., William Revelle, Iftah Yovel, and Wen Li. 2005. "Cronbach's α, Revelle's β, and Mcdonald's ω_h : Their Relations with Each Other and Two Alternative Conceptualizations of Reliability." *Psychometrika* 70 (1): 123–33. https://doi.org/10.1007/s11336-003-0974-7.

VNIVERSIDAD NACIONAL
AVTÓNOMA DE
MEXICO