



Programa  
Universitario  
de Estudios  
del **Desarrollo**  
UNAM

# Validity / Validez

Héctor Nájera

Curso internacional de teorías y métodos contemporáneos para la  
medición de la pobreza multidimensional

29 Nov 2021



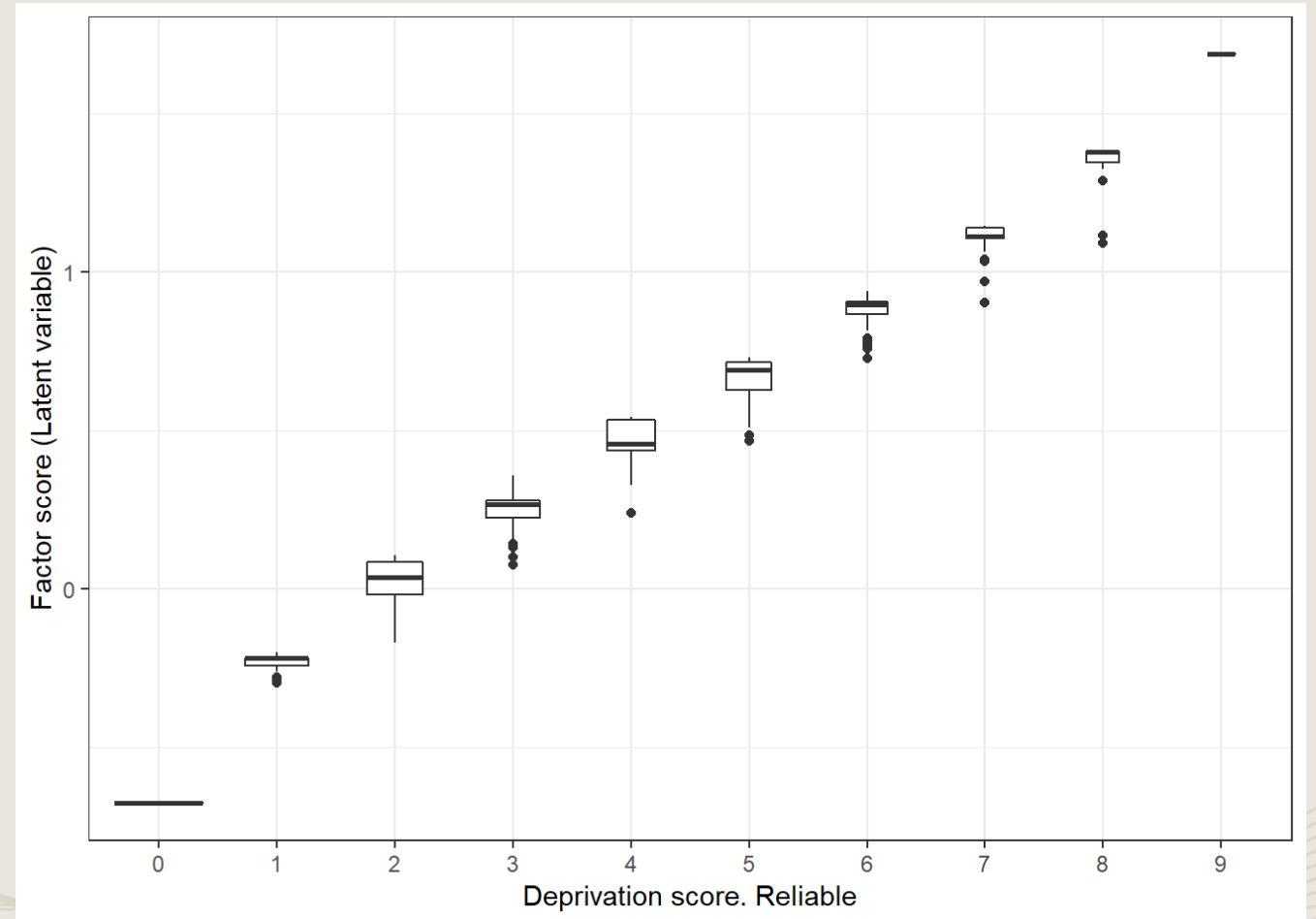
# Validez, evidencia y scores latentes

¿Hay alguna evidencia que justifique mi interpretación de los puntajes como puntajes de pobreza / privación?

¿Sobre qué bases puedo concluir que 9 significa más pobre que 8, ... , 0?

Los modelos de medición suponen que las puntuaciones permiten hacer tales inferencias.

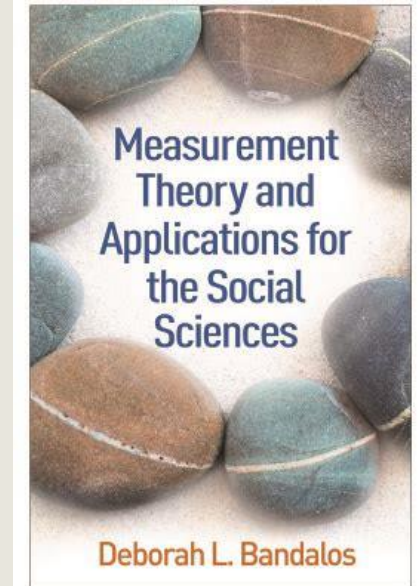
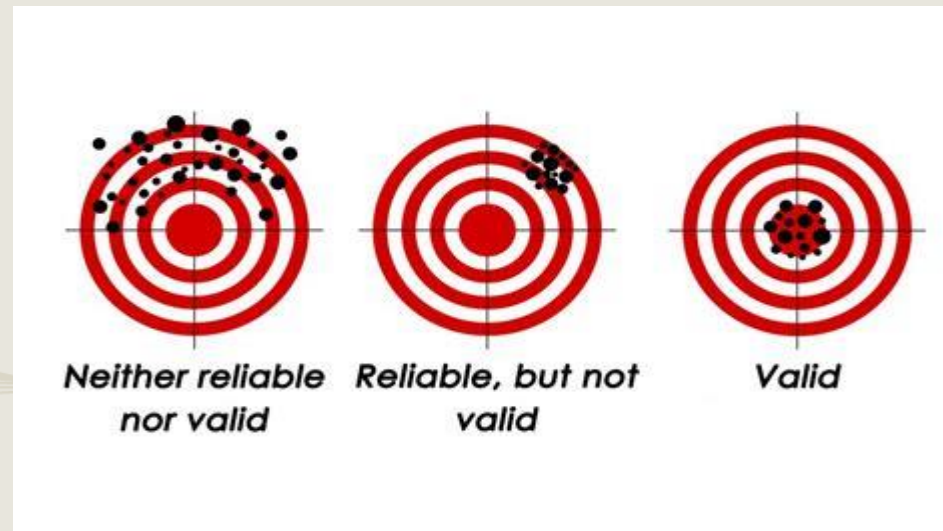
Los modelos de medición exigen evidencia para tales interpretaciones



# Validez: Concepción moderna

- Es toda la evidencia que:
  - Justifica las interpretaciones que hacemos de un índice
  - Sostiene la interpretación de los scores de un índice
  - Detecta error sistemático

Standards for Educational and Psychological measurement (APA, AERA)





# Validez, hipótesis y tipos de evidencia

- La validez es un conjunto de evidencias respecto a una serie de hipótesis del modelo de medición.
  - Los indicadores capturan tienen la misma fuente (Confiabilidad) y dicha fuente es (Pobreza)
  - La estructura del modelo de medición es una adecuada representación de los datos (D, I)
  - Las dimensiones tienen sentido teórico y empírico – existen-
  - Los indicadores representan el fenómeno (pobreza) de quienes la viven (población pobre)
  - La línea de pobreza me lleva a la mejor separación posible de acuerdo al modelo de medición –Townsend breaking point-
  - Los indicadores de la medida A apuntan en la misma dirección que los de la medida B



Producción de evidencia/validación

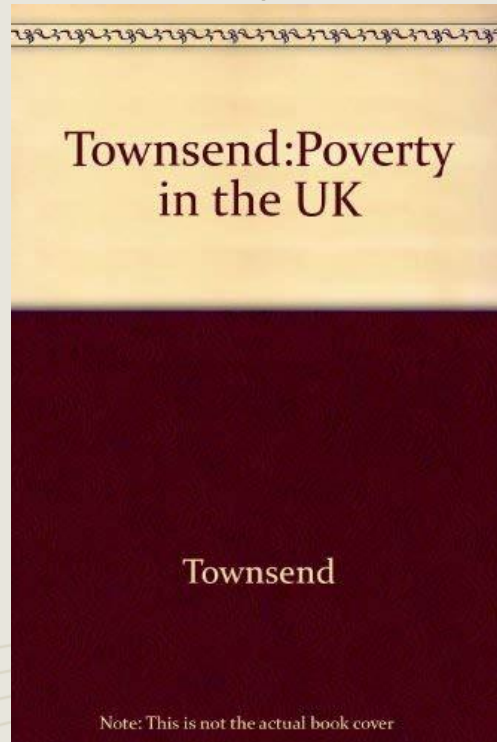
*Criterio, contenido y  
constructo / modelo latente*

*Cara, criterio y concurrente*



# Content validity: Validez de contenido

- El grado en el que la teoría de pobreza ofrece una serie de enunciados lógicos -no contradictorios-, consistentes y testeables
- El grado en el que la teoría de pobreza distingue entre sus causas y sus posibles consecuencias, de manera tal que existe claridad respecto al modelo de medición y el modelo explicativo
- El grado en el que el modelo de medición es *a priori* una representación razonable de la pobreza



Desigualdades estructurales >  
Sistemas distribución de recursos >  
Control de recursos en el tiempo >  
Carencias -pobreza relativa &  
modelo de medición-



# Face validity: Validez de cara

- El grado en el que la población reconoce que el modelo de medición y sus contenidos es una buena representación de su situación
  - Población general
  - Paneles
  
- Métodos:
  - Entrevistas a profundidad
  - Grupos focales
  - Encuestas representativas, e.g. identificación de necesidades socialmente percibidas o FIES scale
  - Legislación/Derechos sociales, e.g. la ley representa la voluntad de la población



**SAGE researchmethods** Browse | Research Tools Search

---

**Face Validity** 

In: [Encyclopedia of Research Design](#)

 ENCYCLOPEDIA

**Edited by:** Neil J. Salkind

**Published:** 2010

**DOI:** <https://dx.doi.org/10.4135/9781412961288.n147>



# Validez de criterio

H: Los indicadores capturan tienen la misma fuente (Confiabilidad) y dicha fuente es (Pobreza)

Si los indicadores capturan pobreza, éstos deberían asociarse a causas de la pobreza: Crisis económica, discriminación, evaluación del nivel de vida, evaluación del estado de salud, posición en el mercado de trabajo.

La teoría de pobreza debe delinear los términos de las hipótesis

Correlación bi-variada  
Regresión multivariada  
Modelos de ecuaciones estructurales

La clave, como en el quehacer científico, es que los resultados apunten en la misma dirección.

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

Shailen Nandy · Marco Pomati

H1: Los indicadores son manifestaciones de pobreza, es decir, tener carencia correlaciona con marcadores de pobreza

	Difficult compared to good or more or less OK (relative risk with 95 % CIs)
<b>Validator 1—evaluation of household income status</b>	
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)

	Goes into debt compared to able to save (relative risk with 95 % CIs)
<b>Validator 2—current financial situation</b>	
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)

	Bottom quintile compared to top quintile (relative risk with 95 % CIs)
<b>Validator 4—asset index quintiles</b>	
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)

Evidencia: Las personas desfavorecidas son más propensas a manifestar que situación de ingresos es difícil, endeudarse y pertenecer al quintil inferior –Índice de riqueza-

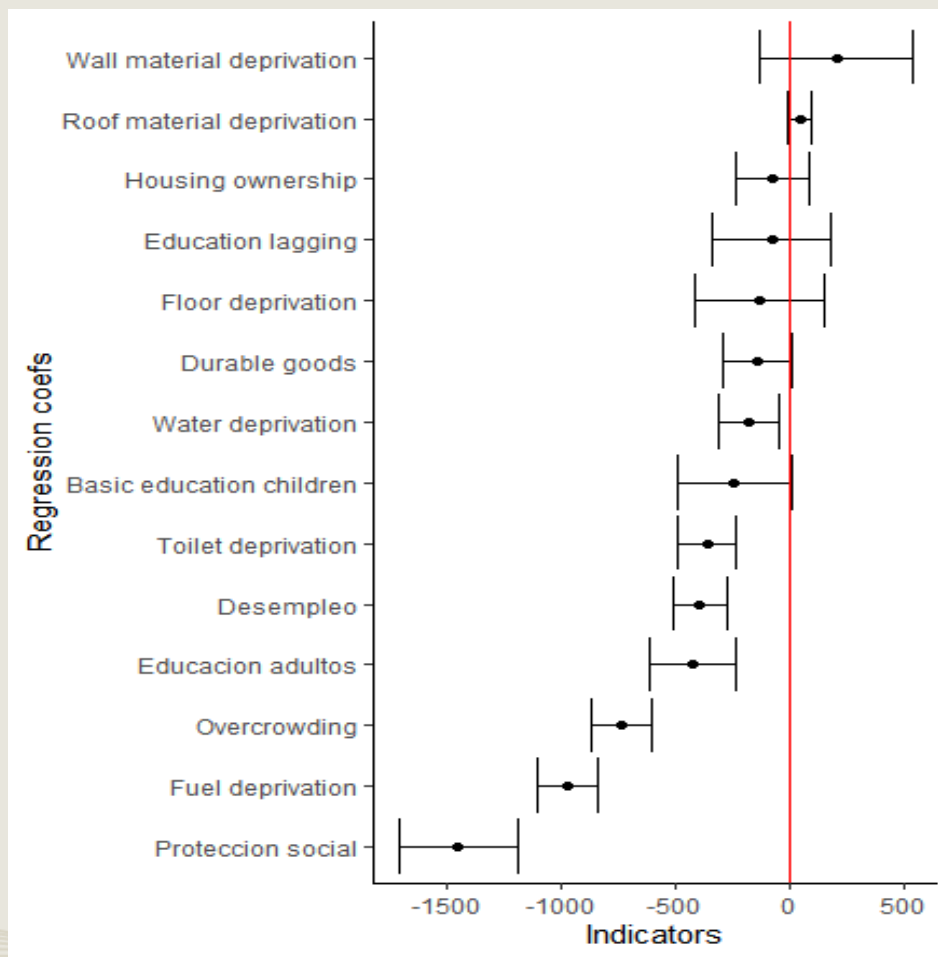




# Validez de criterio: Ejemplo

H1: Los indicadores son manifestaciones de pobreza, es decir, tener carencia se correlaciona con marcadores de pobreza

No todos se correlacionan con "ingresos". Quizás no sea el mejor validador, pero aún así...



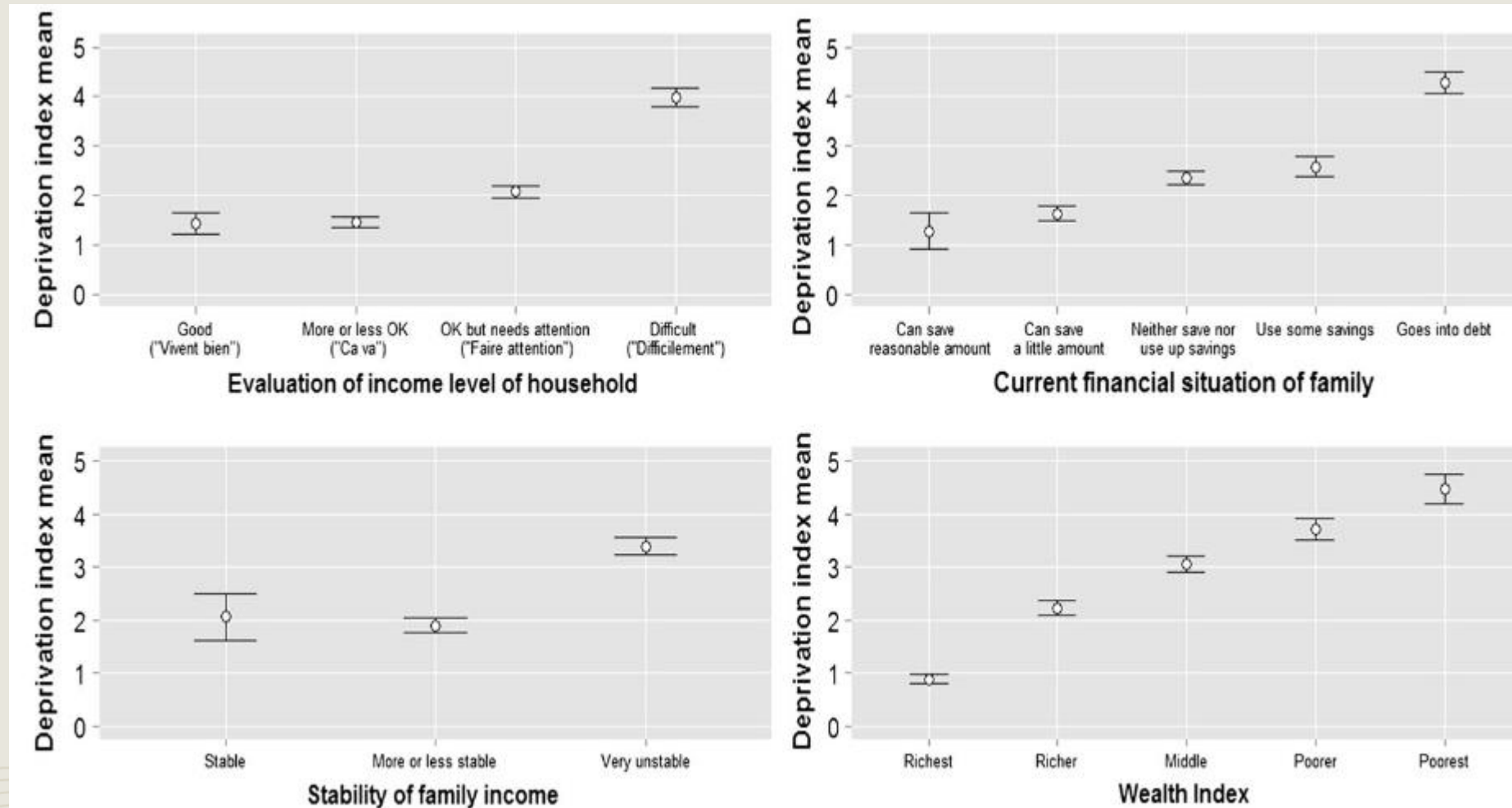
Data from the MPI-LA. Mexico 2014



# Validez de criterio: Ejemplos

H: Mayores puntajes de privación significan mayor pobreza

Evidencia: Las experiencias de las personas son peores para aquellos con puntajes de privación más altos



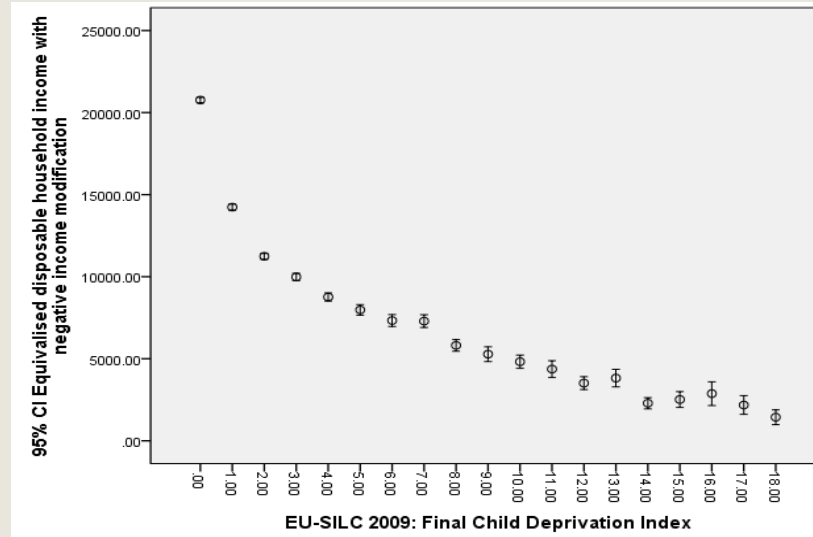
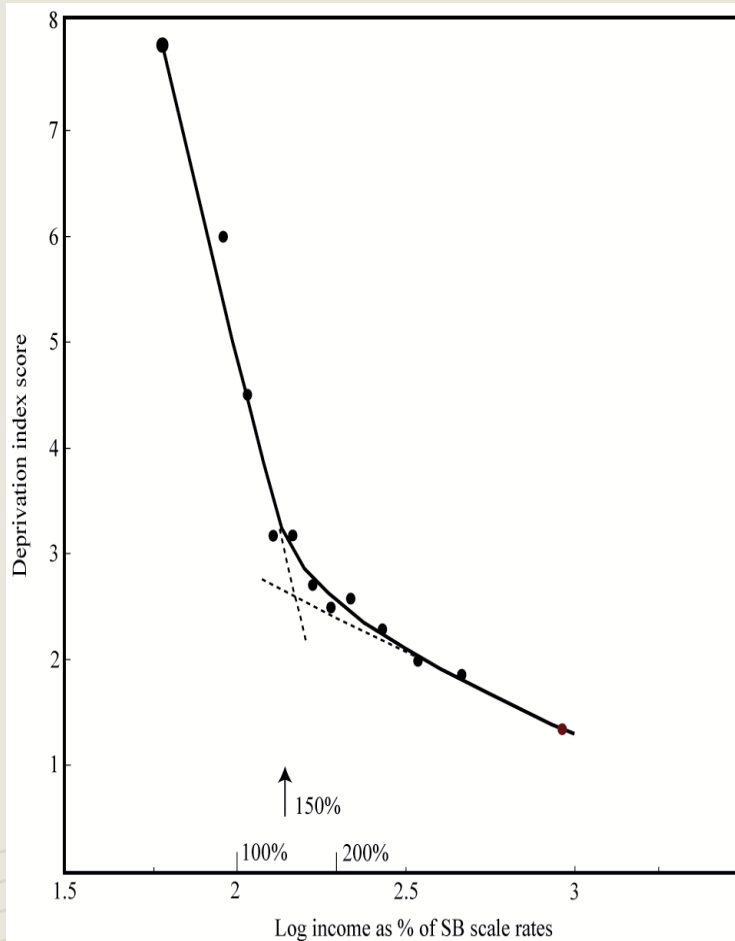
**Fig. 1** Testing scale validity. *Source:* Calculated from Benin DHS 2006



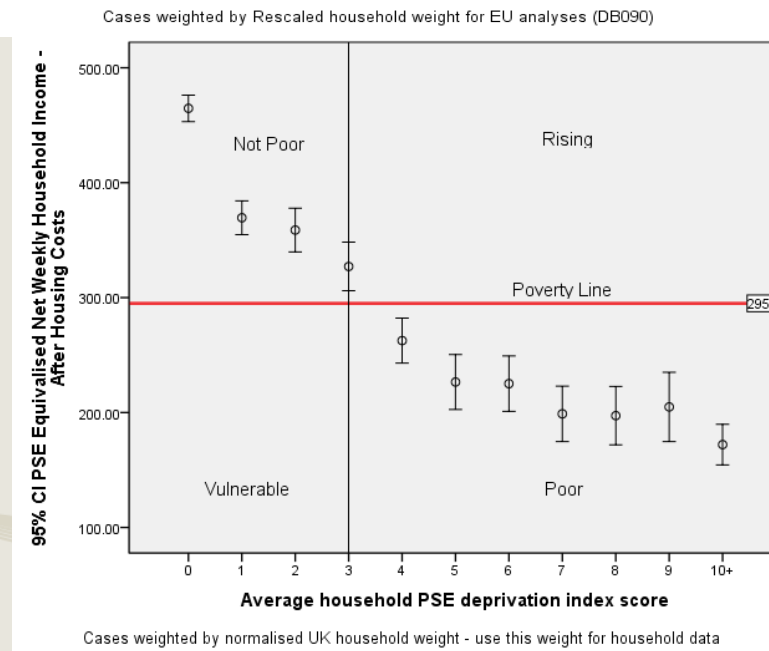


# Validez: Hipótesis específicas. Línea de pobreza

Hipótesis del punto de quiebre de Townsend: Hay un nivel de recursos a partir del cual la privación material múltiple aumenta sustancialmente -línea de pobreza-



Europe

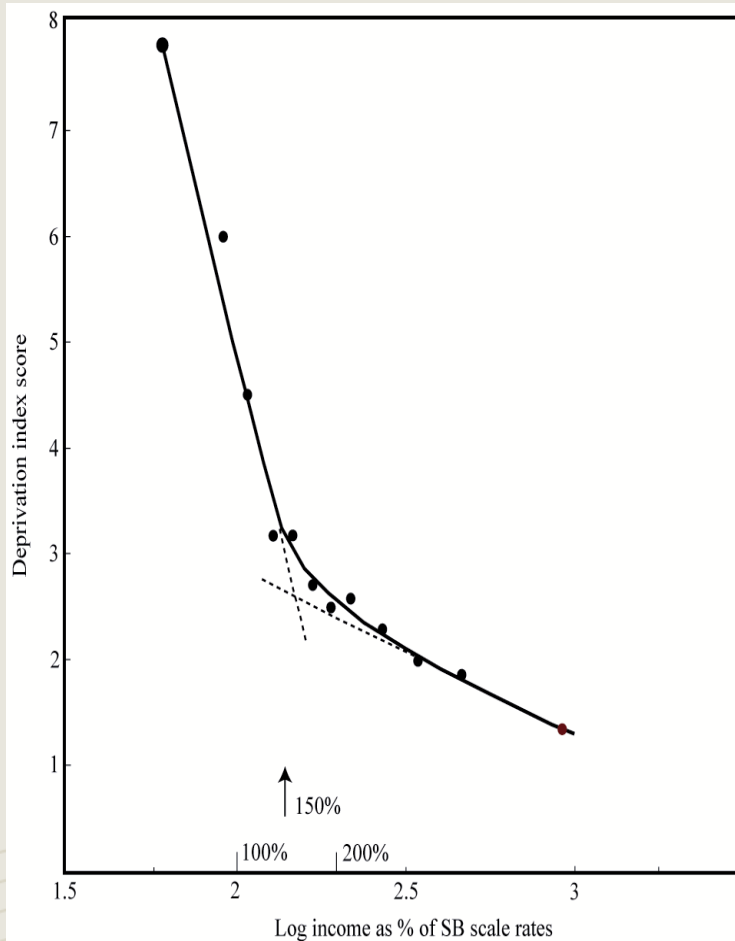


United Kingdom

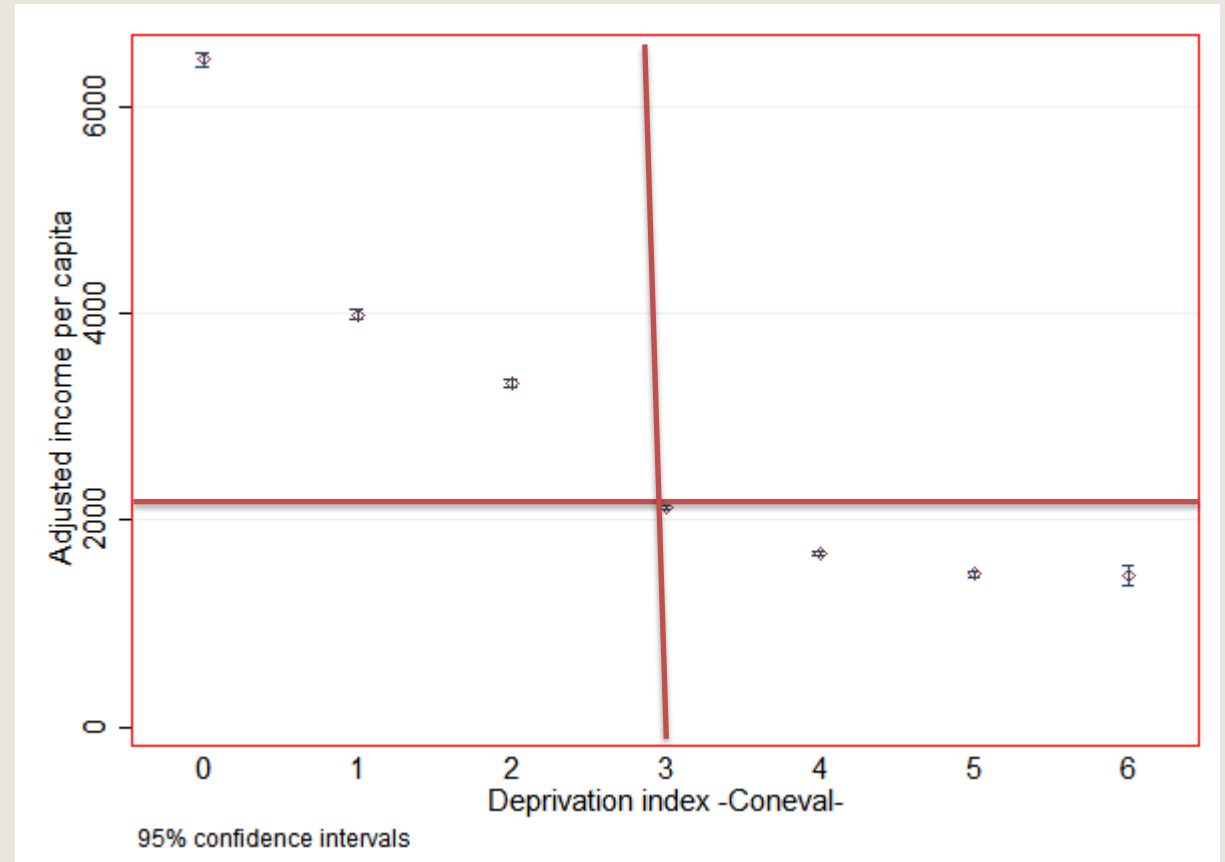


# Validity: Specific hypothesis. Poverty Line

Hipótesis del punto de quiebre de Townsend: Hay un nivel de recursos a partir del cual la privación material múltiple aumenta sustancialmente -línea de pobreza-



Mexico





# Validity: Poverty Line

---

Gordon: Modelos Logit y ANOVA -La división que maximiza el ajuste del modelo-

Finite mixture models -Cruce de curvas latentes. Cuando no hay medida de recursos-

Análisis de clases latentes: cuenta basada en modelos. Con y sin recursos medir-

Modelo de error de clasificación de Hausman -Con y sin medida de recursos-





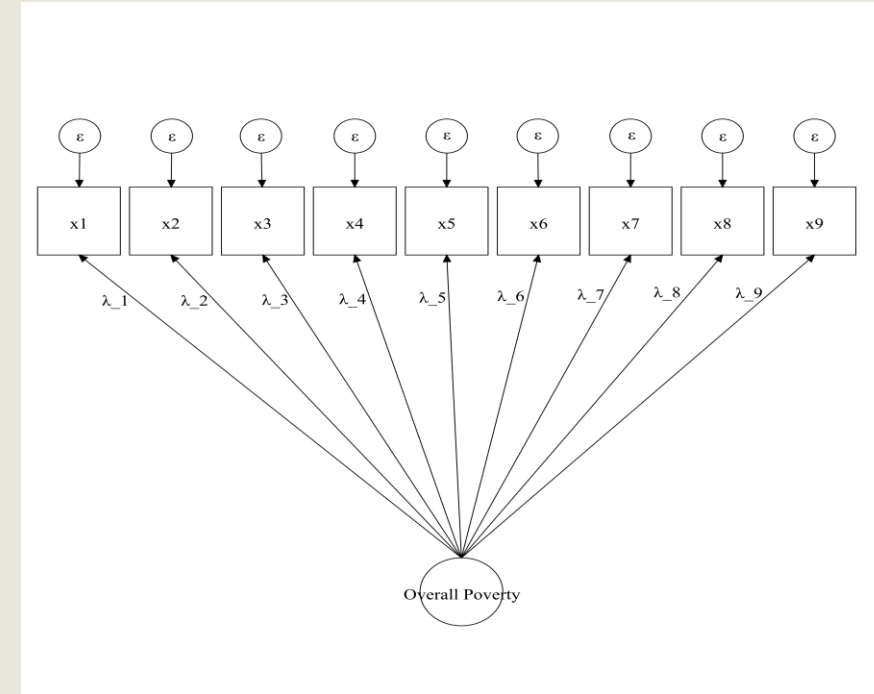
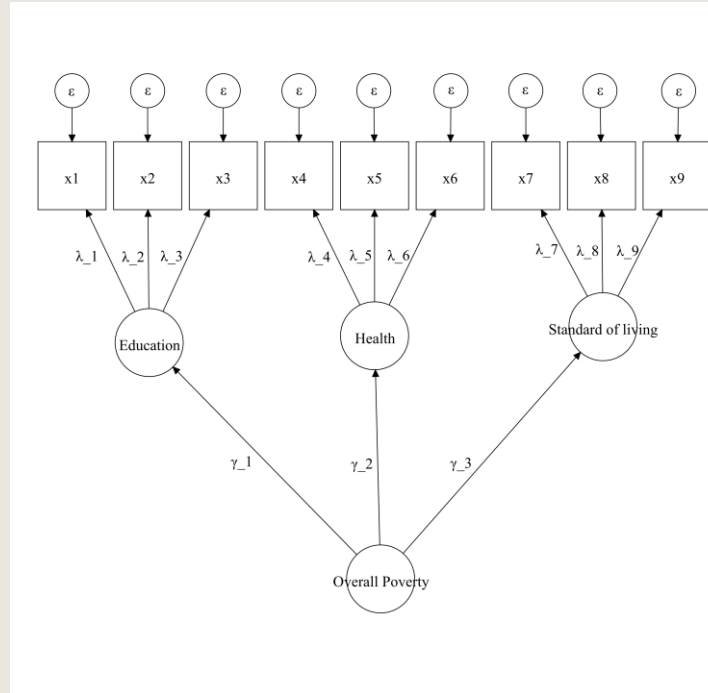
# Validez de constructo/modelo latente

Mismos datos/indicadores.  
Diferentes modelos.

Mismo número de indicadores  
Estructura dimensional  
diferente

Hay varias alternativas:

Diferentes indicadores, misma  
estructura dimensional  
Diferentes indicadores,  
diferente estructura  
dimensional  
Diferentes pesos (?)





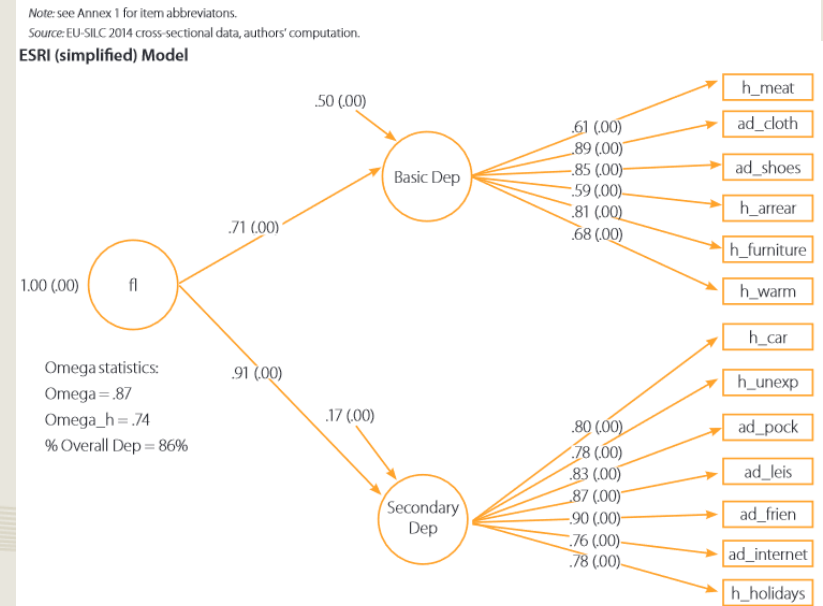
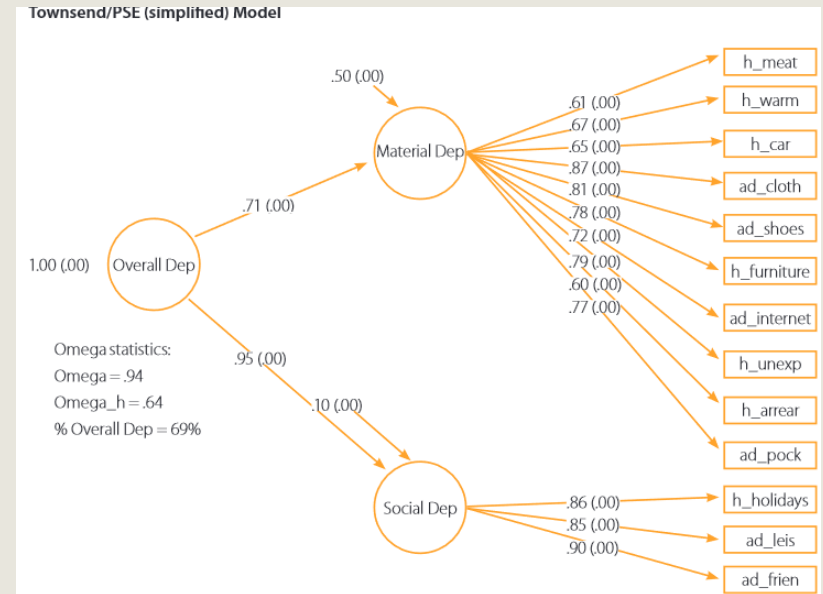
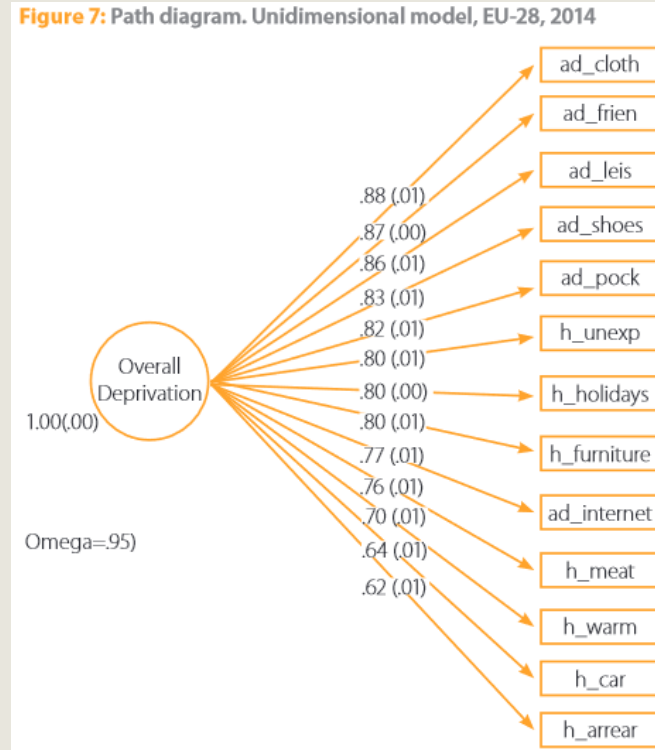
# Validez de constructo/modelo latente

Mismos datos/indicadores.  
Diferentes modelos.

Mismo número de indicadores  
Estructura dimensional diferente

Hay varias alternativas:

Diferentes indicadores, misma estructura dimensional  
Diferentes indicadores, diferente estructura dimensional  
Diferentes pesos (?)





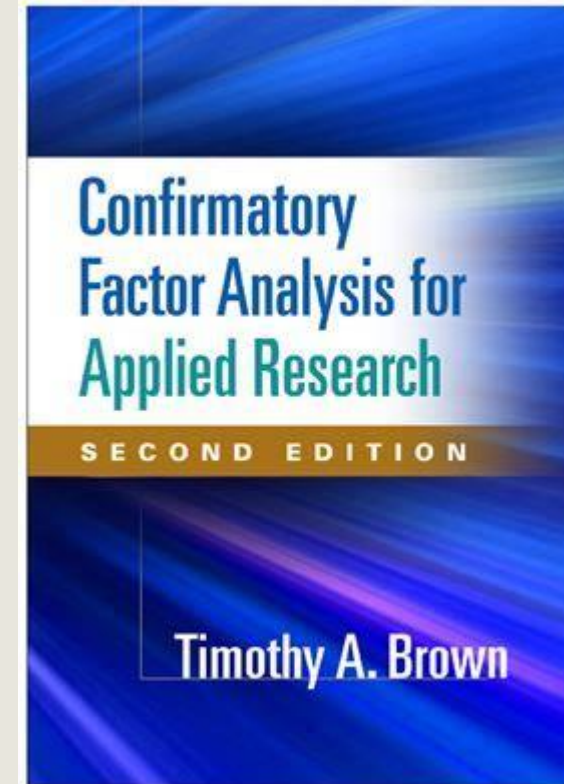
# Métodos: Validez de constructo/modelo latente

El modelado de ecuaciones estructurales (SEM) es el método del siglo XXI para el escrutinio de dichos modelos.

- Estadísticas globales de ajuste (RMSEA, TLI, CFI,  $\chi$ -SQUARE)
- La estructura dimensional es una representación adecuada de los datos: TLI o CFI > .95/.90; RMSEA < .06
- El modelo nulo -unidimensional- es mejor que el multidimensional: TLI\_1 vs TLI\_2

Estadísticas específicas para la prueba de hipótesis:  $\lambda$ 's,

- Cargas iguales:  $\lambda_1 = \lambda_2$
- Valor de la señal del elemento -Es más del 50% de la varianza explicada por el constructo latente-:  $\lambda^2 > .5$







# Ejemplos: Validez de constructo/modelo latente

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

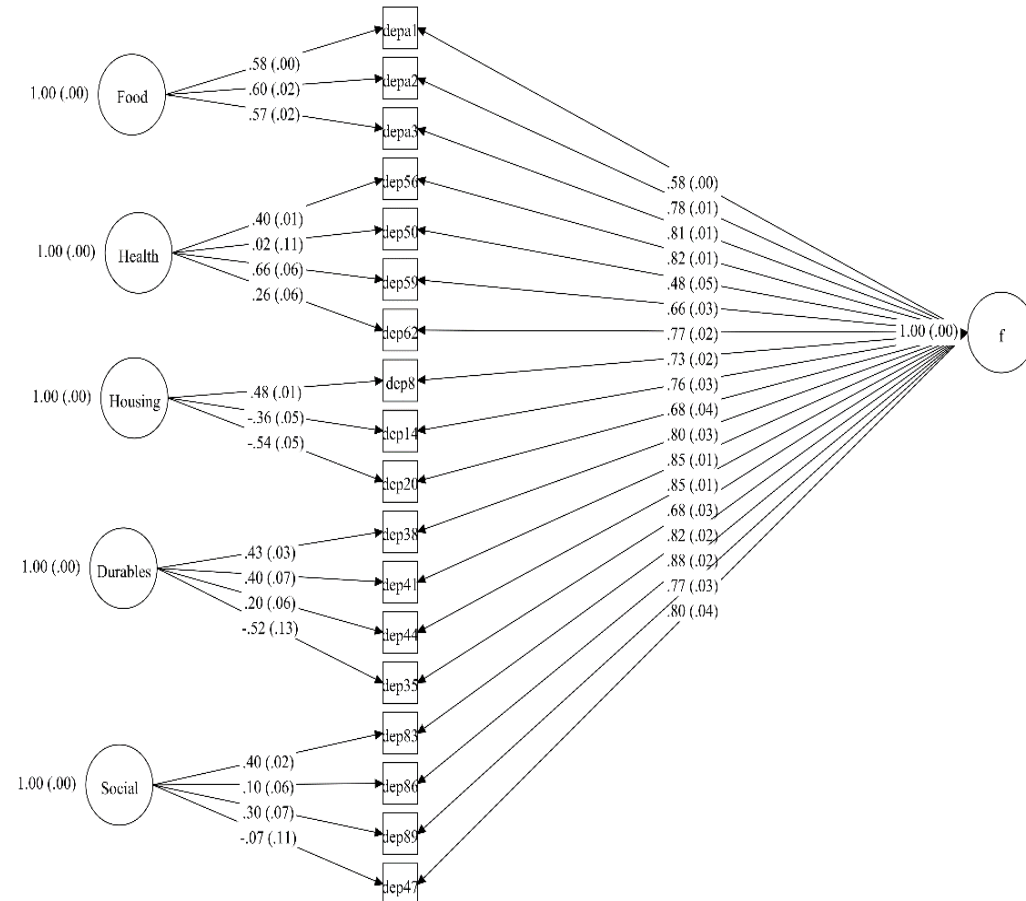
**Table 14:** Maximum Likelihood global statistics of fit. Model 1-4, 2014

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)





# Ejemplos: Validez de constructo/modelo latente

## 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
<b>MPI-LA. Baseline model. Unidimensional model</b>						
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
<b>MPI-LA free weights</b>						
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
<b>MPI-LA (fixed weights)</b>						
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

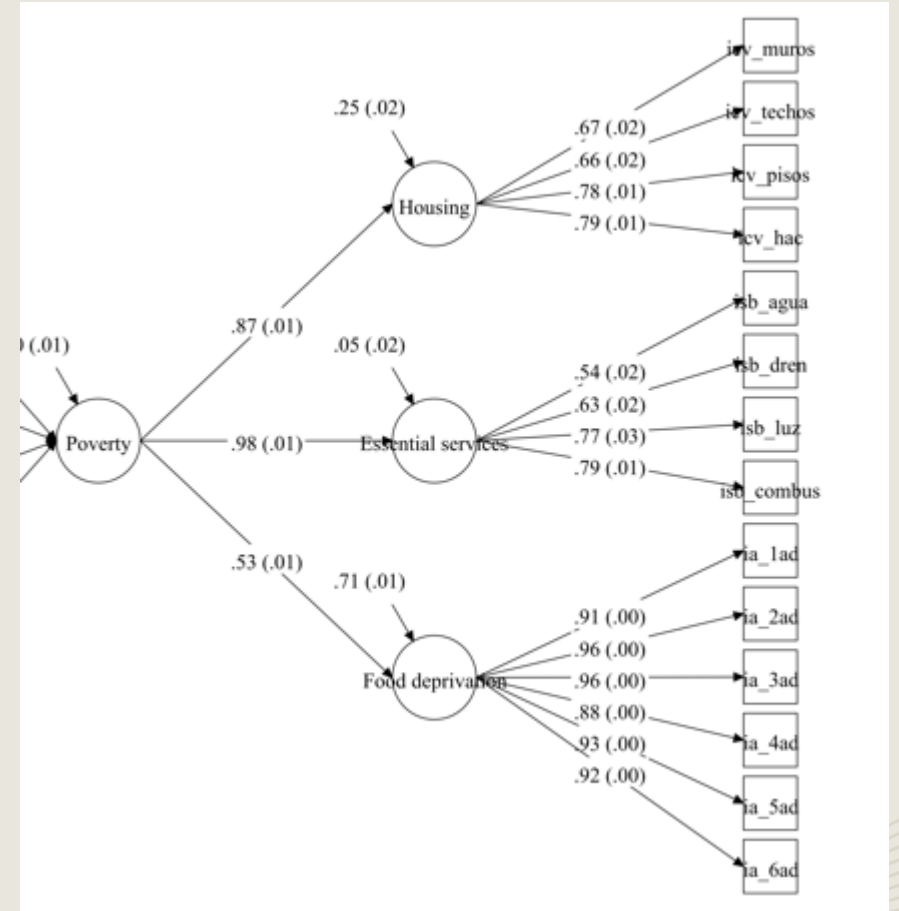
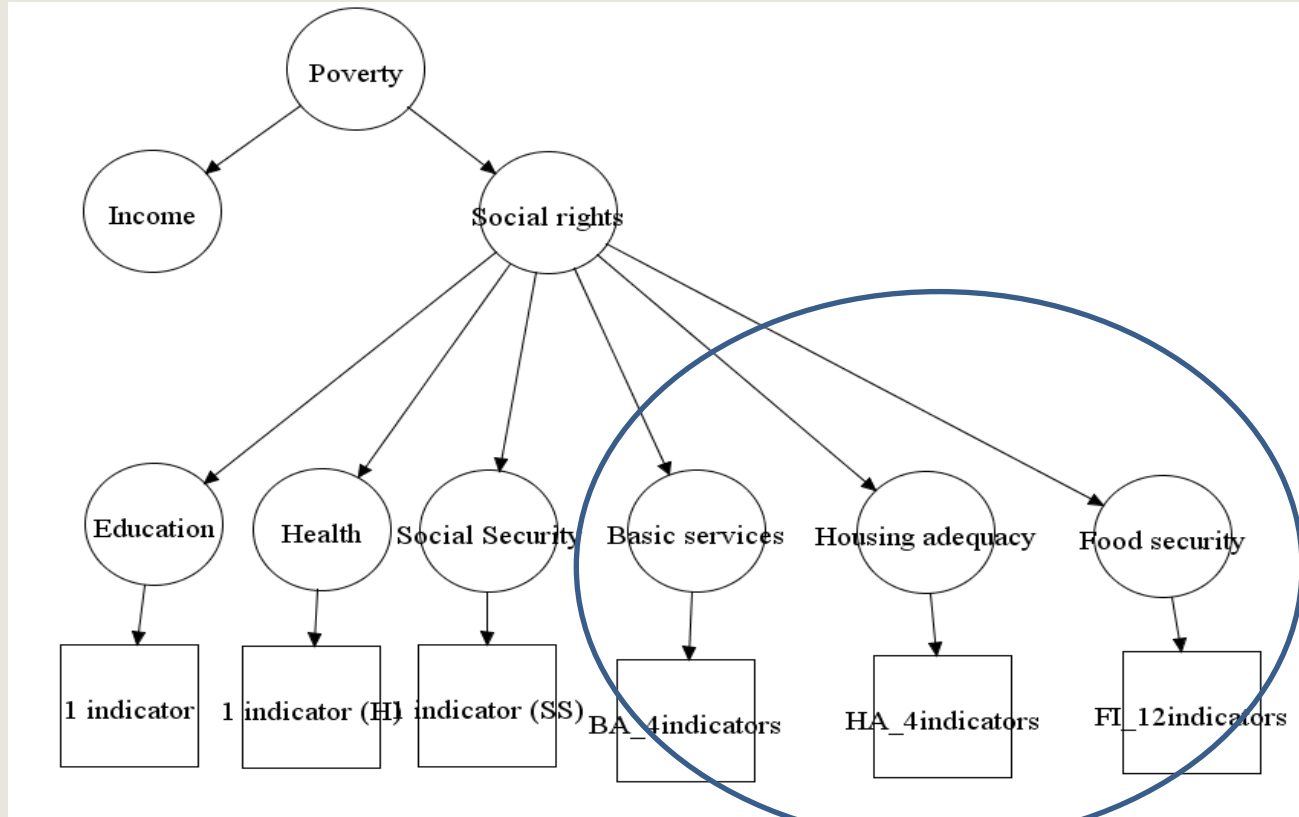
Table 3. Confirmatory Factor Analysis. Model Fit statistics

Country	Statistic	H1	H2	H3	H4
		Fixed item loadings	Fixed dimension loadings and free item loadings	Full free weights	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

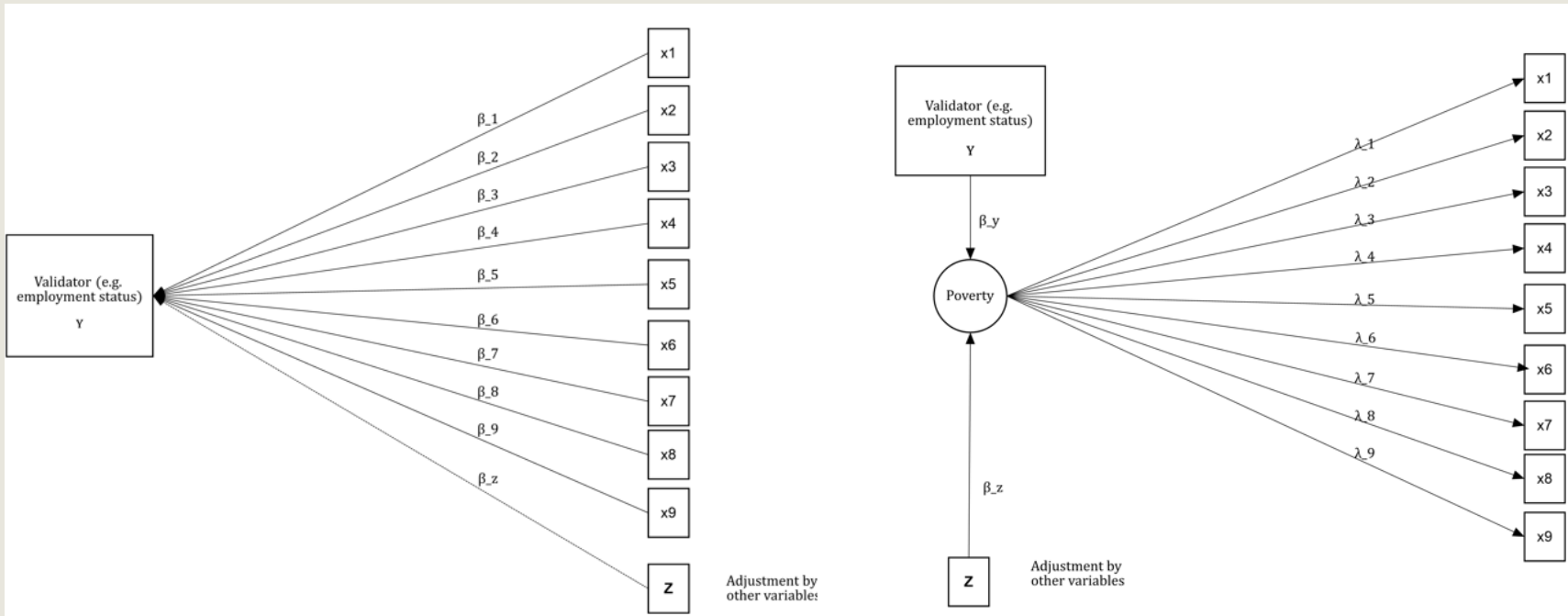


# Ejemplos: Validez de constructo/modelo latente

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

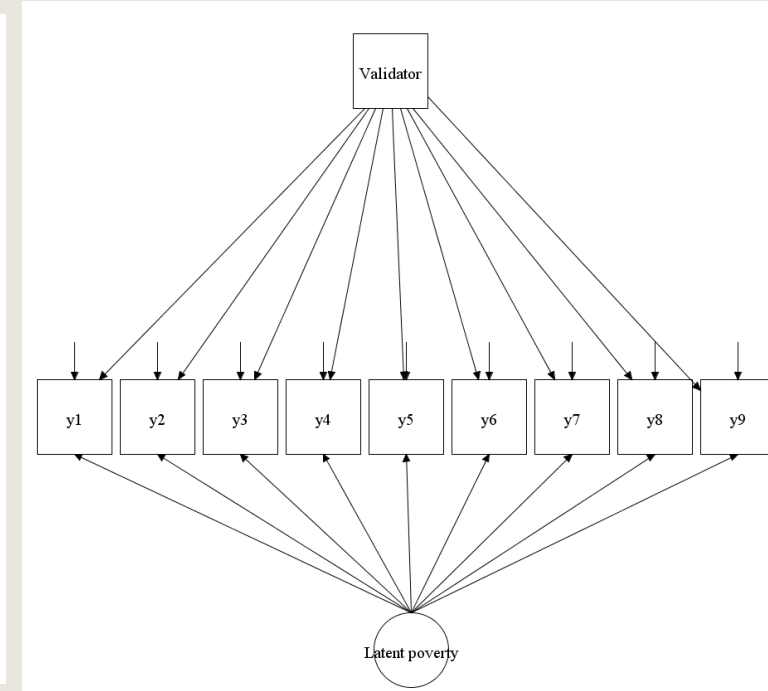


# Advanced: Latent construct + Criterion validity



Item validation

Latent scores validation



Latent scores and item validation

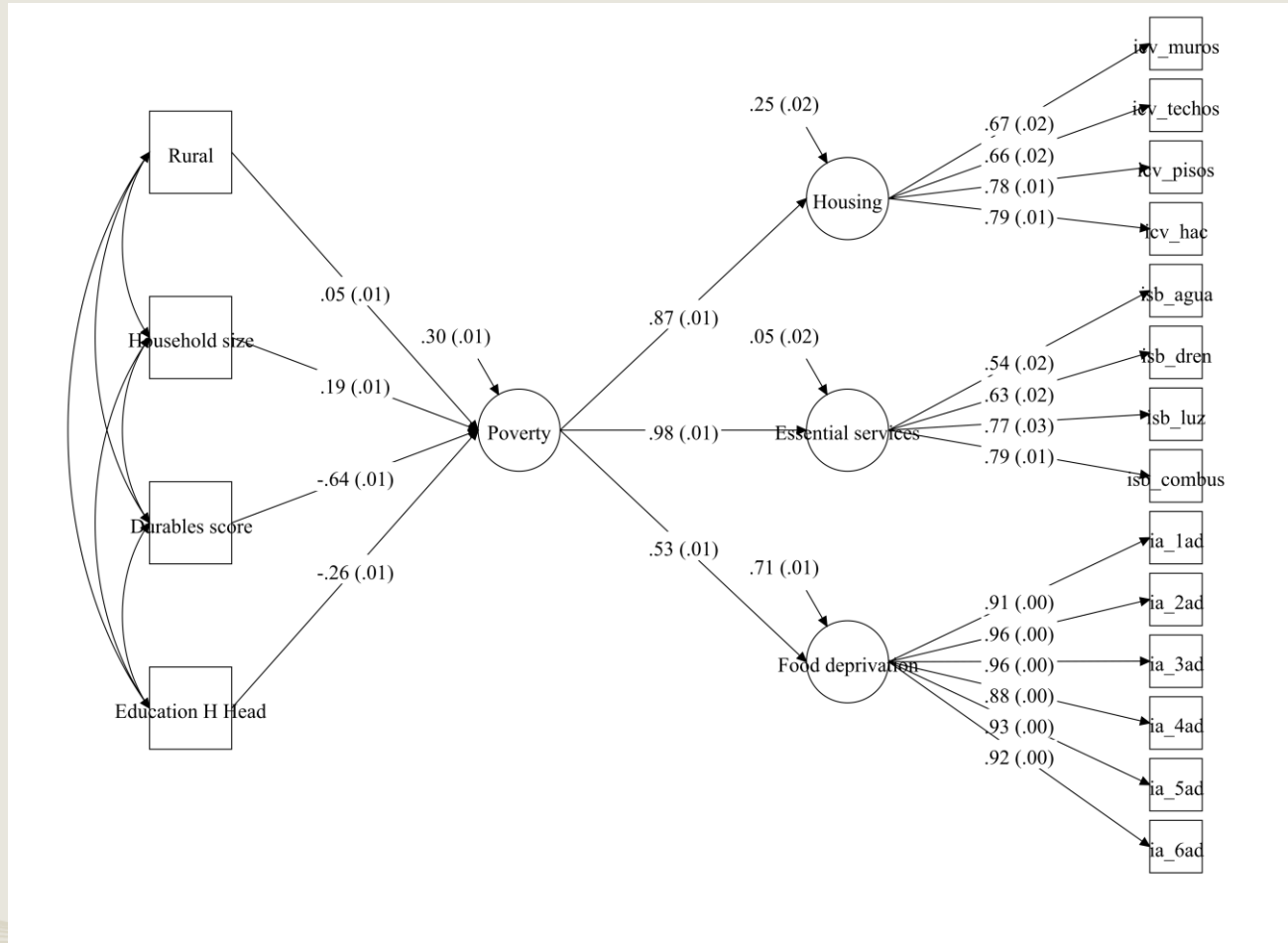


# Advanced: Latent construct + Criterion validity

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)





# Construct validity – Validez de constructo

- La teoría moderna define a la validez de constructo como el paraguas de la validez
- Es la acumulación de los distintos tipos de evidencia y justifica que los scores puedan interpretarse como queremos hacerlo.

## Construct validity / Validez de constructo

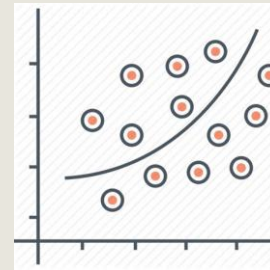
Content validity  
validity



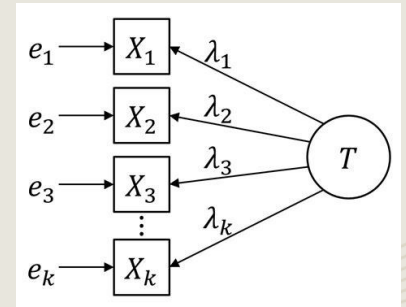
Face validity



Criterion validity



Model/latent construct





# Conclusiones

- Cada actividad de medición tiene tanto un instrumento de medición como un modelo de medición subyacente.
- Los modelos de medición surgen de la serie de supuestos que planteamos para representar la pobreza a través de datos observados.
- Tanto la confiabilidad como la validez son necesarias para los índices de pobreza con bajo error de medición
- El error de medición afecta al error de clasificación y se propaga a través de todas las inferencias (tabulaciones cruzadas para avanzar en los modelos de regresión)
- Tanto la confiabilidad como la validez son propiedades de las puntuaciones -no de las escalas que se refieren- y son relativas al contexto en el que se implementa la medición.



# 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." *American 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, Richard E., William Revelle, Iftah Yovel, and Wen Li. 2005. "Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega_h$ : 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>.