# Validity / Validez

Héctor Nájera

29th Nov 2021

Validez, evidencia y scores latentes

Is there any evidence to justify my interpretation of the scores as poverty/deprivation scores?

On what basis can I conclude that 9 means poorer than 8, ... , 0?

Measurement models assume that the scores lead to such judgements.

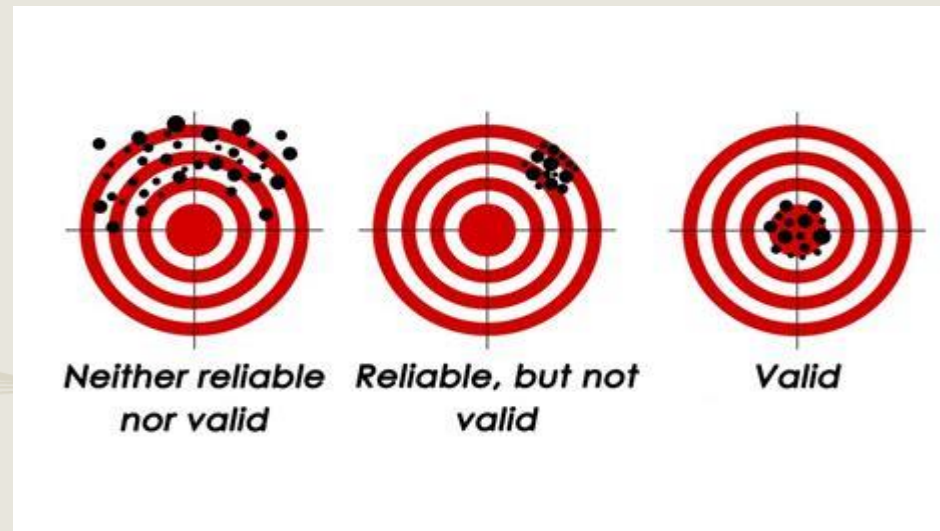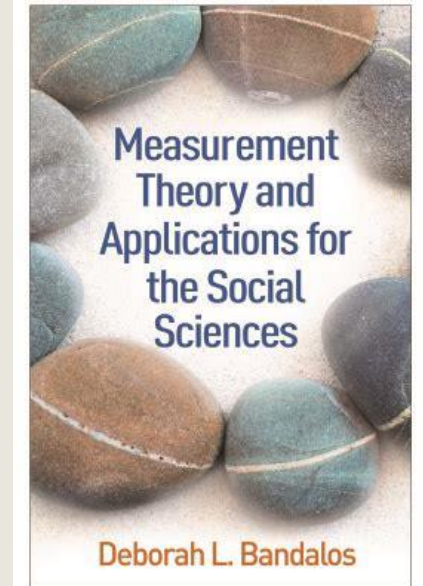Measurement models demand evidence for such interpretations
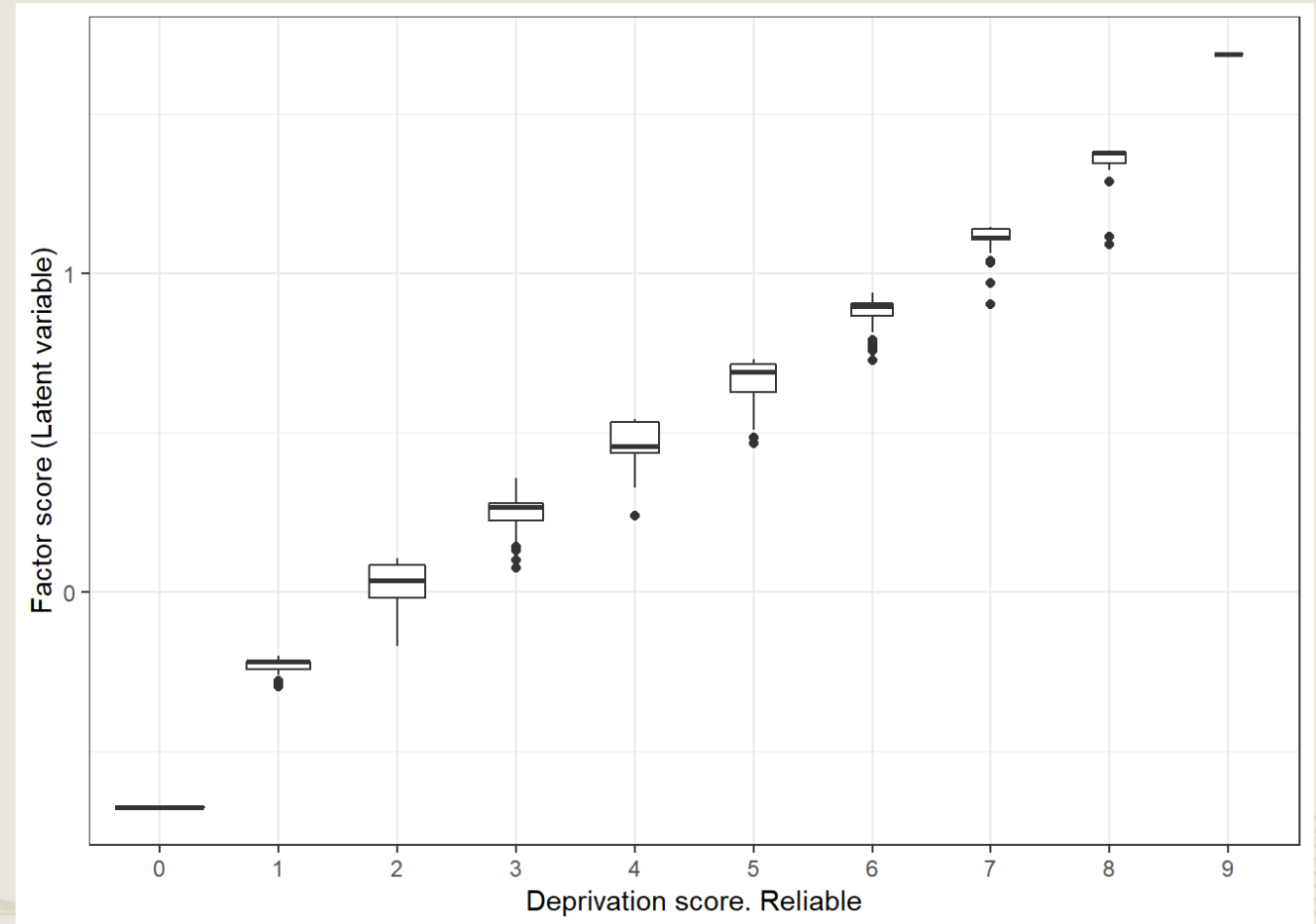
- It is all the evidence that:
- Justifies our interpretations of an index
- Supports the interpretation of the scores of an index
- Detects systematic error

·

Standards for Educational and Psychological measurement (APA, AERA)

Measurement Theory and Applications for the Social Sciences

Deborah L. Bandalos

Neither reliable nor valid

Reliable, but not valid

Valid

Validity, hypotheses and types of evidence

- Validity is a set of evidence regarding a series of hypotheses of the measurement model.
- The indicators capture have the same source (Reliability) and that source is Poverty
- The structure of the measurement model is an adequate representation of the data (D, I)
- Dimensions have theoretical and empirical meaning – they exist–
-

    – The indicators represent the phenomenon (poverty) of those who live it (poor population)
    – The poverty line leads me to the best possible separation according to the Measurement Model -Townsend breaking point-
    – The indicators in measure A point in the same direction as those in measure B

Evidence production/validation

Criterion, content and latent construct/model

Face, criterion and concurrent

- The degree to which a theory of poverty offers a series of logical statements – not contradictory – consistent and testable
- The degree to which a theory of poverty distinguishes between its causes and its possible consequences, so that there is clarity regarding the measurement model and the explanatory model
- The degree to which the measurement model is a priori a reasonable representation of poverty
-

Townsend: Poverty in the UK

Townsend

Note: This is not the actual book cover
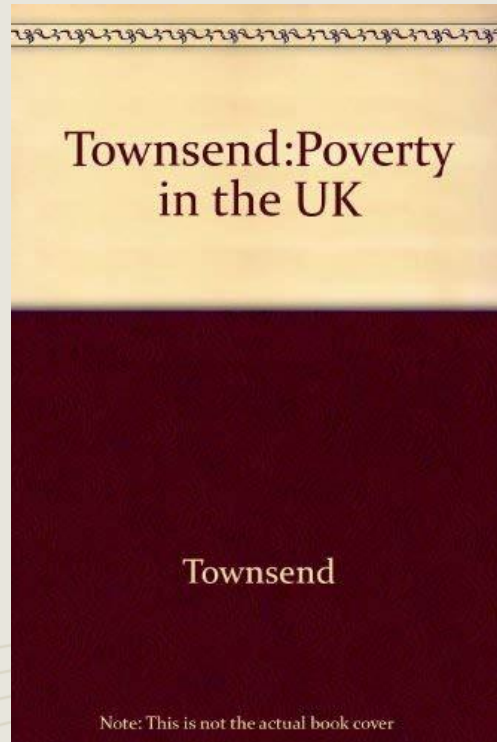
Structural inequalities >
> resource distribution systems
Resource control over > time
Deprivation – relative poverty & measurement model –

- The degree to which the population recognizes that the measurement model and its contents are a good representation of their situation:
- General population
- Expert Panels

- Methods:
- In-depth interviews
- Focus groups
- Representative surveys, e.g. identification of socially perceived needs or FIES scale
- Legislation/Social rights, e.g. the law represents the will of the population
-

Criterion validity

H: The indicators capture have the same source (Reliability) and that source is (Poverty)

If the indicators capture poverty, they should be associated with causes of poverty: economic crisis, discrimination, assessment of living standards, assessment of health status, position in the labour market.

Poverty theory must outline the terms of the hypotheses

Bi-varied correlation

Multivariate regression

Structural equation models

The key, as in the scientific enquiry, is that the results are consistent

Criterion validity: Examples

H1: Indicators are manifestations of poverty, that is, deprivation correlates with poverty markers

Validator 1—evaluation of household income status	Difficult compared to good or more or less OK (relative risk with 95 % CIs)
Number of meals every day	14.7 (10.8–20.1)
Consumption of cereals and tubers every day	9.4 (6.9–12.8)
Clothing	6.3 (5.0–7.9)
Shoes	6.0 (4.8–7.5)
A good meal on festivities/celebrations (Sunday, ceremony, etc.)	5.7 (4.3–7.5)
Consumption of meat or fish every day	5.6 (4.5–6.9)
Cleanliness/personal hygiene	4.1 (3.4–5.1)
Housing	3.8 (3.3–4.4)
Care in case of sickness	3.6 (3.1–4.3)
Self-care products in the house	3.5 (3.0–4.1)
Education for children	2.8 (2.3–3.2)
Availability of transport	2.5 (2.3–2.7)
Availability of leisure	2.2 (2.0–2.5)
Furniture in the house	2.2 (2.0–2.4)
Availability of drinking water	1.9 (1.7–2.1)
Availability of electricity	1.6 (1.5–1.7)

Validator 2—current financial situation	Goes into debt compared to able to save (relative risk with 95 % CIs)
Number of meals every day	8.3 (6.0–11.5)
A good meal on festivities/celebrations (Sunday, ceremony, etc.)	7.6 (5.3–10.9)
Consumption of cereals and tubers every day	7.6 (4.9–11.8)
Consumption of meat or fish every day	6.3 (4.8–8.1)
Shoes	5.4 (4.1–7.1)
Clothing	5.1 (4.0–6.7)
Cleanliness/personal hygiene	3.7 (3.0–4.7)
Care in case of sickness	3.2 (2.7–3.8)
Education for children	3.1 (2.5–3.8)
Housing	3.0 (2.6–3.6)
Self-care products in the house	2.9 (2.5–3.5)
Availability of leisure	2.5 (2.1–2.8)
Availability of transport	2.3 (2.1–2.6)
Furniture in the house	2.2 (2.0–2.4)
Availability of drinking water	1.8 (1.6–2.1)
Availability of electricity	1.6 (1.5–1.7)

Validator 4—asset index quintiles	Bottom quintile compared to top quintile (relative risk with 95 % CIs)
Availability of electricity	9.8 (8.3–11.6)
Shoes	7.5 (5.7–9.8)
Self-care products in the house	7.2 (5.6–9.2)
Clothing	7.0 (5.4–9.1)
Cleanliness/personal hygiene	6.9 (5.3–9.1)
Consumption of meat or fish every day	6.9 (5.2–9.0)
A good meal on festivities/celebrations (Sunday, ceremony, etc.)	6.8 (5.0–9.2)
Education for children	6.0 (4.8–7.4)
Care in case of sickness	5.6 (4.6–6.7)
Availability of drinking water	5.0 (4.2–6.0)
Number of meals every day	4.9 (3.7–6.6)
Furniture in the house	4.5 (3.9–5.2)
Consumption of cereals and tubers every day	4.1 (3.0–5.5)
Housing	3.7 (3.1–4.3)
Availability of transport	2.9 (2.5–3.2)
Availability of leisure	2.8 (2.4–3.3)

Evidence: Disadvantaged people are more likely to say that their income situation is difficult, get into debt and belong to the bottom quintile -Wealth Index-
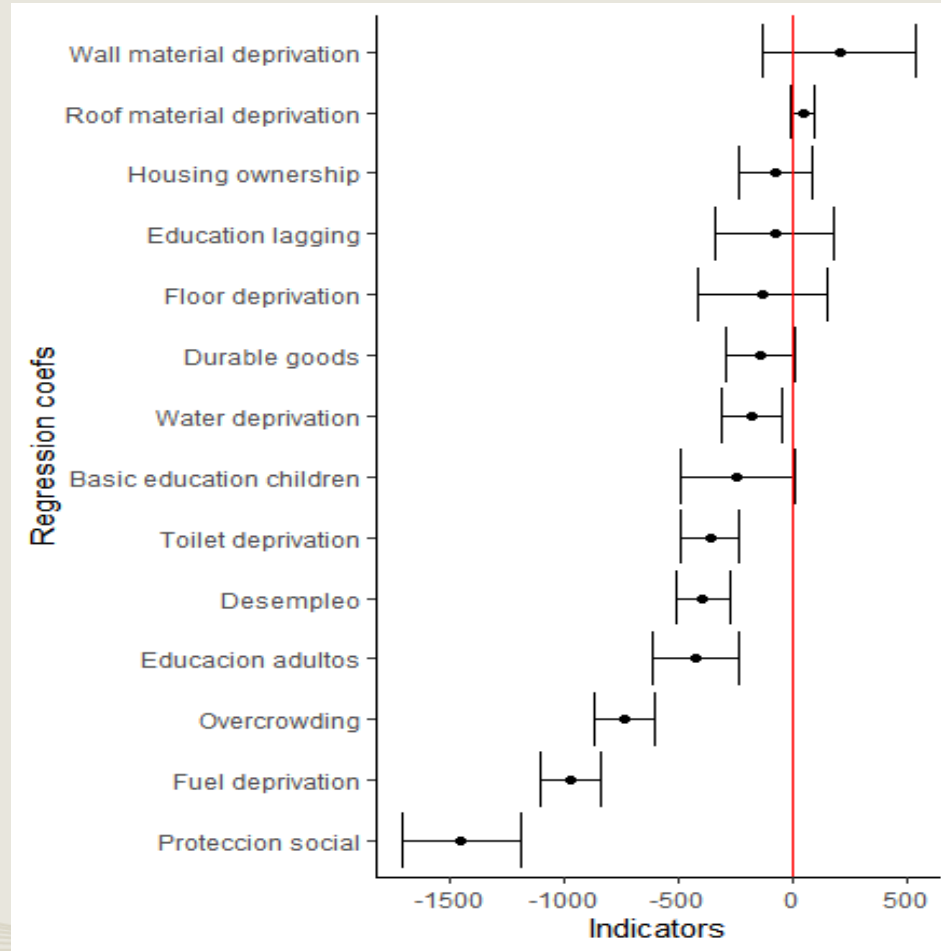
Criterion validity: Example

H1: Indicators are manifestations of poverty, that is, having lack correlates with poverty markers

Not all of them correlate with "income." Maybe it's not the best validator, but still...
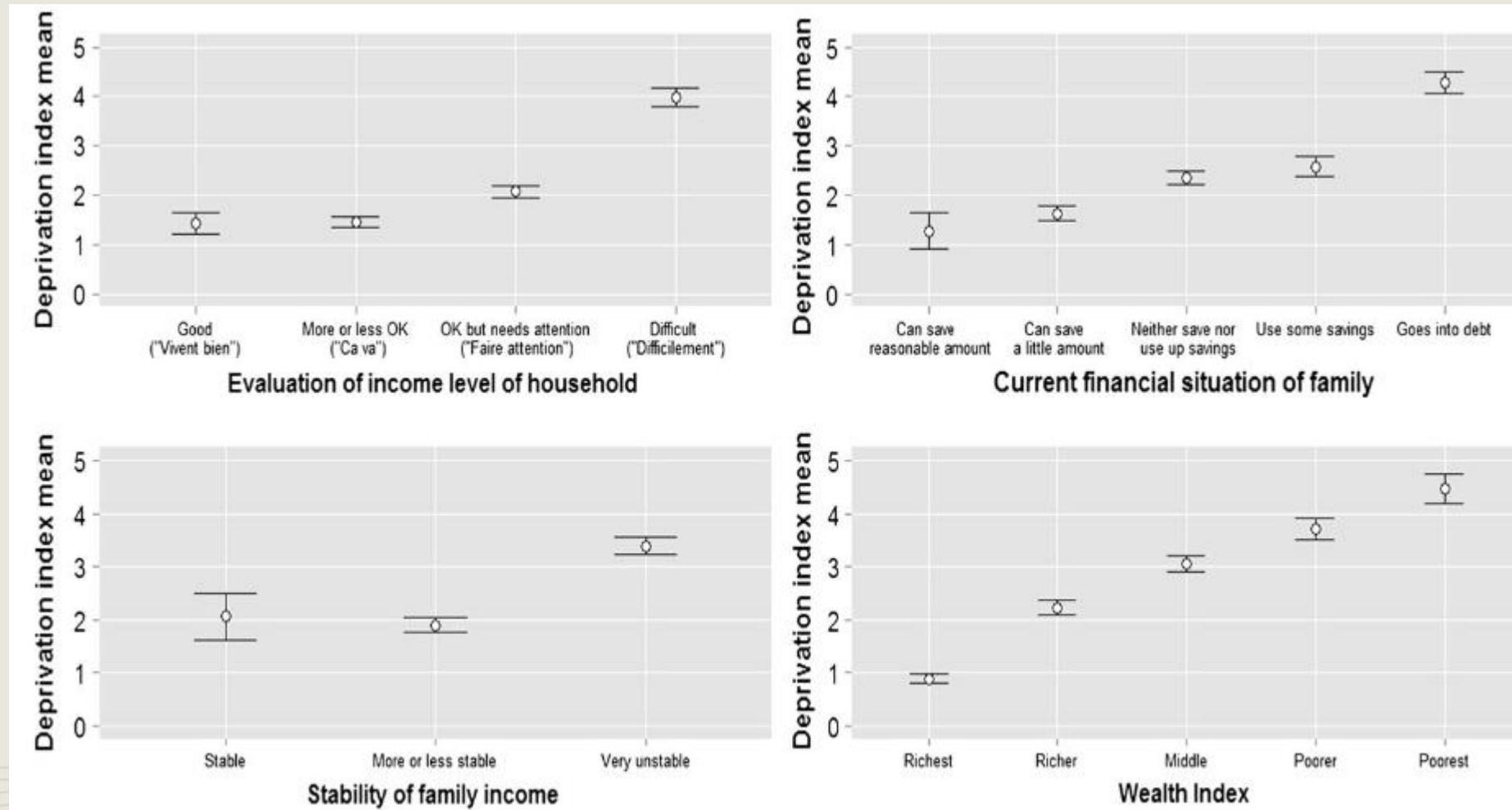

Data from the MPI-LA. Mexico 2014

Criterion validity: Examples

H: Higher deprivation scores mean higher poverty

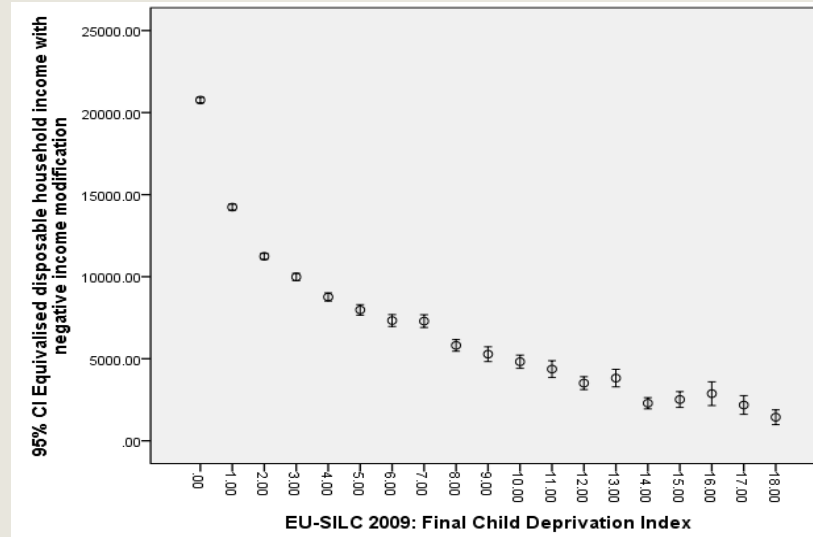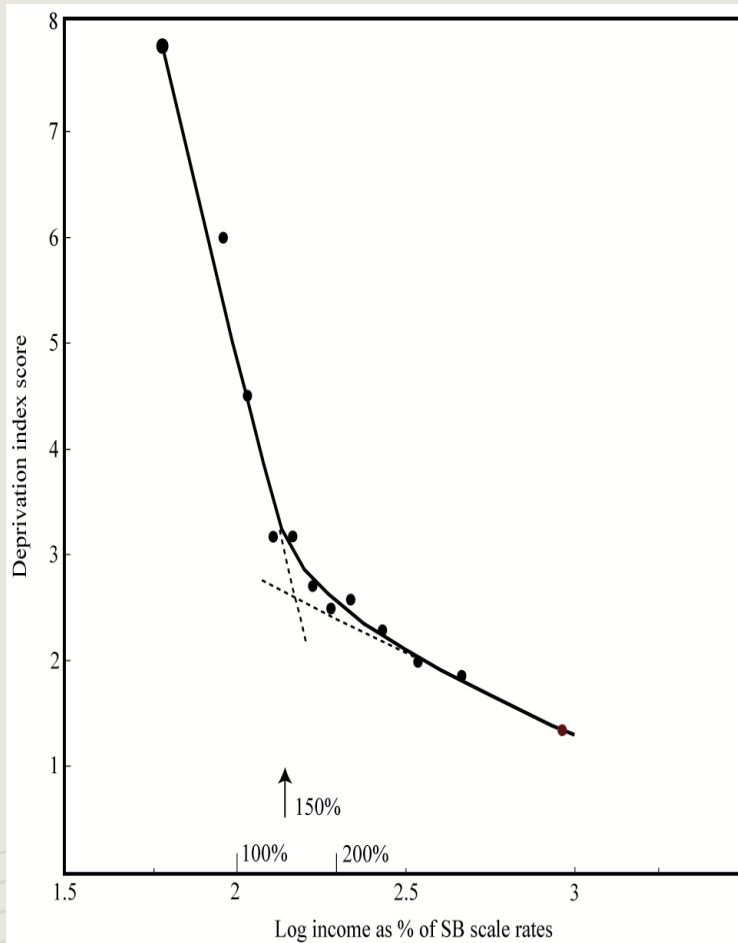Evidence: People's experiences are worse for those with higher deprivation scores

CrossMark

Applying the Consensual Method of Estimating Poverty in a Low Income African Setting

Shailen Nandy · Marco Pomati

Fig. 1 Testing scale validity. *Source*: Calculated from Benin DHS 2006

Townsend breaking point hypothesis: There is a level of resources from which multiple material deprivation increases substantially – poverty line.

Europe
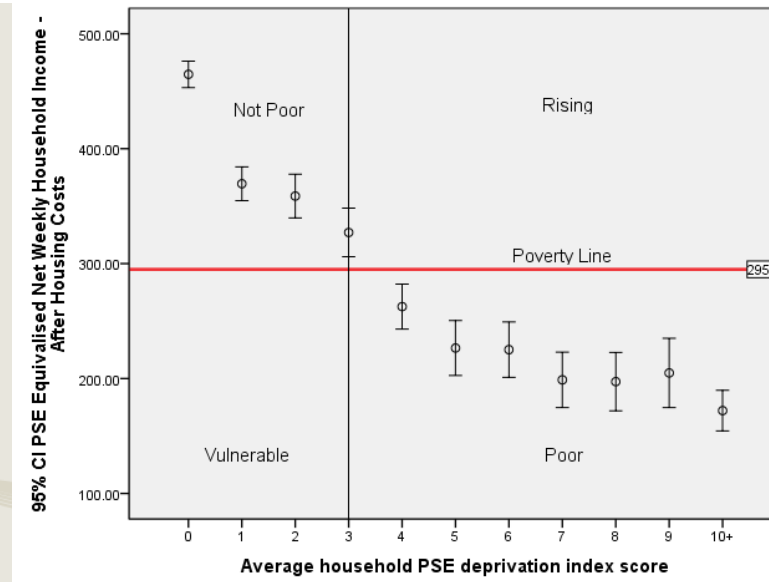
United Kingdom

Townsend breaking point hypothesis: There is a level
of resources from which multiple material deprivation
increases substantially – poverty line.

Mexico
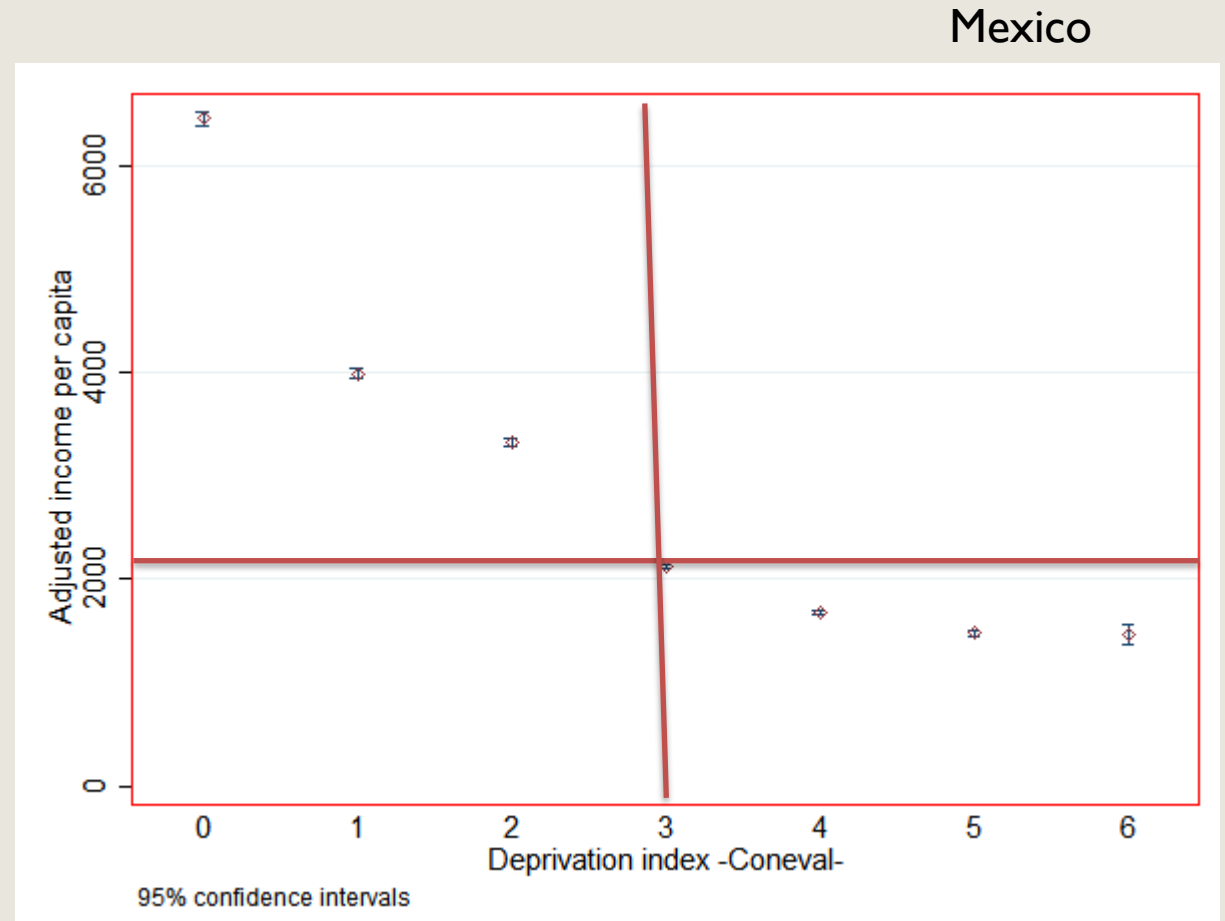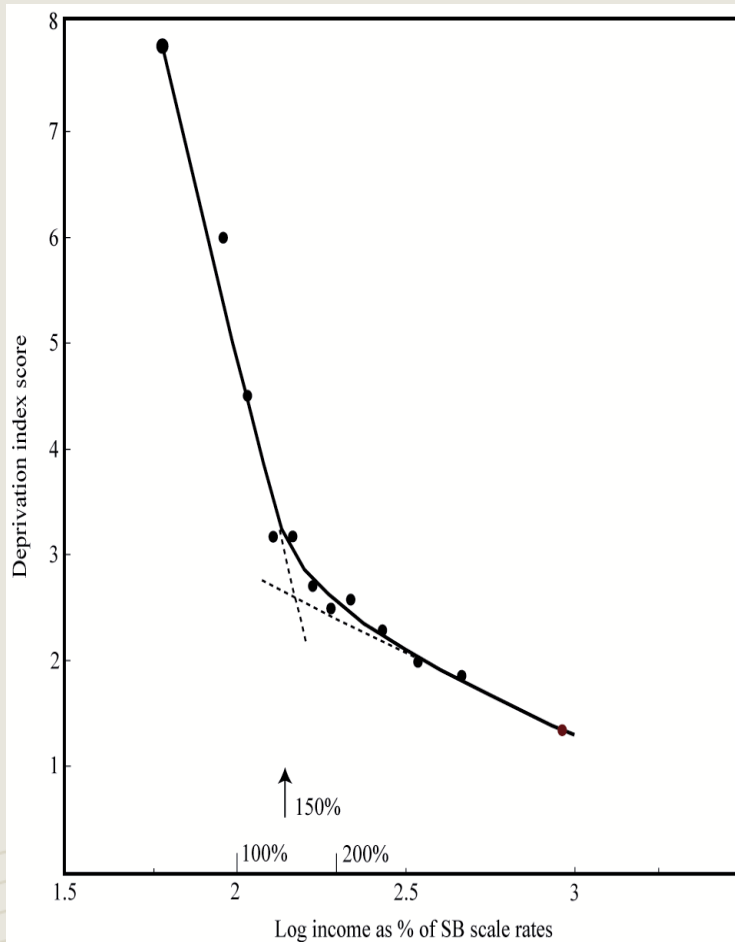

95% confidence intervals

Gordon: Logit and ANOVA Models – The deprivation split that Maximizes Model Fit–

Finite mixture models – Crossing of latent curves. When there is no measure of resources-

Latent class analysis: model-based account. With and without resources measure-

Hausman classification error model –With and without resource measure-

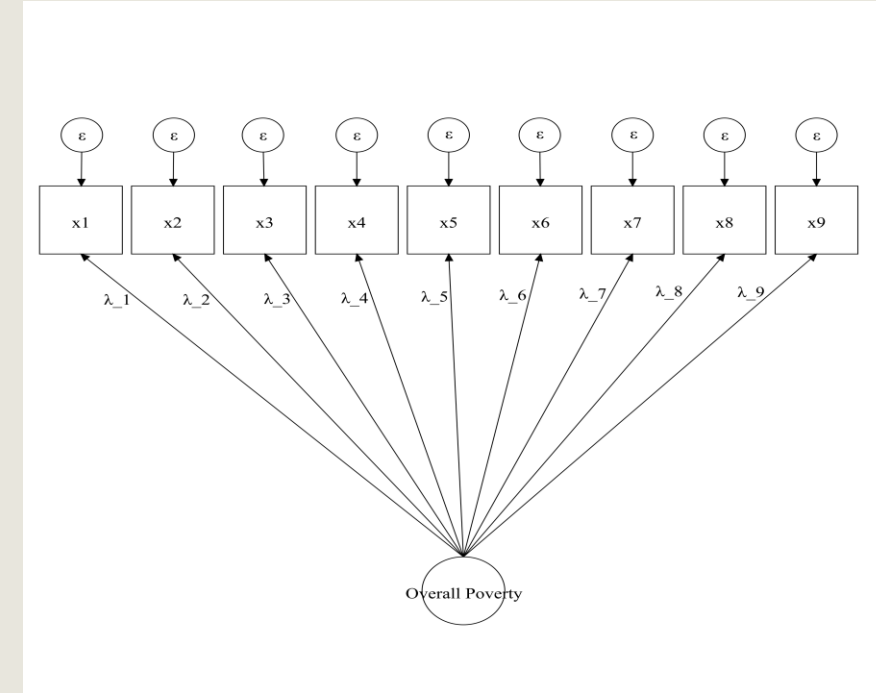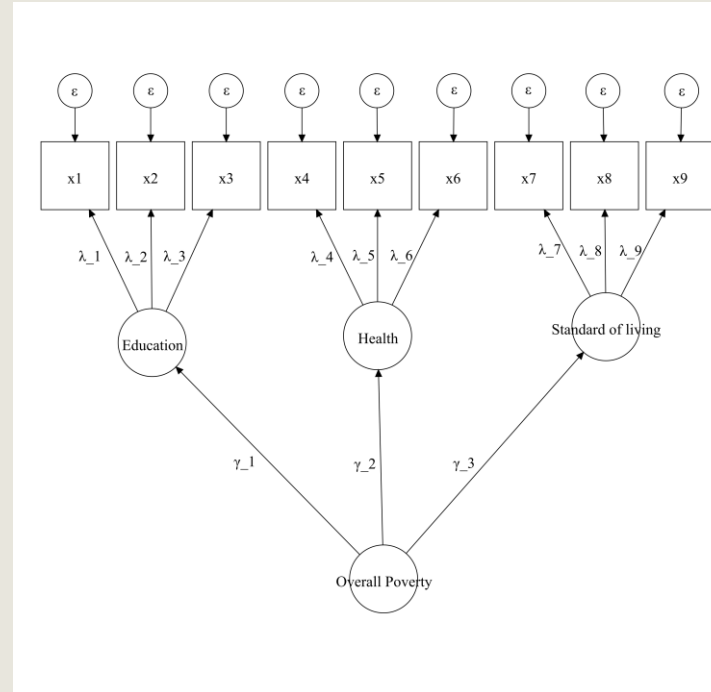Same data/indicators.
Different models.
Same number of indicators
Different dimensional
structure

There are several alternatives:

Different indicators, same
dimensional structure
Different indicators, different
dimensional structure
Different weights (?)

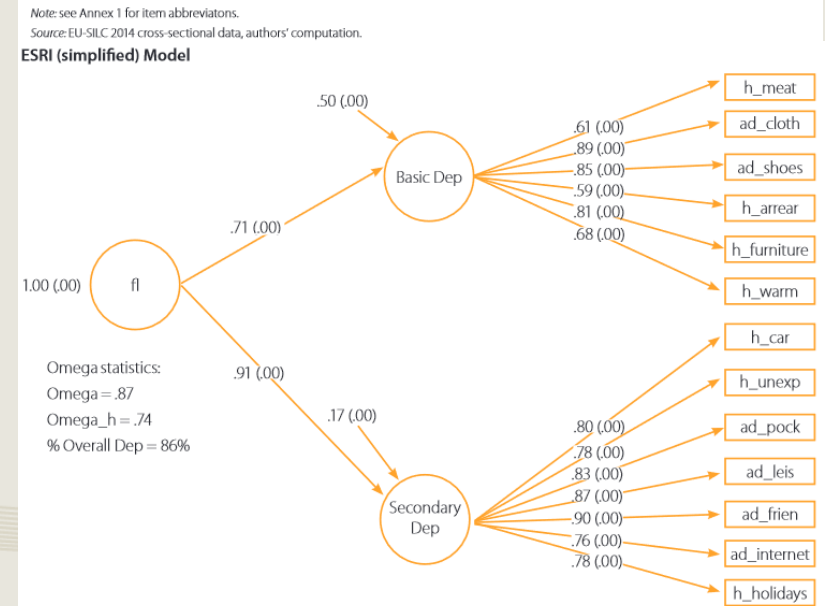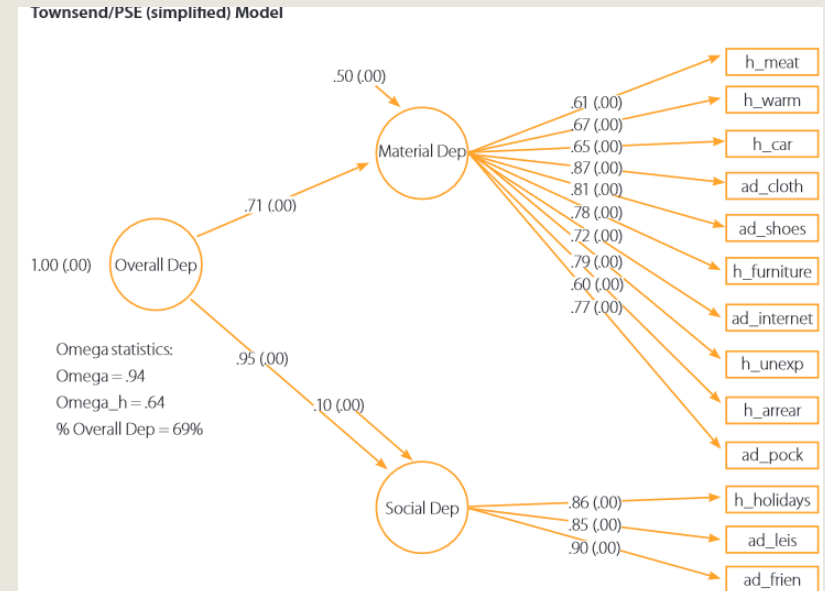Same data/indicators. Different models.

Same number of indicators

Different dimensional structure

There are several alternatives:

Different indicators, same dimensional structure
Different indicators, different dimensional structure
Different weights (?)

Figure 7: Path diagram. Unidimensional model, EU-28, 2014

Structural equation modeling (SEM) is the twenty-first century method for the scrutiny of such models.

- Global fit statistics (RMSEA, TLI, CFI, χ-SQUARE)
- The dimensional structure is an adequate representation of the data: TLI or CFI > .95/.90; RMSEA<.06
- The null -one-dimensional- model is better than the multidimensional one: TLI_1 vs TLI_2
-
- Specific statistics for hypothesis testing: λ's,
- Cargas iguales: λ_1= λ_2
- Value of the element signal –It is more than 50% of the variance explained by the latent construct–: $\lambda^2 > .5$
-

Confirmatory Factor Analysis for Applied Research

SECOND EDITION

Timothy A. Brown
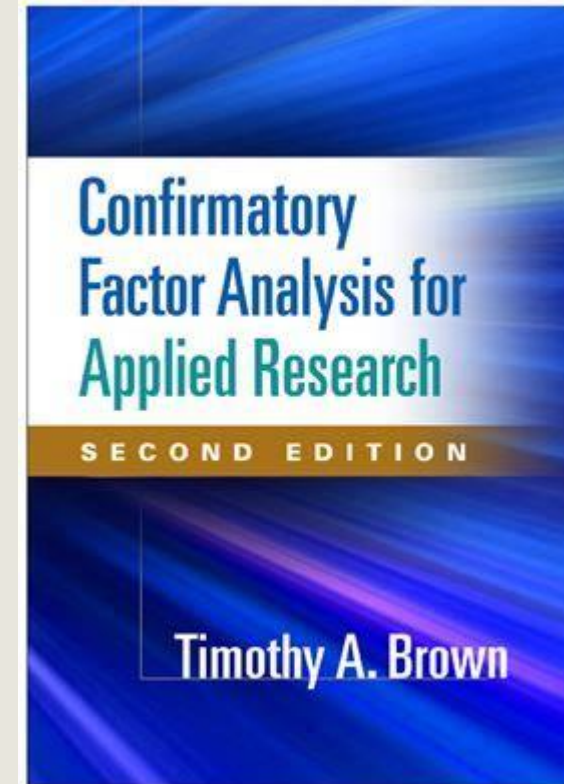
Examples: Latent construct/model validity

European model. Relative deprivation (Guio et al., 2017)

Table 14: Maximum Likelihood global statistics of fit. Model 1-4, 2014

Omega statistics using measurement models			
	Omega	Omega_h	BIC
Unidimensional	0.95	0.95	4613267
Townsend	0.94	0.64	4708170
Empirical	0.91	0.71	4717885
ESRI Model	0.87	0.74	4711749

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Ciudad de Buenos Aires model. Relative deprivation (Beccaria et al., (Forthcoming)

Confirmatory Factor Model. Adults. (TLI=.98; CFI= .98; RMSEA<.005)
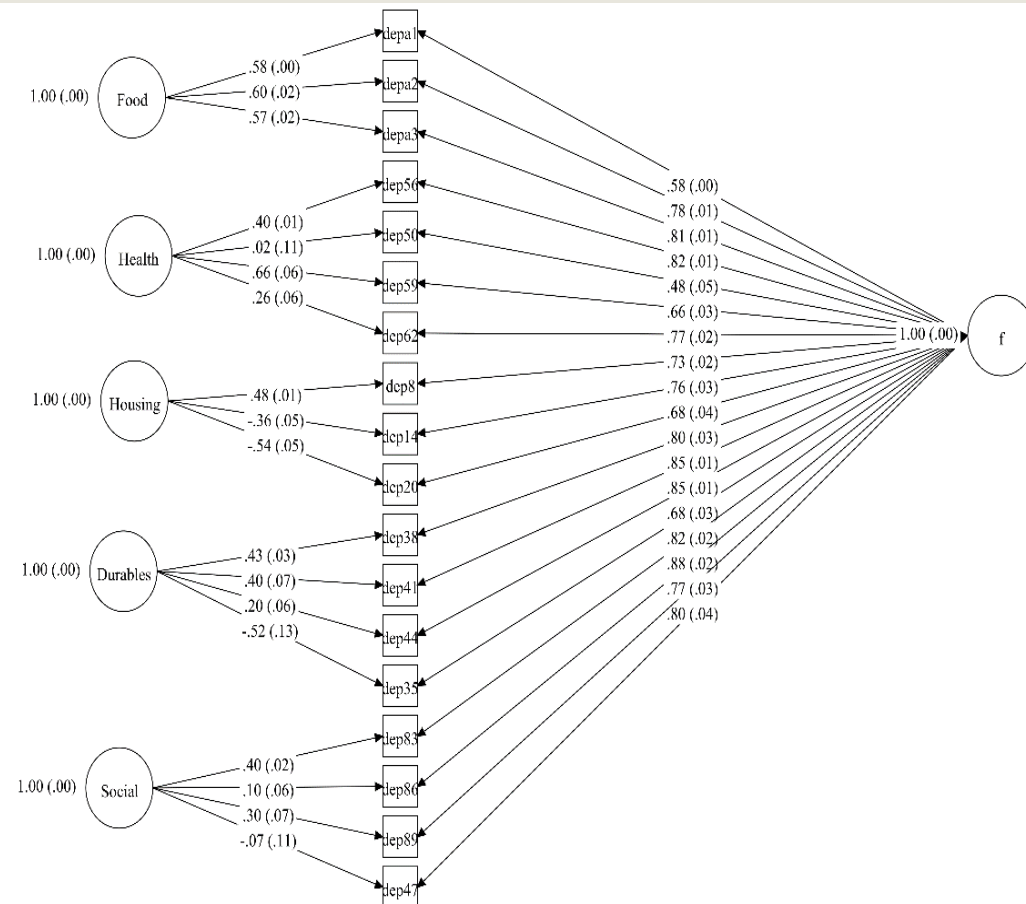
MPI-LA. Najera & Gordon (2019)

Table 3. CFA analysis. MPI-LA with weights and free weights

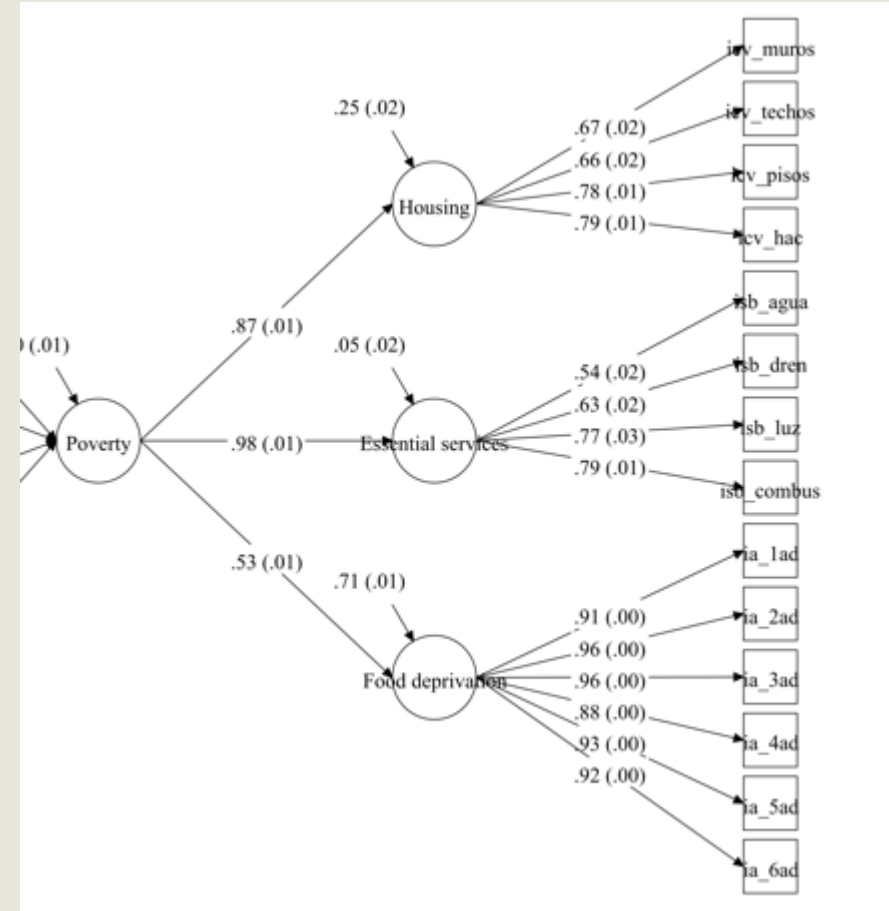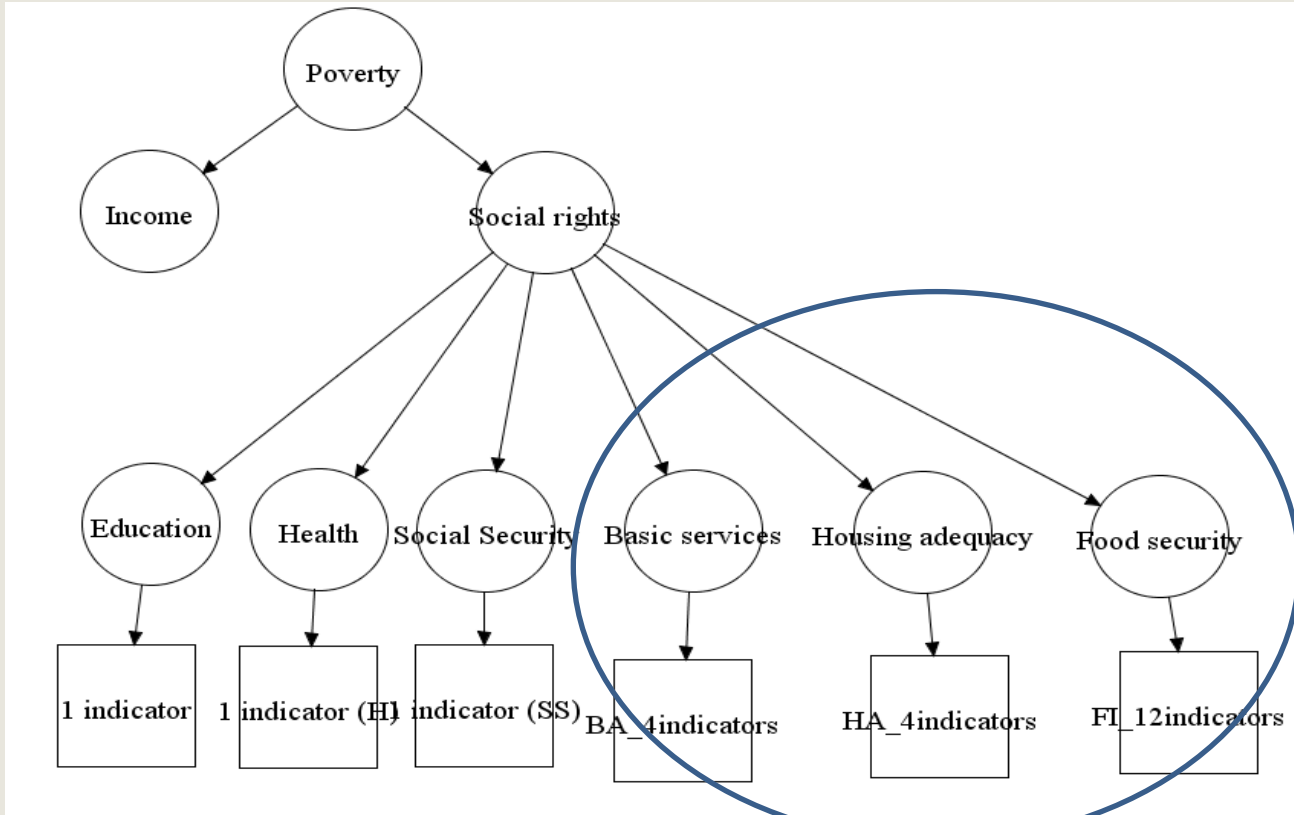| Country	First year			Second year		
	CFI	TLI	RMSEA	CFI	TLI	RMSEA
MPI-LA. Baseline model. Unidimensional model						
Argentina	0.68	0.61	0.02	0.81	0.76	0.01
Bolivia	0.18	0	0.06	0.68	0.62	0.07
Brazil	0.44	0.32	0.04	0.27	0.12	0.04
Chile	0.64	0.57	0.03	0.53	0.43	0.02
Mexico	0.7	0.64	0.04	0.77	0.73	0.05
Uruguay	0.88	0.86	0.03	0.78	0.74	0.03
MPI-LA free weights						
Argentina	0.52	0.36	0.03	0.63	0.51	0.02
Bolivia	0.46	0.27	0.07	0.69	0.6	0.07
Brazil	0.51	0.42	0.04	0.46	0.3	0.04
Chile	0.65	0.54	0.01	0.68	0.58	0.01
Mexico	0.7	0.6	0.05	0.79	0.73	0.05
Uruguay	0.91	0.88	0.02	0.82	0.77	0.02
MPI-LA (fixed weights)						
Argentina	0.4	0.24	0.03	0.59	0.48	0.02
Bolivia	0.1	na	0.08	0.58	0.53	0.08
Brazil	0.54	0.43	0.04	0.32	0.23	0.04
Chile	0.59	0.49	0.01	0.63	0.54	0.01
Mexico	0.68	0.61	0.05	0.8	0.75	0.05
Uruguay	0.9	0.88	0.02	0.82	0.77	0.02

Table 3. Confirmatory Factor Analysis. Model Fit statistics

Country	Statistic	H1 Fixed item loadings	H2 Fixed dimension loadings and free item loadings	H3 Full free weights	H4 Unidimensional
Uganda	TLI	0.73	0.94	0.85	0.94
	RMSEA	0.08	0.04	0.06	0.04
Benin	TLI	0.78	0.83	0.68	0.97
	RMSEA	0.06	0.06	0.08	0.02
Cameroon	TLI	0.79	0.91	0.92	0.94
	RMSEA	0.08	0.05	0.05	0.04
Congo	TLI	0.67	0.88	0.75	0.92
	RMSEA	0.06	0.03	0.05	0.03
Congo DR	TLI	0.39	0.96	0.88	0.97
	RMSEA	0.10	0.03	0.04	0.02
Ghana	TLI	0.63	0.82	0.72	0.88
	RMSEA	0.06	0.04	0.05	0.03
India	TLI	0.92	0.89	0.85	0.97
	RMSEA	0.04	0.04	0.05	0.02
Pakistan	TLI	0.89	0.94	0.92	0.97
	RMSEA	0.04	0.03	0.04	0.02
Nigeria	TLI	0.50	0.79	0.79	0.91
	RMSEA	0.06	0.04	0.04	0.03
Guinea	TLI	0.73	0.83	0.69	0.94
	RMSEA	0.07	0.05	0.07	0.03

CONEVAL 2018. Partial model –identified.
TLI>.95 (But two dimensions are a bit noisy)
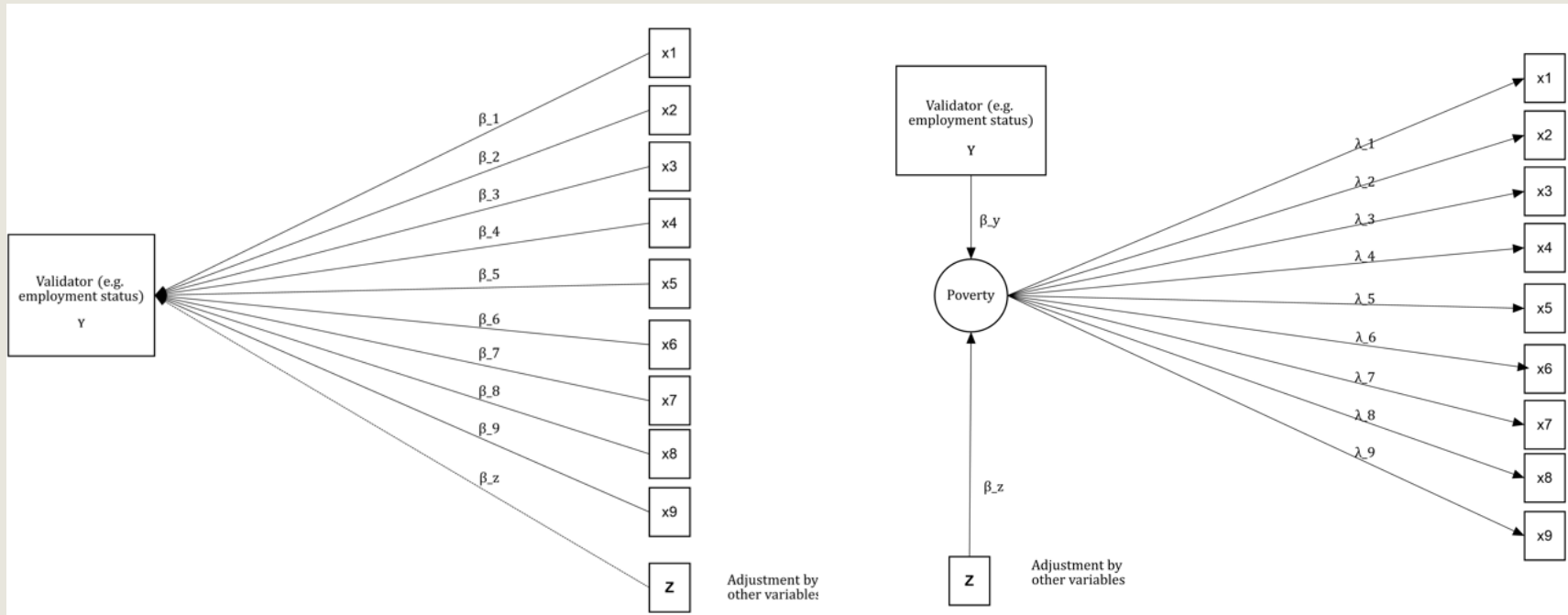
Item validation

Latent scores validation
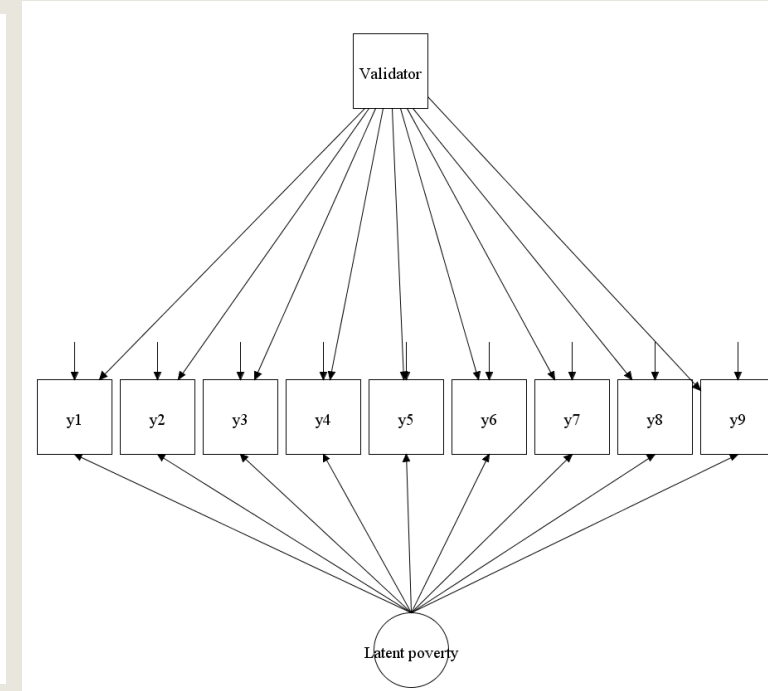
Latent scores and item validation

This is a Multiple indicators and multiple causes (MIMIC) model of a reduced version of the multidimensional Mexican measure.

The model shows that latent poverty is associated by possession of different goods and education attainment of the household head, adjusted by rurality and household size.

Standardised coefficients (Standard error within brackets)

- Modern theory defines construct validity as the umbrella of validity.
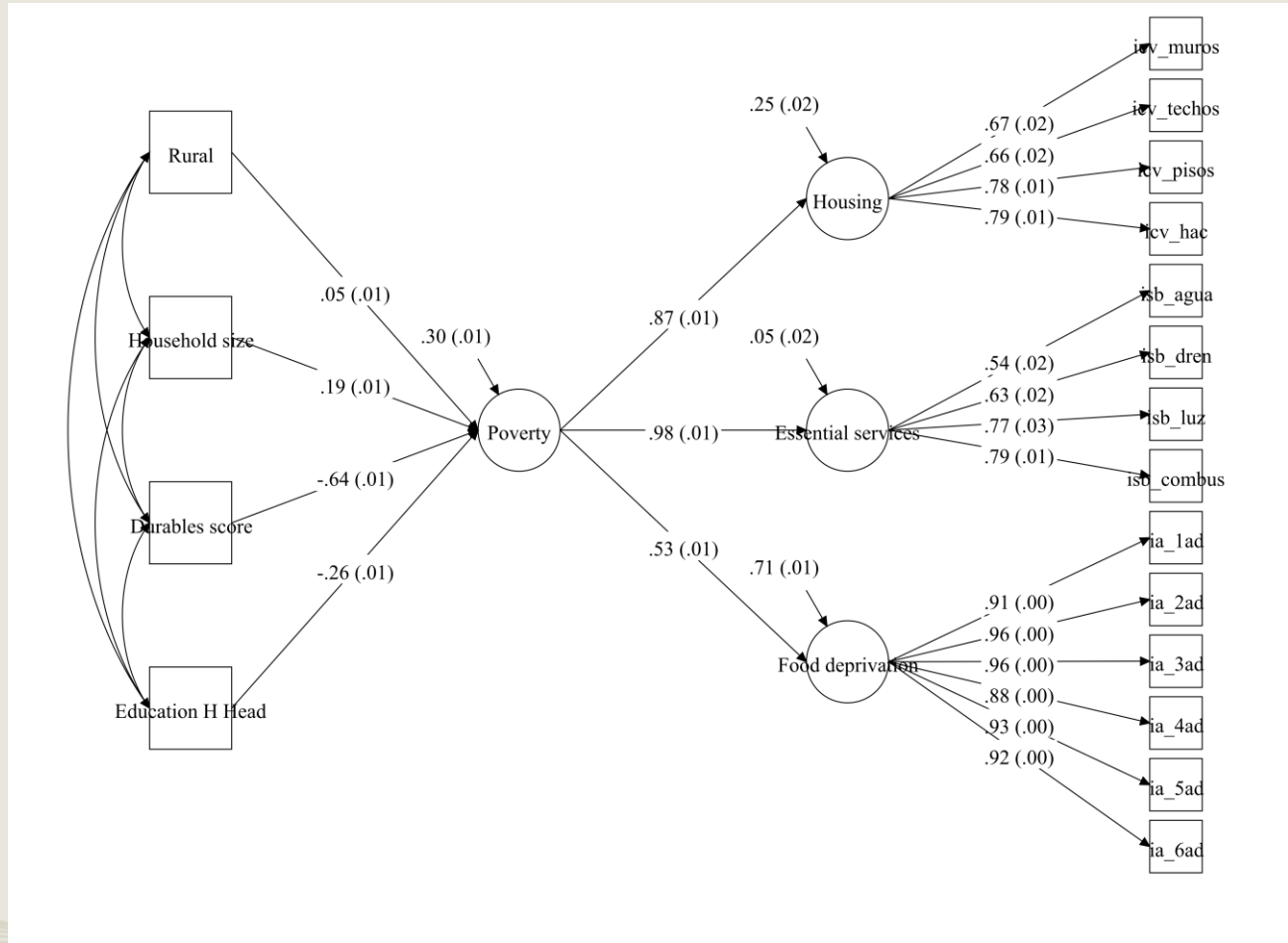
It is the accumulation of the different types of evidence and justifies that the scores can be interpreted as we want to do.

Construct validity / Validez de constructo

Content validity validity

Face validity

Criterion validity
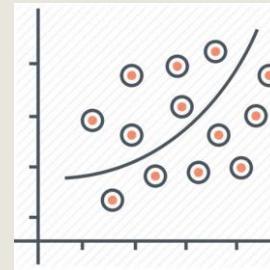
Model/latent construct
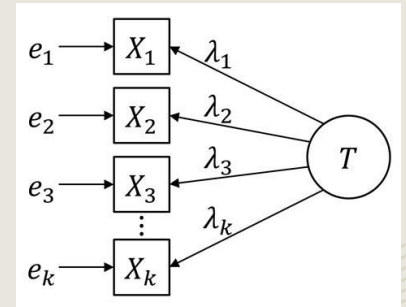
- Every measurement activity has both a measurement instrument and an underlying measurement model

- Measurement models emerge from the series of assumptions we raise to represent poverty via observed data

- Both reliability and validity are necessary for poverty indices with low measurement error

- Measurement error impacts classification error and is propagated across all inference (Crosstabulations to advance regression models)

- Both reliability and validity are properties of the scores –not the scales themselves- and are relative to the context in which the measurement is implemented

Gracias

- Dr. Héctor Nájera
- Investigador Asociado – C PUED-UNAM

AERA, APA and NCME. 2014. "Standards for Educational and Psychological Testing." Edited by American Educational Research (AERA) and American Psychological Association (APA) and National Council on Measurement in Education (NCME) and Joint Committee on Standards for Educational and Psychological Testing (US). Amer Educational Research Assn.

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.

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

Cronbach, Lee J., and Paul E. Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52 (4): 281.

Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16: 297–334.

Decancq, Koen, and María Ana Lugo. 2013. "Weights in Multidimensional Indices of Wellbeing: An Overview." *Econometric Reviews* 32 (1): 7–34. https://doi.org/10.1080/07474938.2012.690641.

Gibbons, Robert D., Jason C. Immekus, R. Darrell Bock, and Robert D. Gibbons. 2007. "The Added Value of Multidimensional Irt Models." *Multidimensional and Hierarchical Modeling Monograph* 1.

Gordon, D. 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.

———. 2018. "Measuring Poverty in the Uk." In *Poverty and Social Exclusion in the Uk*, edited by E Dermott and Gill Main.

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, 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.

Guttman, Louis. 1945. "A Basis for Analyzing Test-Retest Reliability." *Psychometrika* 10 (4): 255–82. https://doi.org/10.1007/BF02288892.

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.

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

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

Messick, Samuel. 1987. "Validity." *ETS Research Report Series* 1987 (2): i–208.

Muthén, Beng. 2013. "IRT in Mplus." Mplus. http://www.statmodel.com/download/MplusIRT2.pdf.

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, 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.

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.

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.

Spearman, C. 1904. "The Proof and Measurement of Association Between Two Things." *Américan 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.

Streiner, David L., Geoffrey R. Norman, and John Cairney. 2015. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, USA.

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

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.