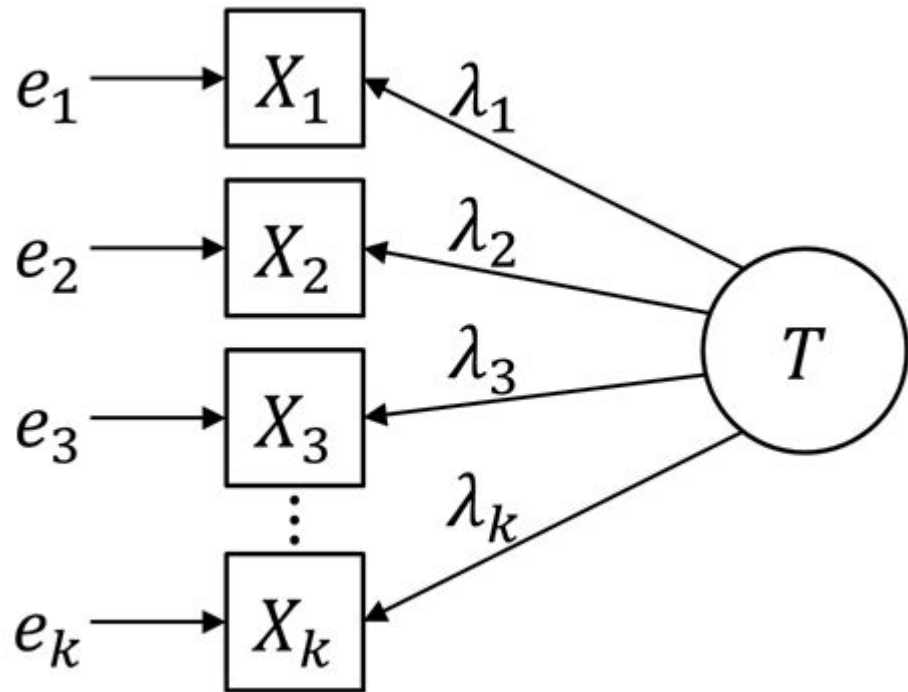


# Validez en medición

Héctor Nájera

# Variables latentes y confiabilidad



¿Qué sabemos de  $T$ ?

¿En qué términos hablamos de  $T$ ?

¿Es  $T$  es el fenómeno que me interesa?

¿Si los scores crecen entonces creció el fenómeno que me interesa?

# Variables latentes repensada por Goldstein (1989)

- Una persona con una habilidad  $\theta$  se enfrenta a un ítem, que tiene cierta dificultad y discriminación, y el resultado es una respuesta binaria
- Si la respuesta es correcta o incorrecta depende de la teoría ¿Pero de qué teoría? La teoría del TRI es estadística (*modelo estadístico*) no psicológica, de pobreza, exclusión, etc..
- ¿Por qué tenemos que restringir la respuesta al parámetro latente de habilidad?



# Validez

- La confiabilidad es una condición necesaria para una buena medición, pero no es suficiente
- No podemos hacer medición sin validez
- La pregunta es ¿Qué es validez?
- Siguiendo la crítica de Goldstein: Necesitamos un marco que le dé sentido a  $\Theta$



Sistema bajo  
medición

Fenómenos  
(ante los ojos)

≠

Objetos científicos

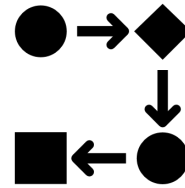


Resultados de  
Medición

Fenómenos  
(ante los ojos)

≠

Observación  
(codificada)



Modelo de  
medición

Puntajes

≠

Objetos científicos

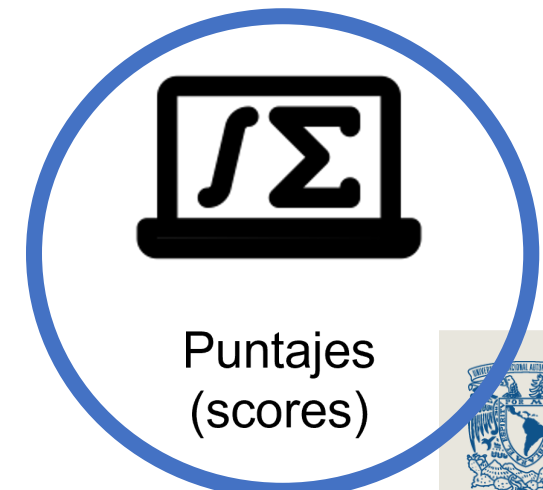


Indicaciones  
instrumentales

Datos

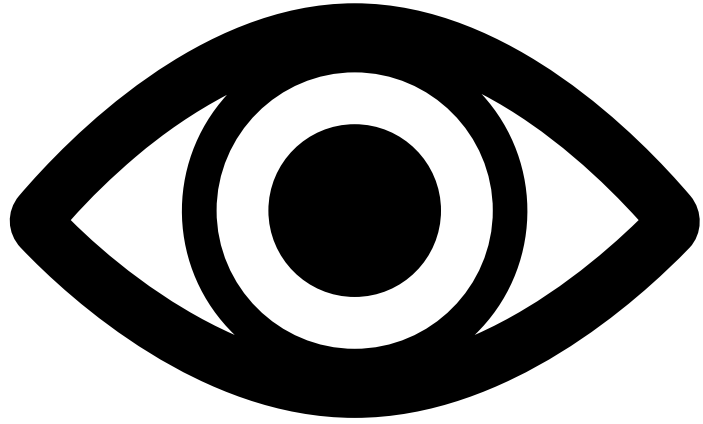
≠

Estimadores



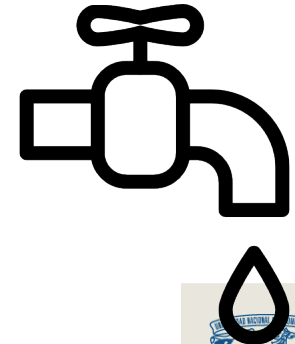
Puntajes  
(scores)





# Sistema bajo medición

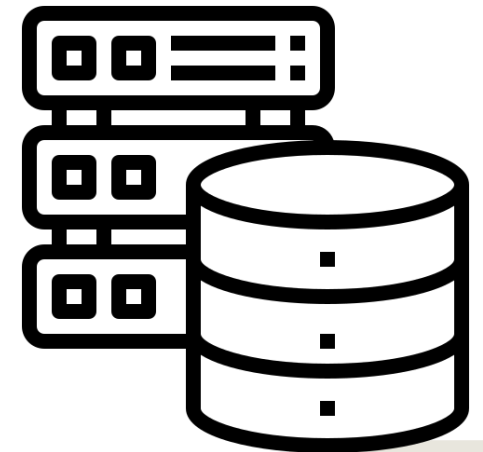
- Los referentes
- El mundo (natural) allá afuera
- Los fenómenos (ante los ojos)

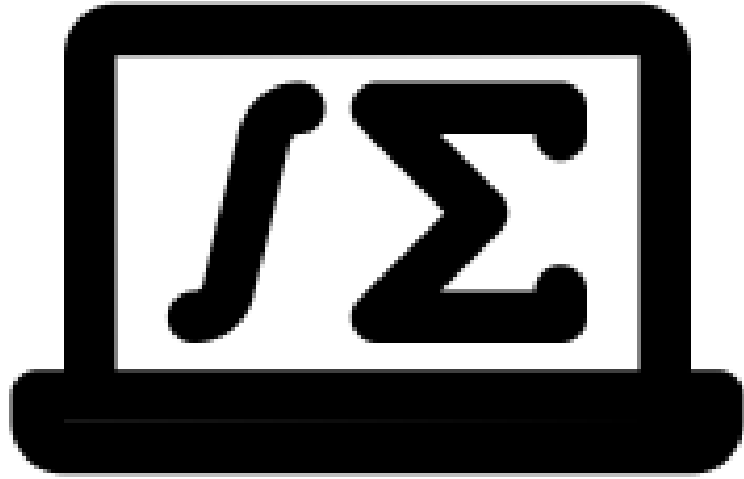




## Indicaciones instrumentales

- Lecturas de los instrumentos
- Generación/fuente de datos
- Indicadores (sin compromiso)
- Variables en bases





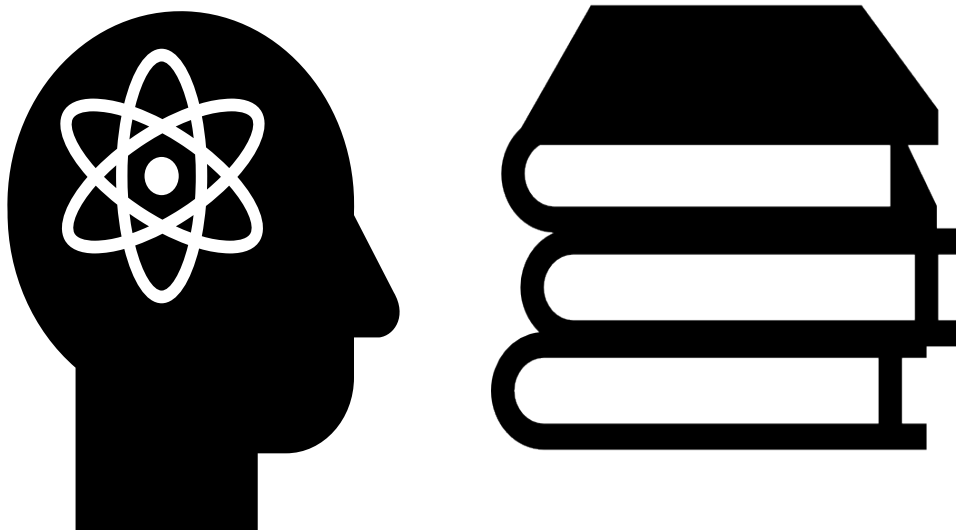
# Puntajes (scores)

- Procesamiento/transformación/ajuste de datos
- Modelaje estadístico
- Método de agregación

$$M_0 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^d w_j g_{ij}^0(k)$$

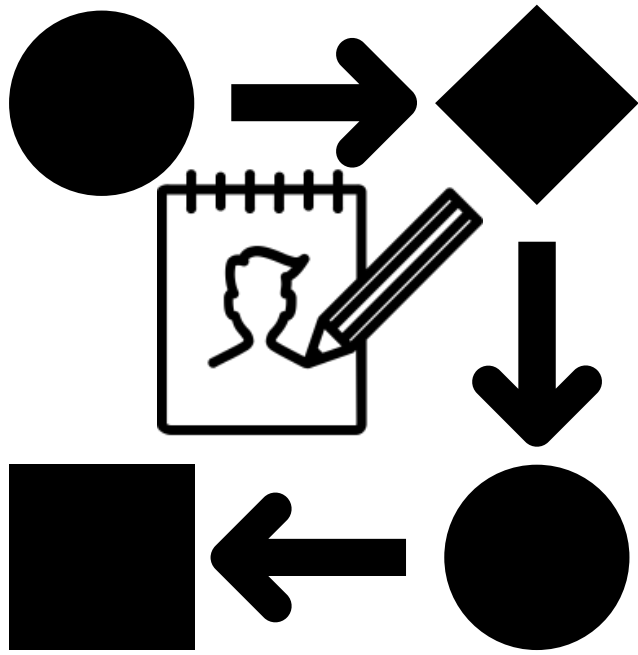
$$X_{ik} = [ a_k + b_k (T_i) ] + E_{ik}$$





# Resultados de Medición

- **Afirmación de conocimiento** acerca de una o más cantidades atribuidas al sistema bajo medición
- Formuladas en clave de objetos científicos, conceptos abstractos y universales -e.g. masa, corriente, temperatura, duración, pobreza

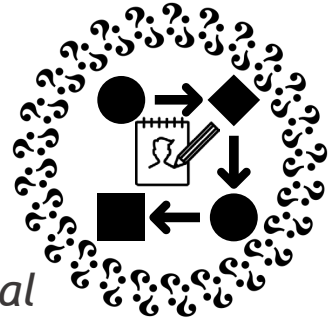


# Modelo de medición

- Modelo teórico o estadístico del proceso de medición mismo
- Una representación abstracta y local construida a partir de supuestos simplificadores
- Sólo bajo el modelo es posible evaluar la *interpretabilidad representacional* de los puntajes (su validez)
  - Coherencia de los supuestos con las teorías contextuales relevantes
  - Consistencia mutua de resultados con diferentes instrumentos, ambientes y modelos
- Sin modelo no hay medición

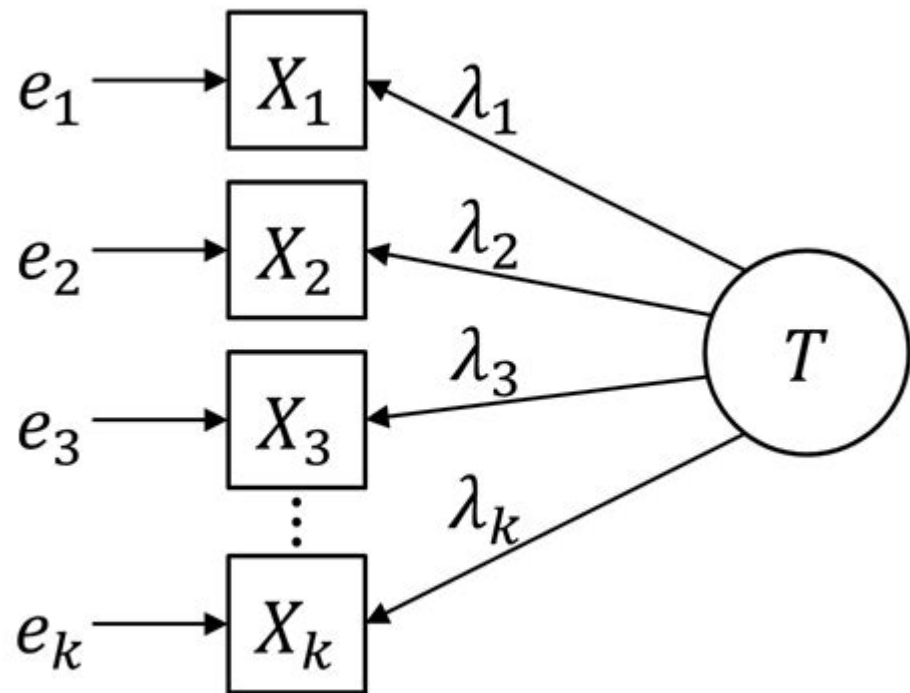
# Confiabilidad y validez

## Sólo bajo el modelo de medición



- 1. Mental health screening of preschool children: Validity and reliability of ABLE. *American Journal of Orthopsychiatry*, 77(3), 402-418, 402.
  - *ABLE (Attention, Behavior, Language, and Emotions), a new screening tool, was used to estimate the prevalence and the severity of concerns parents and teachers have about children's school adjustment and evaluate their need for services. Data obtained from parents and teachers of children randomly selected from public Pre-K classrooms in 6 states (N = 415) and from a mental health screening of rural and urban children (N = 5,577) support the validity and reliability of ABLE. Barbarin, O. A. (2007).*
- 2. Hides, L., Dawe, S., Young, R. McD., & Kavanagh, D. J. (2006). The reliability and validity of the Severity of Dependence Scale for detecting cannabis dependence in psychosis. *Addiction*, 102, 35-40.
  - *The SDS [Severity of Dependence Scale] is a brief, valid and reliable screen for cannabis dependence among people with psychosis.*

# Variables latentes



Si  $\Omega > .8$  y es una escala de precariedad laboral

Podemos decir que scores más altos implican mayor precariedad

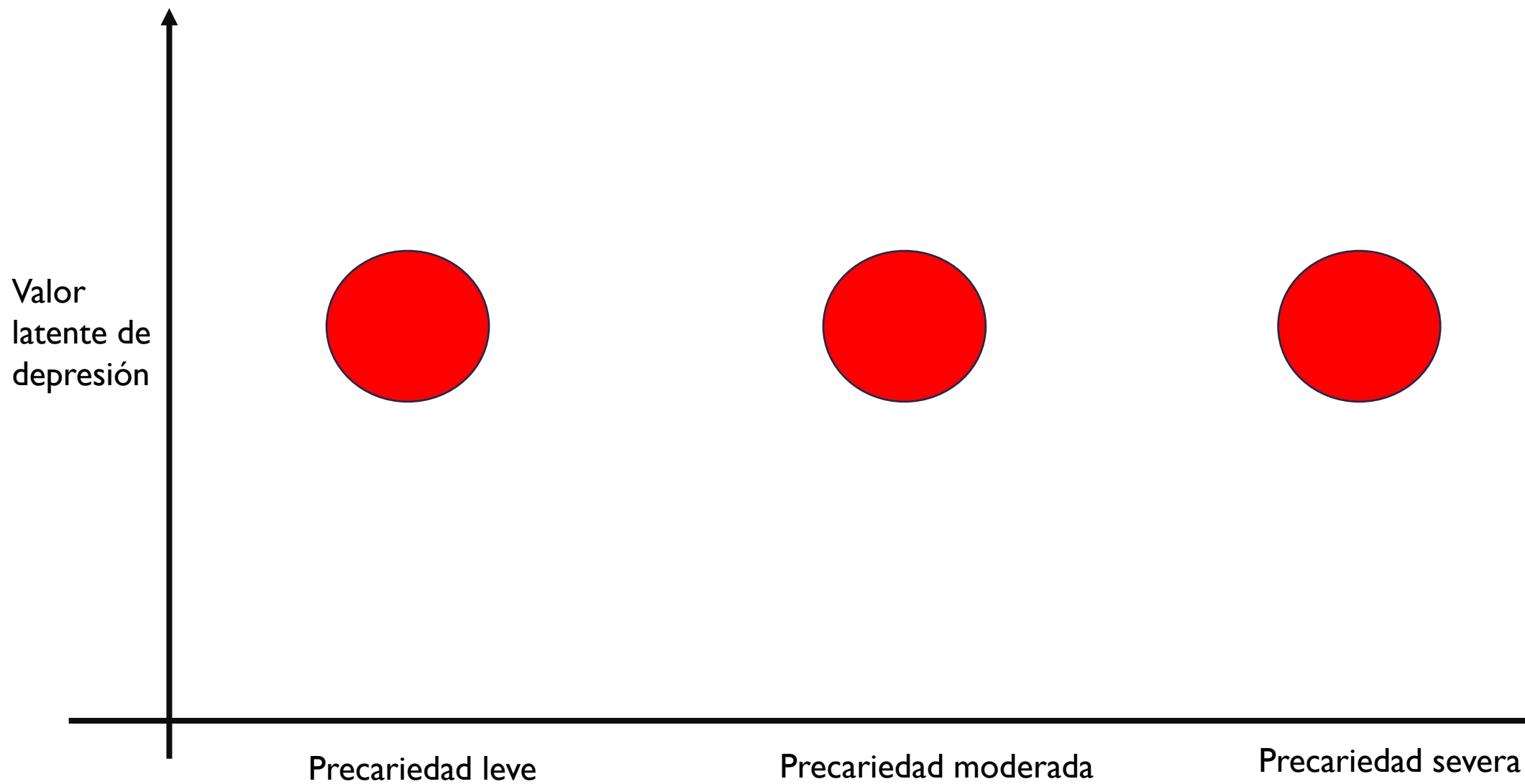
Podemos decir que scores más altos implican mayor -lo que sea que signifique la variable latente-

# Considere lo siguiente

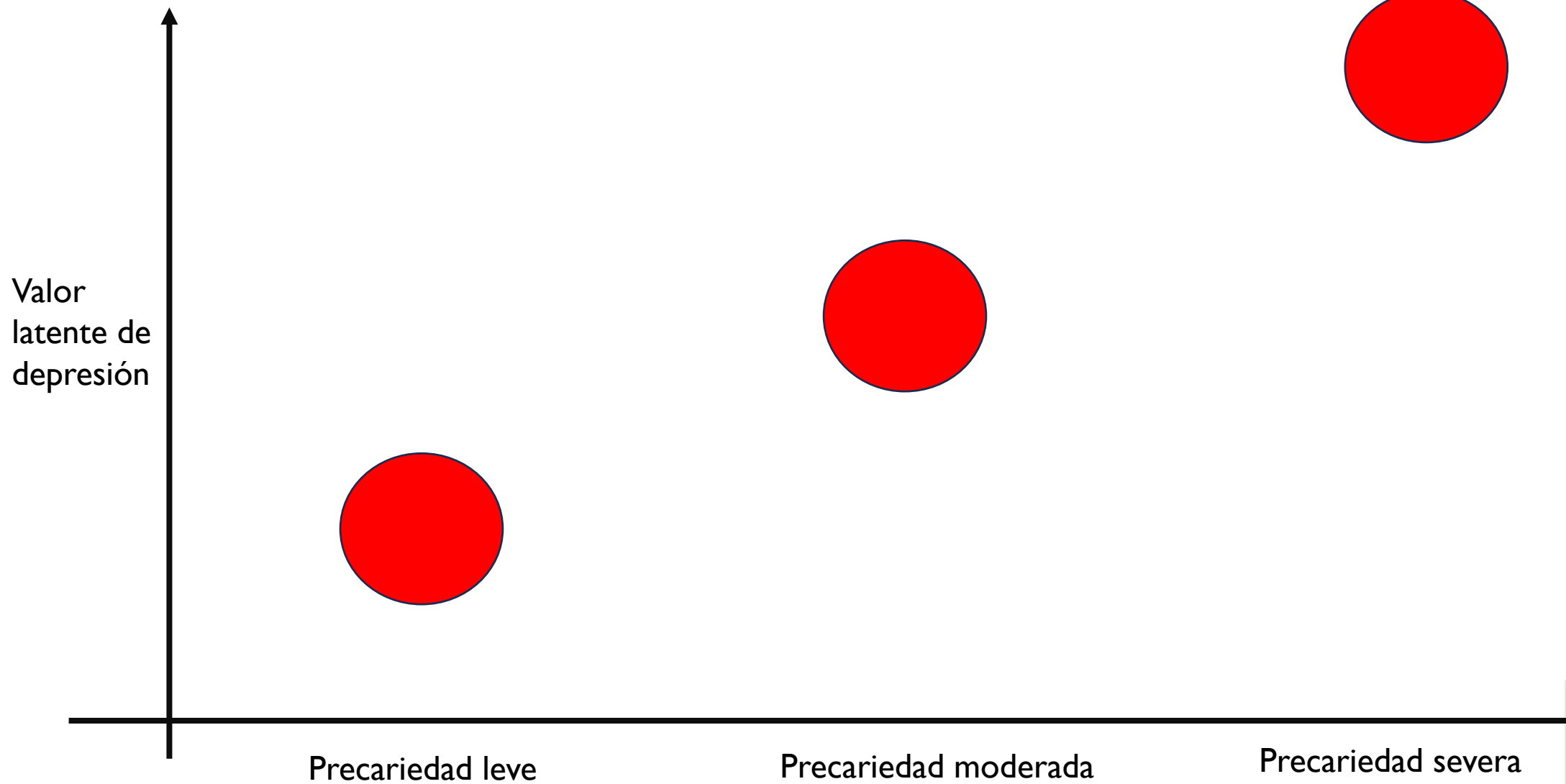
- Una persona con alto nivel educativo, riqueza y alto nivel de salud y a otra que, por el contrario, tiene bajo logro educativo, muy poca riqueza y sufre de problemas recurrentes de salud.
- Imagine que comparamos sus niveles de severidad de pobreza dado un conjunto de indicadores y encontramos un resultado inesperado:
  - La primera persona tiene claramente mayor nivel de severidad de pobreza que la segunda.



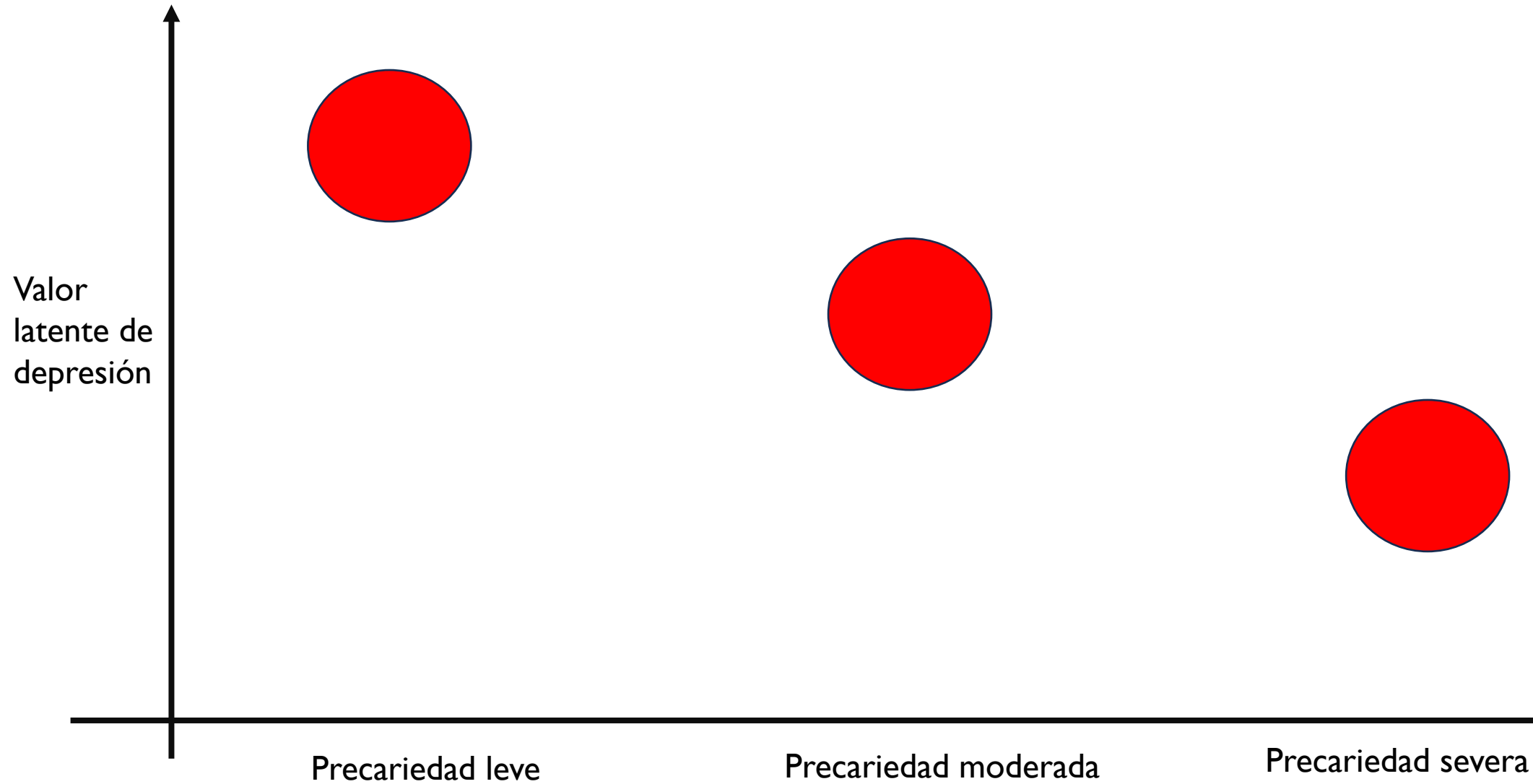
# Expectativa



# Expectativa

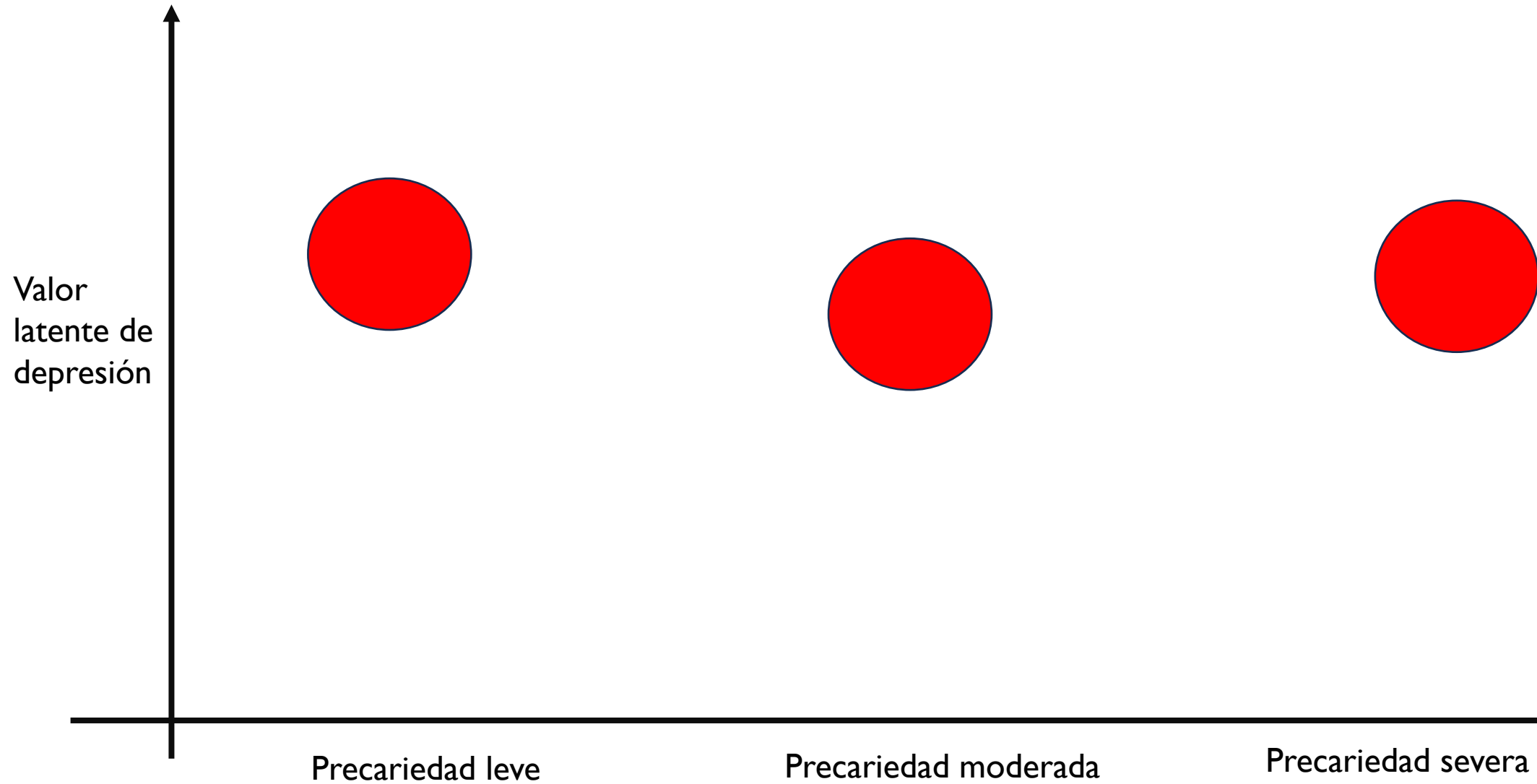


# Resultado

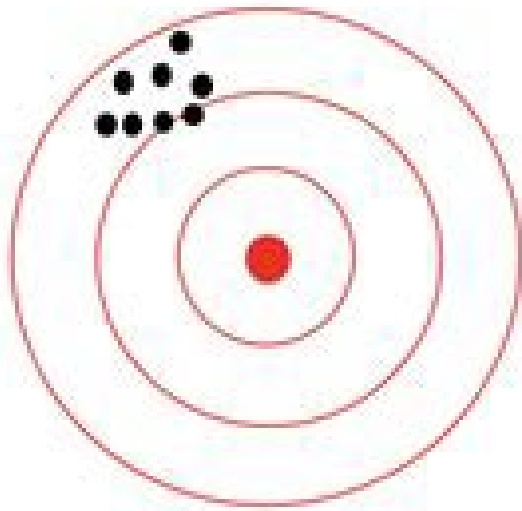




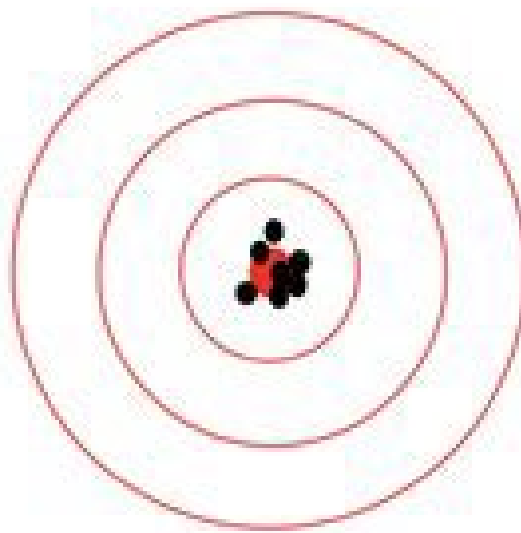
# Resultado



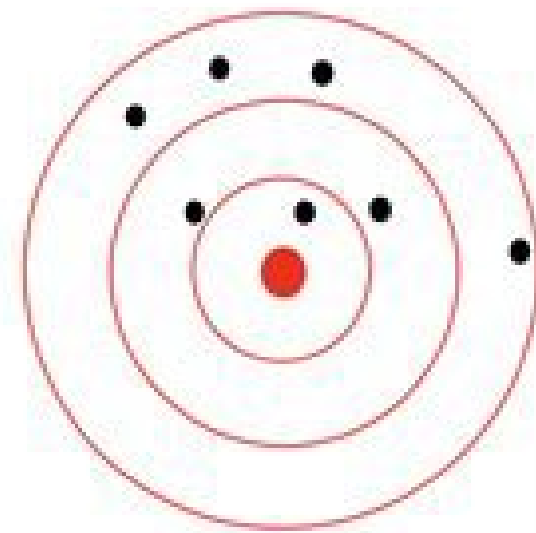
# Confiabilidad y validez



Reliable  
Not Valid



Valid  
and Reliable



Not Valid  
or Reliable

# Validez

Los scores son válidos cuando nos dicen la naturaleza de lo que se está midiendo y conectan con sus causas



# Validez, experimento y el estándar de oro

Nos interesa medir la ingesta de azúcar con una muestra representativa para la población.

- 1) Podríamos seguir a nuestra muestra varios días para producir un estimado
- 2) Podríamos usar un cuestionario

¿Cuál tiene más error?



# Validez y experimento

## FitNuts Food Log

Track your food for 14 days then complete the assessment to earn 10 Nutrition Points.

Participant Name: \_\_\_\_\_

Email Address

Date: \_\_\_\_\_

[illegible]

**VS**



# Validez y experimento

Pobreza  
Depresión  
Clase Social  
Inflación  
Marginación  
Desarrollo Humano  
Igualdad de Oportunidades  
Igualdad de Género  
Discriminación  
Inversión en capital humano  
Inversión pública  
Productividad

Timestamp	How likely is it that you would recommend SurveyMonkey to a friend or colleague?	What changes would this company have to make for you to give it a higher rating?	What does this company do really well?	What is your gender?	What is your age?	What is your approximate average household income?
10/29/2014	4			Female	75 or older	\$0-\$24,999
10/29/2014	10			Female	35 to 44	\$100,000-\$124,999
10/29/2014	7			Female	65 to 74	\$175,000-\$199,999
10/29/2014	8			Male	45 to 54	\$75,000-\$99,999
10/29/2014	8			Female	65 to 74	\$25,000-\$49,999
10/29/2014	3			Female	18 to 24	\$0-\$24,999
10/29/2014	7			Female	65 to 74	\$75,000-\$99,999
10/29/2014	2			Female	18 to 24	\$125,000-\$149,999
10/29/2014	6			Male	25 to 34	\$25,000-\$49,999
10/29/2014	7			Male	55 to 64	\$125,000-\$149,999
10/29/2014	1			Female	45 to 54	\$50,000-\$74,999
10/29/2014	9			Female	75 or older	\$150,000-\$174,999
10/29/2014	0			Male	55 to 64	\$25,000-\$49,999
10/29/2014	5			Male	65 to 74	\$125,000-\$149,999
10/29/2014	6			Female	18 to 24	\$100,000-\$124,999
10/29/2014	2			Female	55 to 64	\$25,000-\$49,999
10/29/2014	7			Male	35 to 44	\$100,000-\$124,999
10/29/2014	7			Female	65 to 74	\$200,000 and up
10/29/2014	0			Female	65 to 74	\$150,000-\$174,999
10/29/2014	10			Male	45 to 54	\$25,000-\$49,999
10/29/2014	6			Female	75 or older	\$150,000-\$174,999
10/29/2014	5			Female	75 or older	\$150,000-\$174,999
10/29/2014	2			Female	75 or older	\$50,000-\$74,999
10/29/2014	6			Female	25 to 34	\$150,000-\$174,999
10/29/2014	5			Male	55 to 64	\$25,000-\$49,999
10/29/2014	1			Female	18 to 24	\$75,000-\$99,999
10/29/2014	4			Female	75 or older	\$25,000-\$49,999
10/29/2014	9			Female	18 to 24	\$50,000-\$74,999
10/29/2014	8			Female	35 to 44	\$150,000-\$174,999
10/29/2014	10			Male	25 to 34	\$200,000 and up
10/29/2014	6			Female	25 to 34	\$100,000-\$124,999
10/29/2014	2			Male	35 to 44	\$125,000-\$149,999
10/29/2014	4			Female	45 to 54	\$125,000-\$149,999
10/29/2014	2			Female	35 to 44	\$200,000 and up

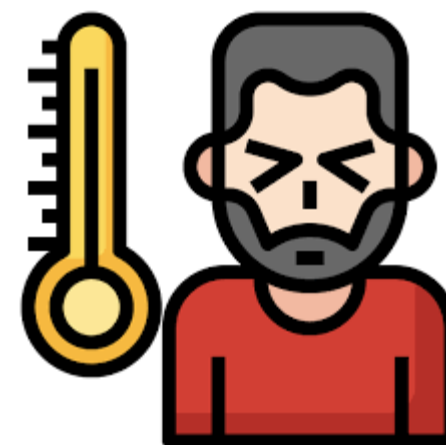
vs ?



# Validez

- La validez es, entonces, una propiedad que se realiza en función del modelo teórico de medición
  - Definición de lo que me interesa
  - Modelo teórico sobre sus causas y consecuencias
  - Modelo estadístico que convierte en parámetros las relaciones de interés

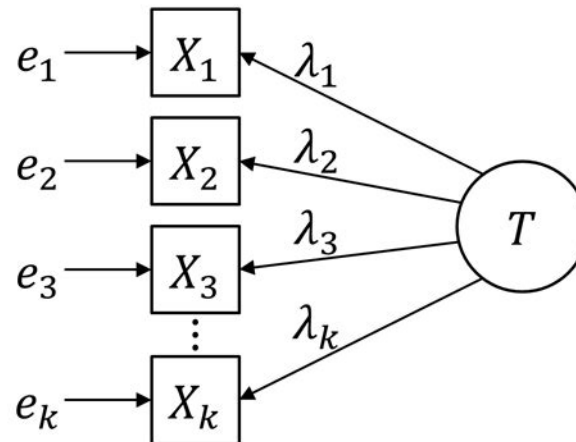
En el caso del termómetro hay un modelo físico sobre la relación entre cambios en la temperatura, expansión del mercurio y distancia.



Demuéstrame que el índice que creaste se puede interpretar de la forma que quieres hacerlo.

- La demostración se le en clave del modelo teórico

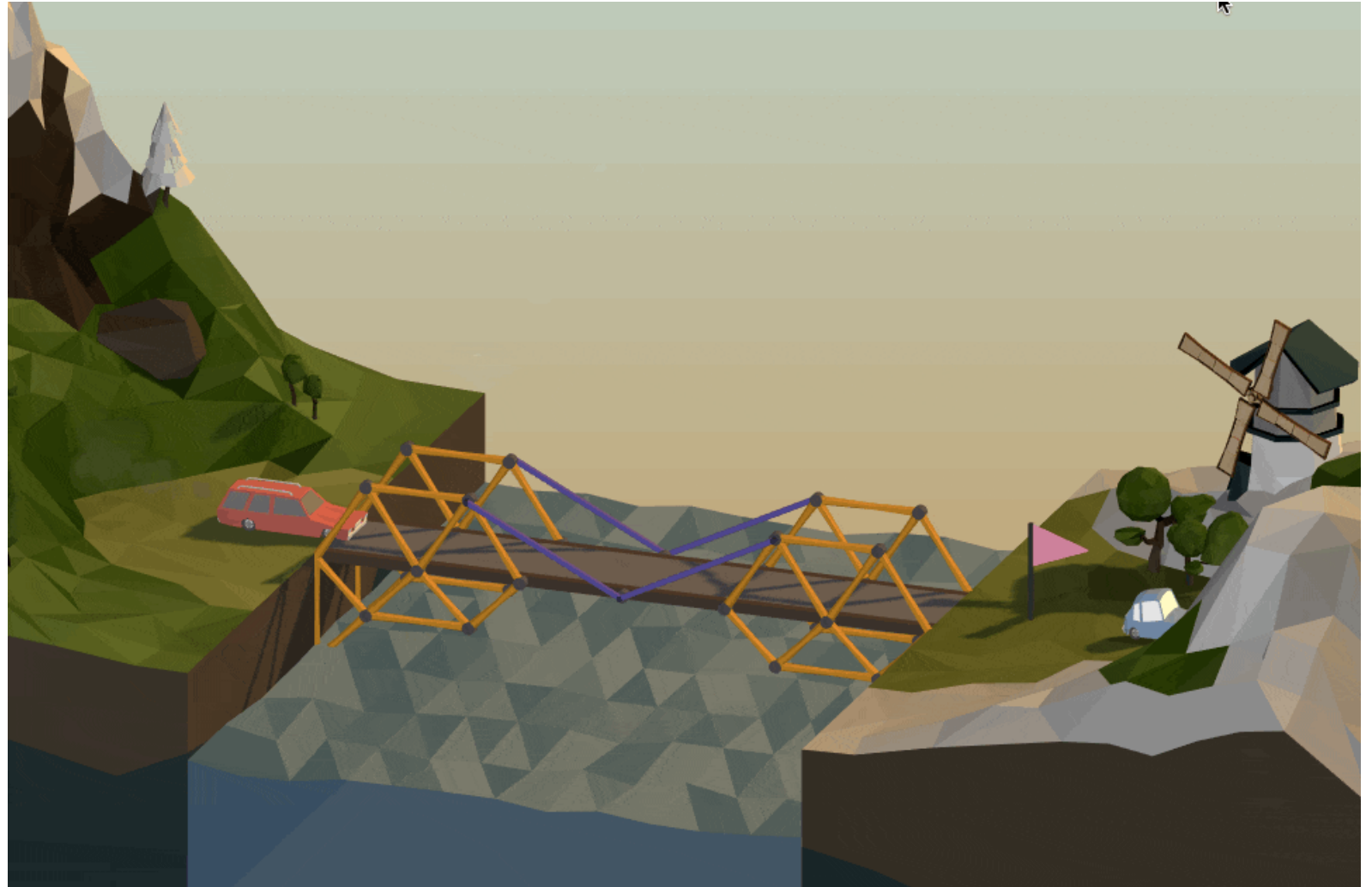
Es decir, ¿Cuál es la evidencia que te permite sostener que poner ciertos indicadores juntos te llevan a una medición de exclusión social y no de depresión?





# Validez

- Contenido
- Cara
- Criterio
- Concorrente
- **Constructo**

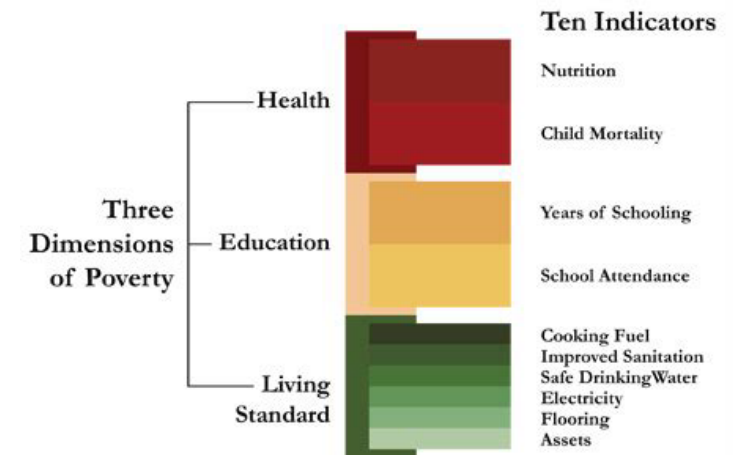


# Validez contenido

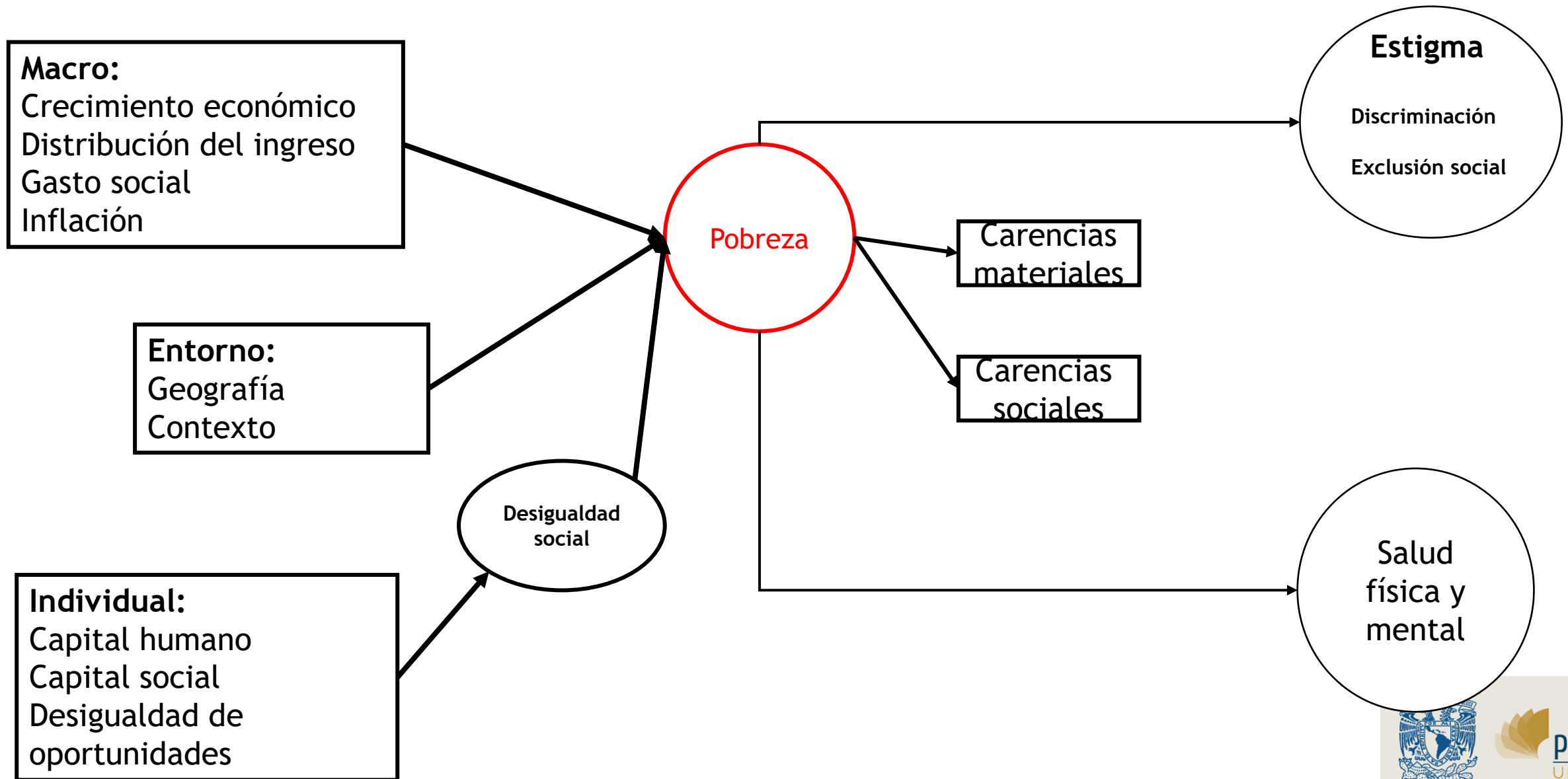
- Refiere a la validez del modelo teórico
- Modelo científico:
  - Definición
  - Distinción entre causas y consecuencias
- La validez de contenido busca atar los ítems a las razones teóricas detrás de su formulación
- En algunos casos se usan marcos normativos (derechos humanos) o teóricos para justificar su inclusión en las escalas

# Validez de contenido

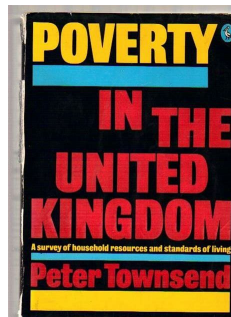
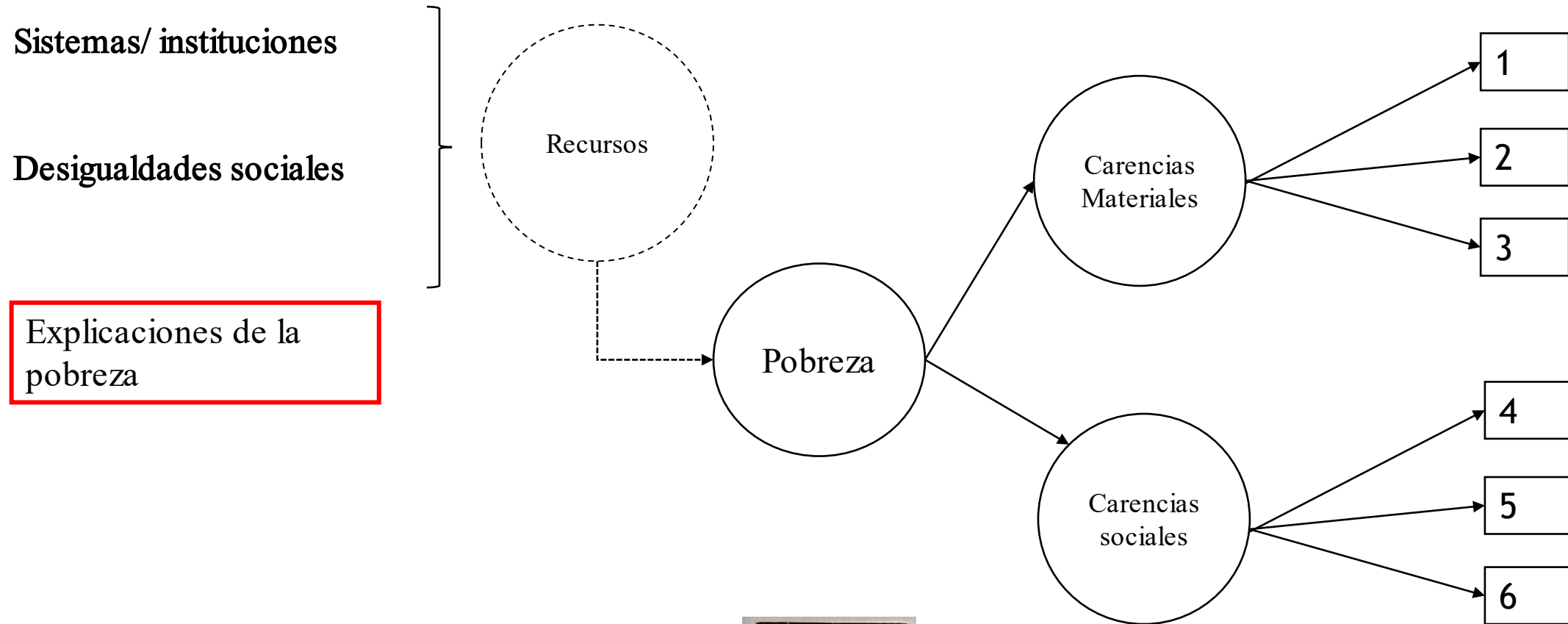
- OK! Las correlaciones entre distintos tests y algunos predictores del fenómeno puede ser útil pero...
- ¿Dónde quedó la teoría detrás de la selección y formulación de cada ítem?
- Como nos decía Goldstein, la TRI no es una teoría de las preguntas sino un modelo probabilístico sobre los 1 y 0's que vemos



# Explicaciones de pobreza factores



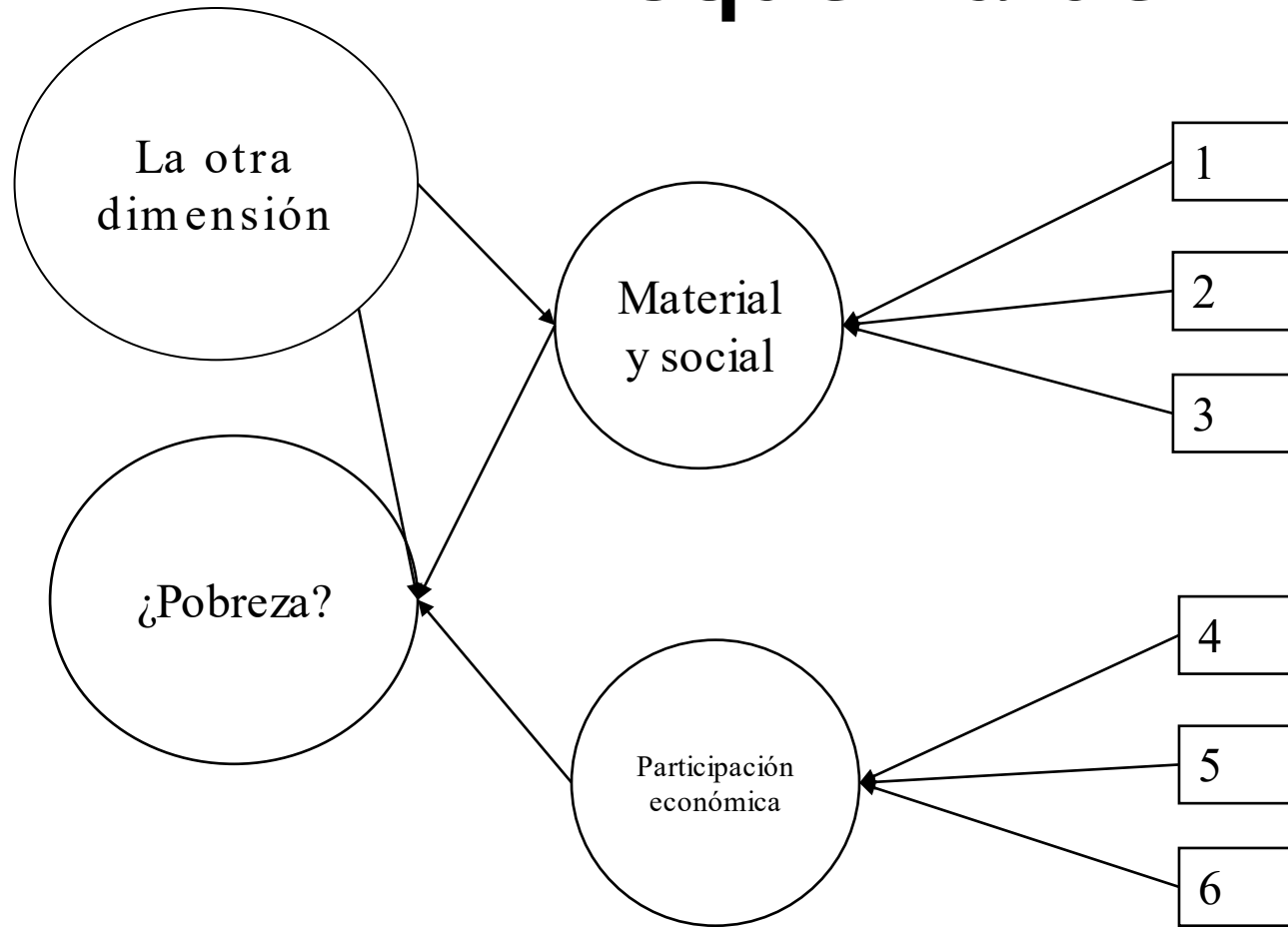
# Sin modelo, no hay medición



Modelo convencional de la teoría contemporánea de la medición



# Esquema de medición no científico



Modelos formativos  $\neq$  modelos científicos de medición

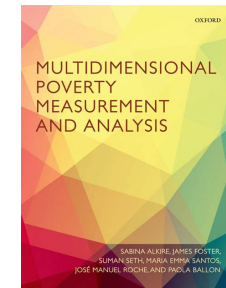
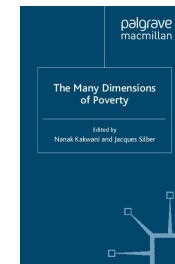
El reloj (instrumento)  $\neq$  el tiempo mismo (fenómeno)

¿Cambian los indicadores, cambia la definición?

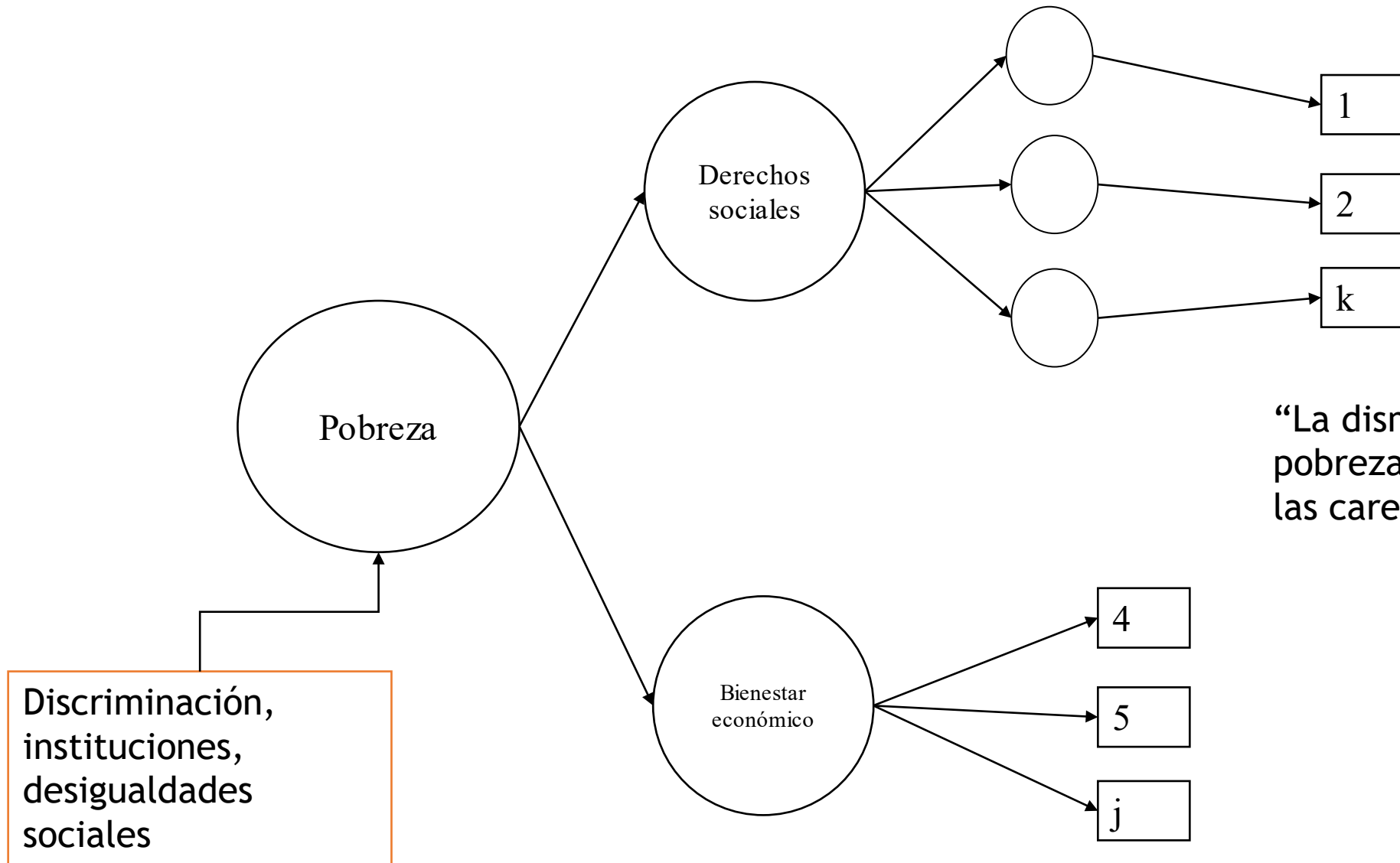
¿Cambian los indicadores, cambian las causas?

¿LDGS? ¿NBI?

La pobreza son los indicadores/ dimensiones que la componen

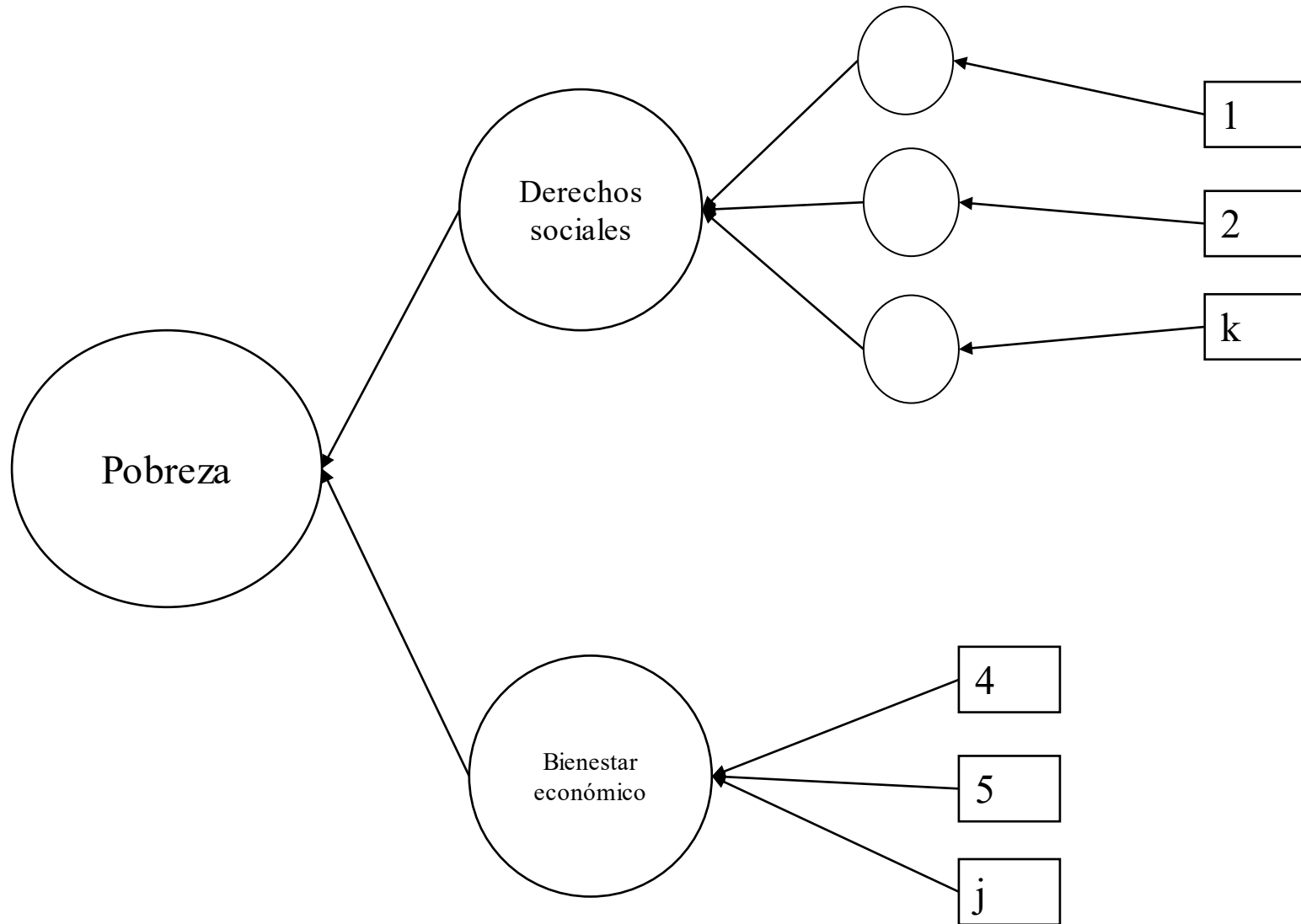


# Modelo mexicano



“La disminución de la pobreza es atribuible a que las carencias mejoraron”

# Modelo mexicano



“La disminución de la pobreza es atribuible a que las carencias mejoraron”

¿Causas = consecuencias?



# Validez de “cara” (face validity)

- Es la medida en la que los scores son interpretados por la población en estudio
- Es decir, la población nos indica si lo que estamos midiendo les hace sentido desde sus vivencias
- Se usa para producir cuestionarios y analizar sus resultados
- Últimamente ha ganado eco por el fracaso que han tenido medidas diseñadas desde el escritorio

# Métodos: Validez de cara

- La validez de cara requiere métodos mixtos: Cualitativos y Cuantitativos
- Cualitativos: Entrevistas a profundidad, grupos focales e incluso etnografía
- Cuantitativos

# Ejemplos



% thinking items a necessity (Sex and Age of Respondent)	Sex		Age group	
	Male	Female	<24 yrs	65+
A visit to a health facility when ill and all the medication prescribed to treat the illness	97	98	98	97
Three meals a day	96	95	95	96
Two sets of clothing	94	93	93	92
Toiletries to be able to wash every day (e.g. soap, hairbrush/comb)	93	93	94	90
All fees, uniform of correct size and equipment required for school e.g. books, school bag, lunch/lunch money, etc.	89	88	88	86
Own blanket	85	85	87	83
Own bed	81	81	82	79
Two pairs of properly fitting shoes, including a pair of all-weather shoes	80	79	82	74
Own room for children over 10 of different sexes	78	75	74	77
Books at home suitable for their age (including reference and story books)	72	71	72	69
Some new clothes (not second hand or handed on/down)	70	69	73	64
Bus/taxi fare or other transport (e.g. bicycle) to get to school	69	68	69	66
To be able to participate in school trips or events that cost money	68	69	70	66
A desk and chair for homework for school aged children	57	54	56	54
Presents for children once a year on special occasions, e.g. birthdays, Christmas, Eid	55	53	57	52
Educational toys and games	54	52	57	50
Some fashionable clothes for secondary school aged children	38	37	43	35
Own cell phone for secondary school aged children	23	21	26	24

# Ejemplo: Medición oficial mexicana

Elección del umbral oficial

Educación: 10%

Paredes: 1 %

Techos: 4%

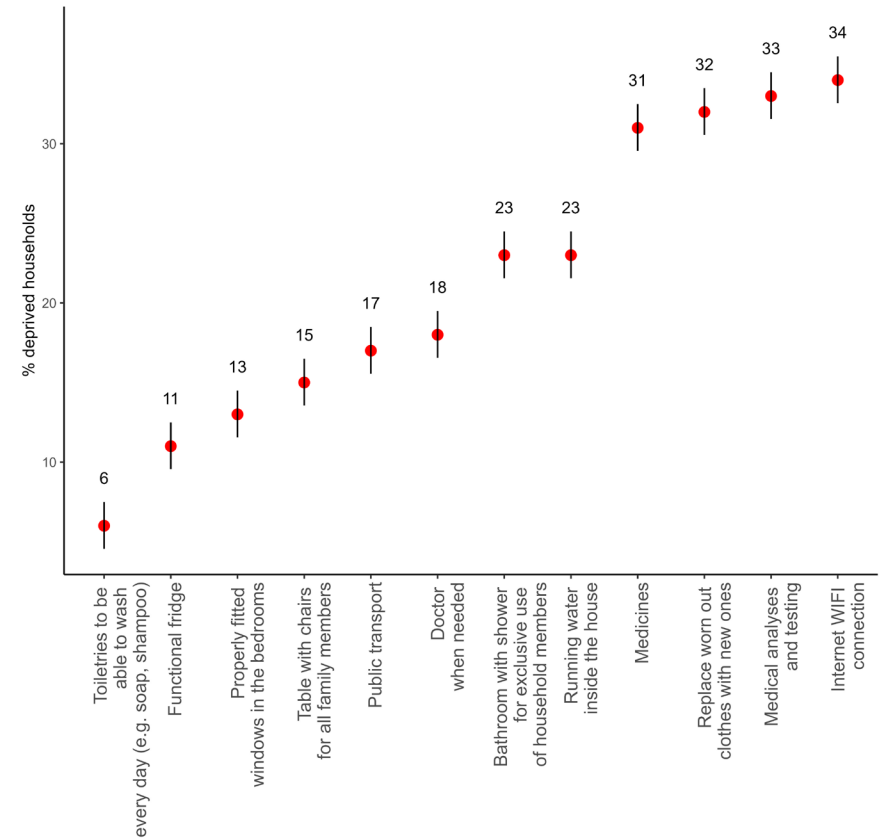
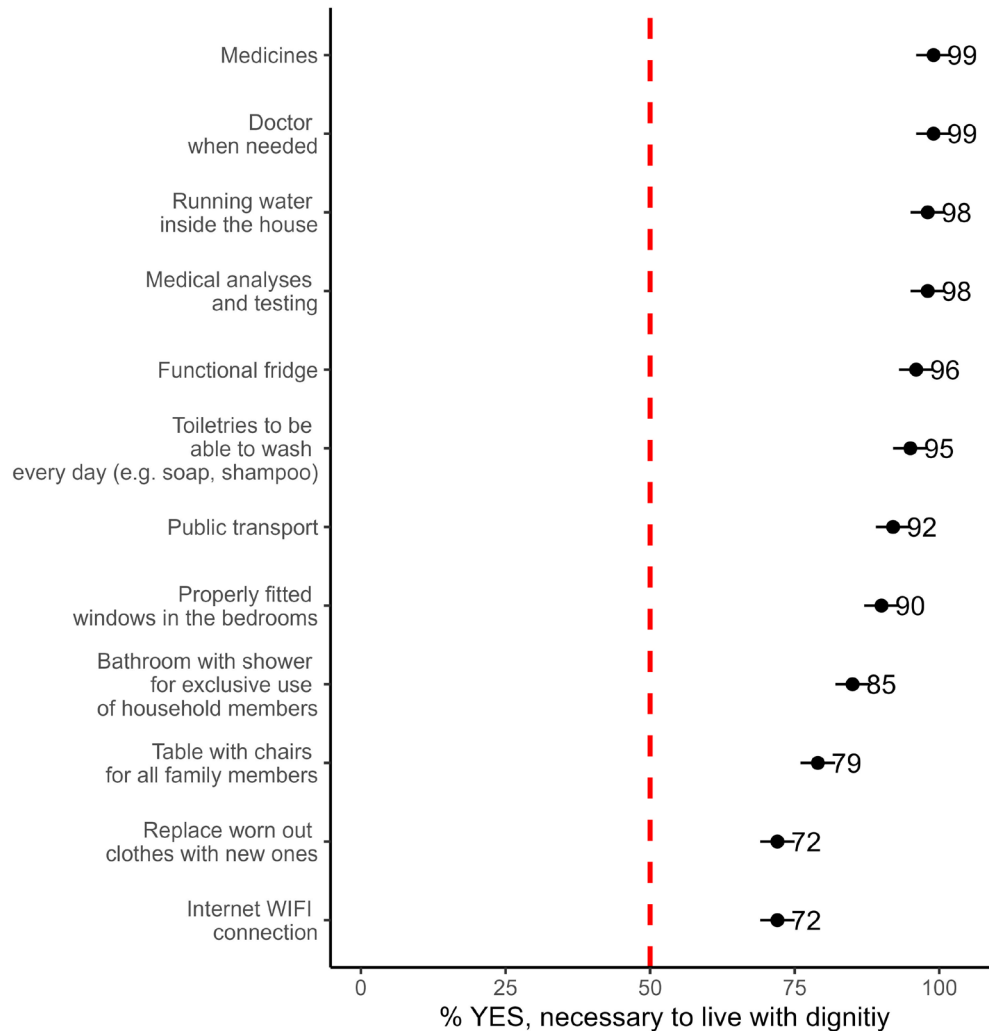
Agua: 9%

Social deprivation indicators	CONEVAL's estimates (%)	Consensual approach
1. Educational gap	19	21
2. Access to health care services	22	22
3. Access to social security	61	65
4. Quality and living space of the dwelling	14	35
5. Access to basic services in the dwelling	21	52
6. Access to food	23	23
7. Access to information	n/a	68

EDUMP 2017



# Ejemplo: Privación relativa



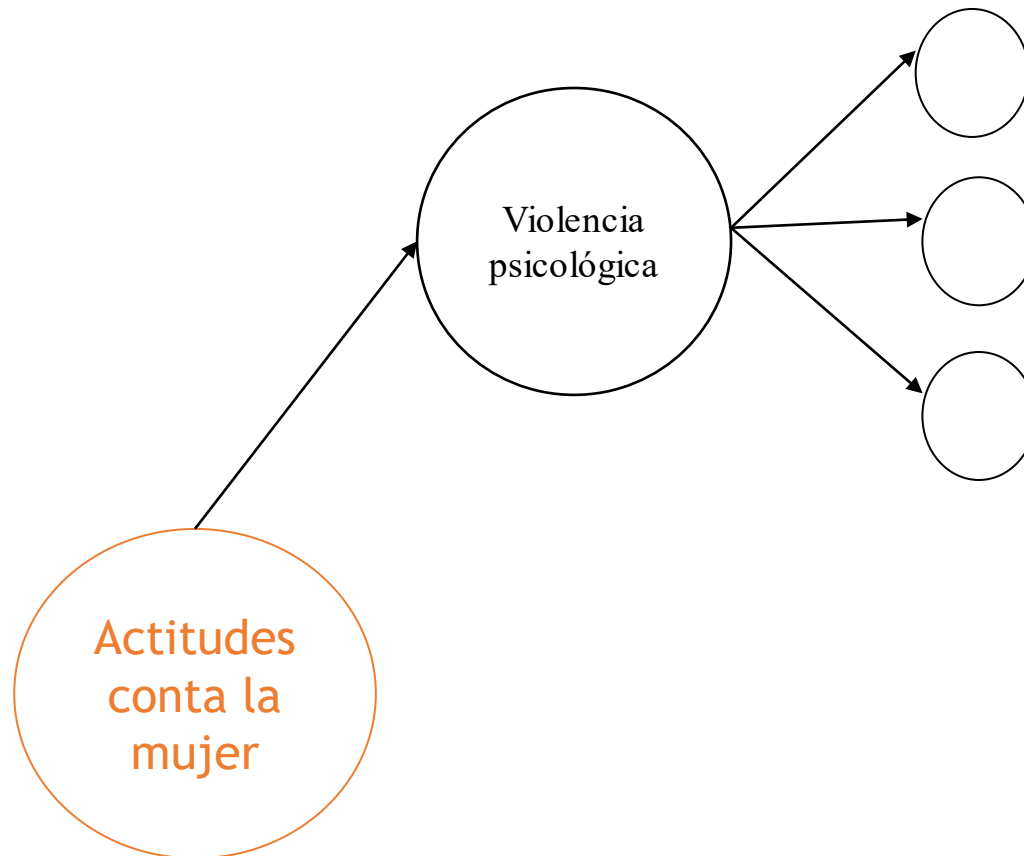
# Validez de criterio: Sus inicios

- En los 30s: Los conceptos son (a) convenientes, creados, constructos explicativos y (b) son inobservables pero inferidos a través de manifestaciones
- La primera forma de validación de los scores consistía en su capacidad de predecir algún “criterio”

Por ejemplo:



# Considere el siguiente modelo



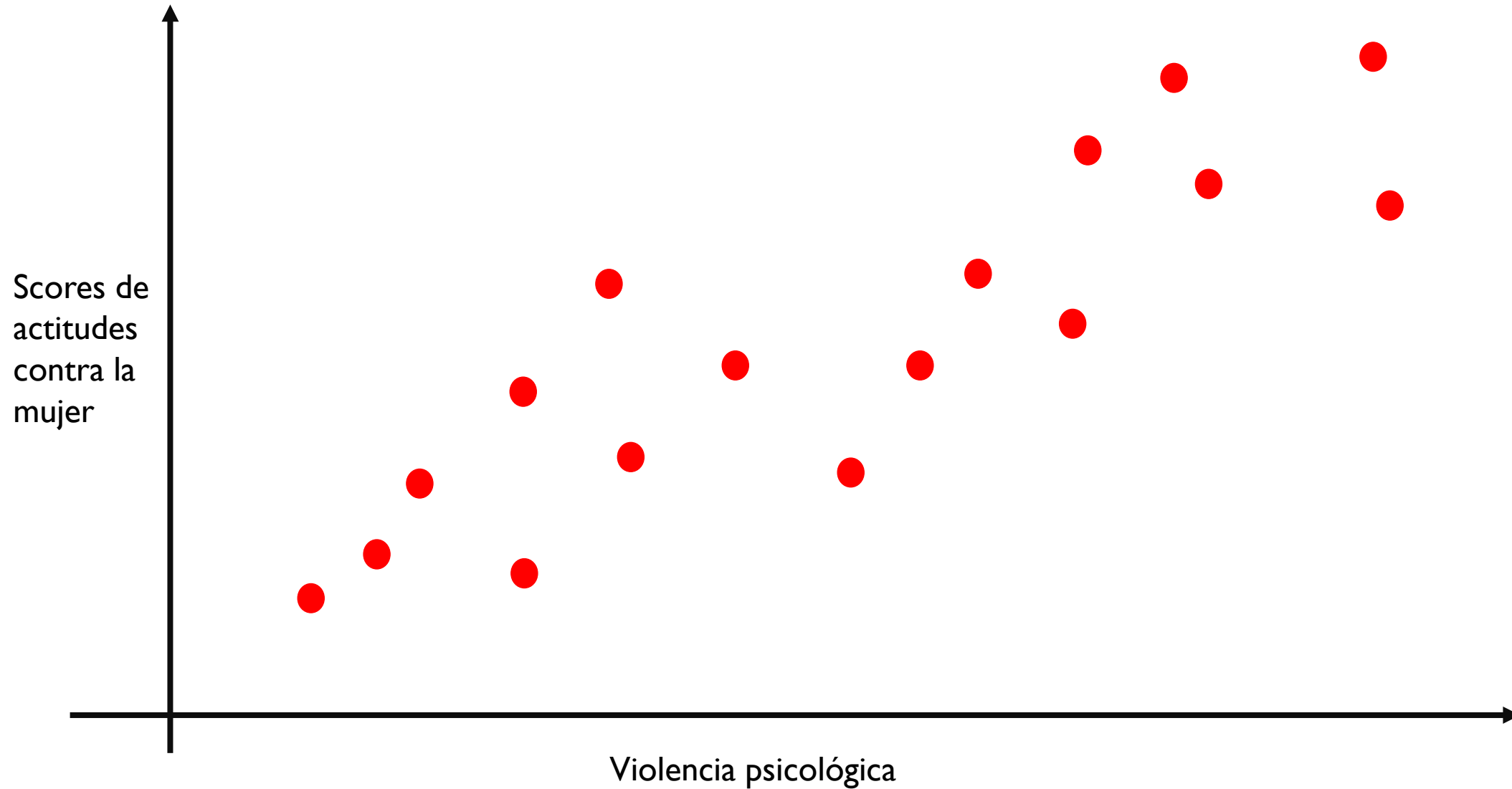
Teoría:

Las normas sociales justifican ciertos juicios sobre lo que es permisible en las relaciones de género. Estas normas se traducen en **actitudes contra la mujer**.

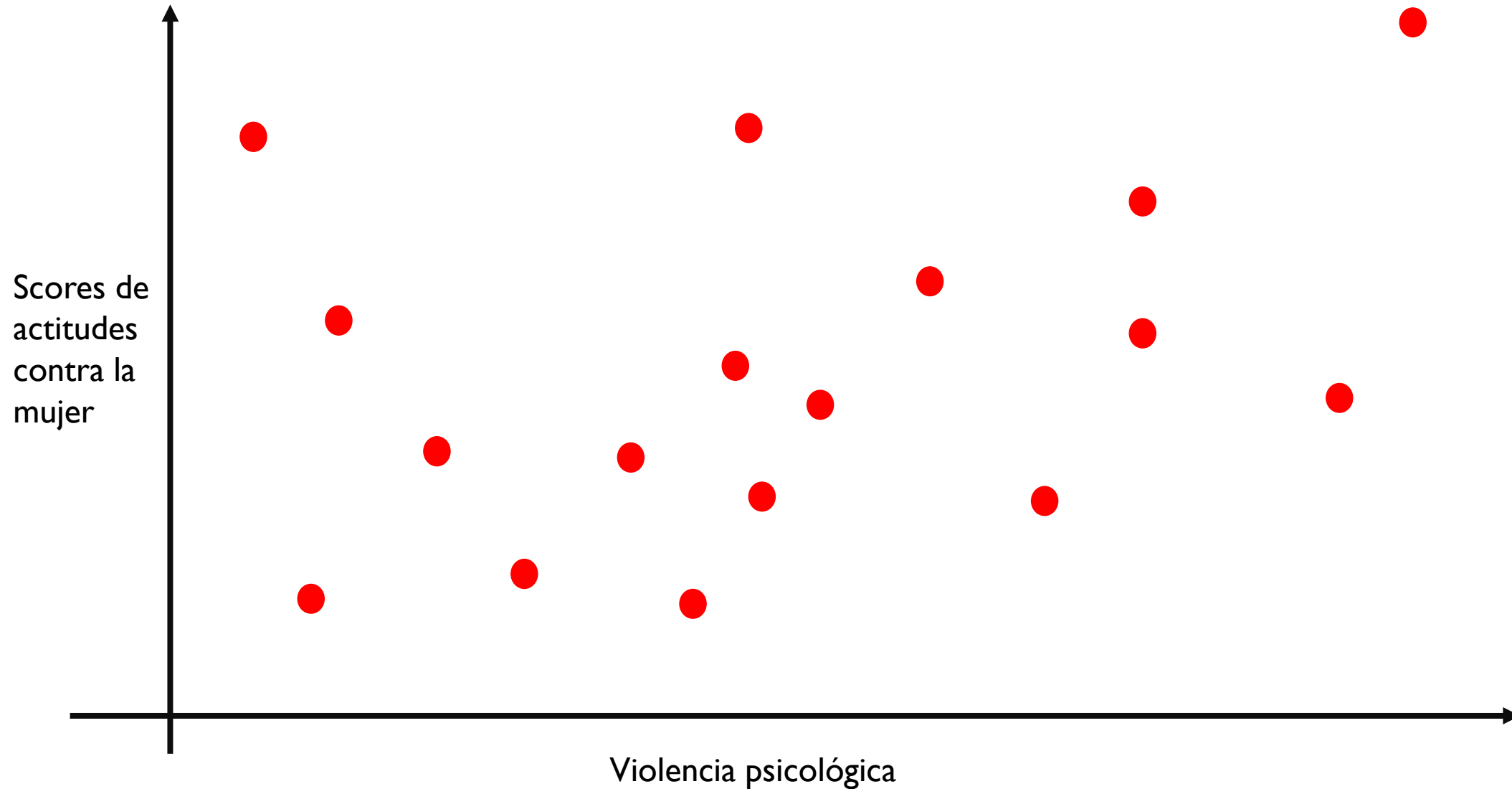
Cuando ciertas actitudes negativas se asientan, el riesgo de perpetrar **violencia psicológica** aumenta.



# Validez de criterio



# Validez de criterio



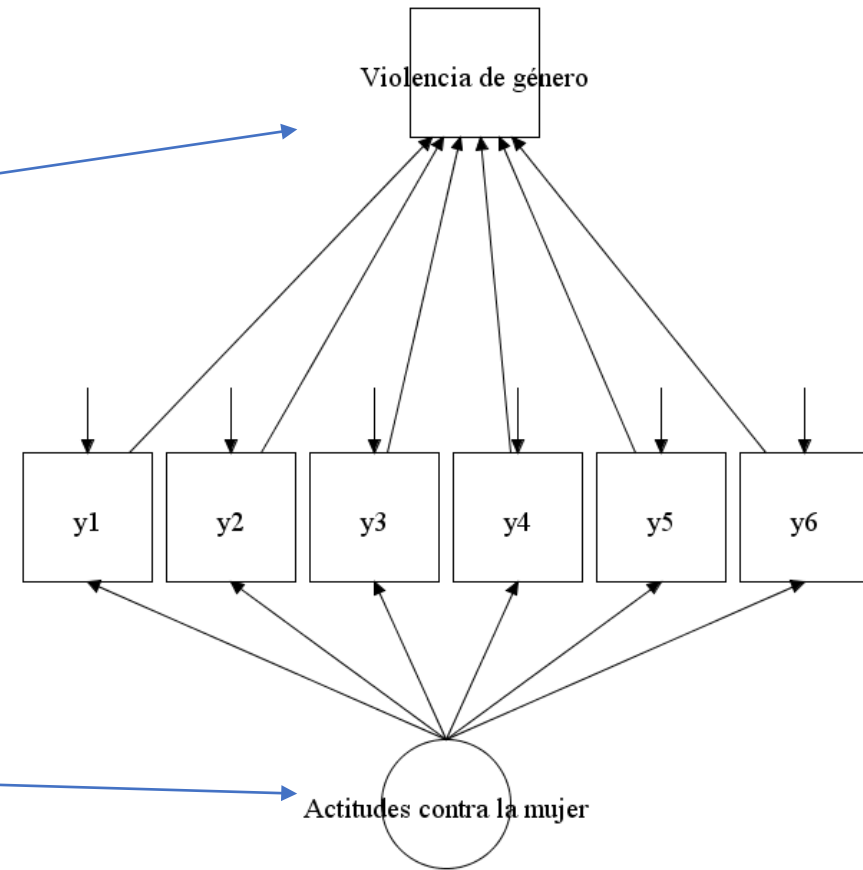
# Validación de criterio a la antigua

Lo que se hacía era correlacionar los indicadores con alguna variable auxiliar

Esto llevo a pensar en “proxies”

Uno tenía correlaciones / pendientes para cada indicador

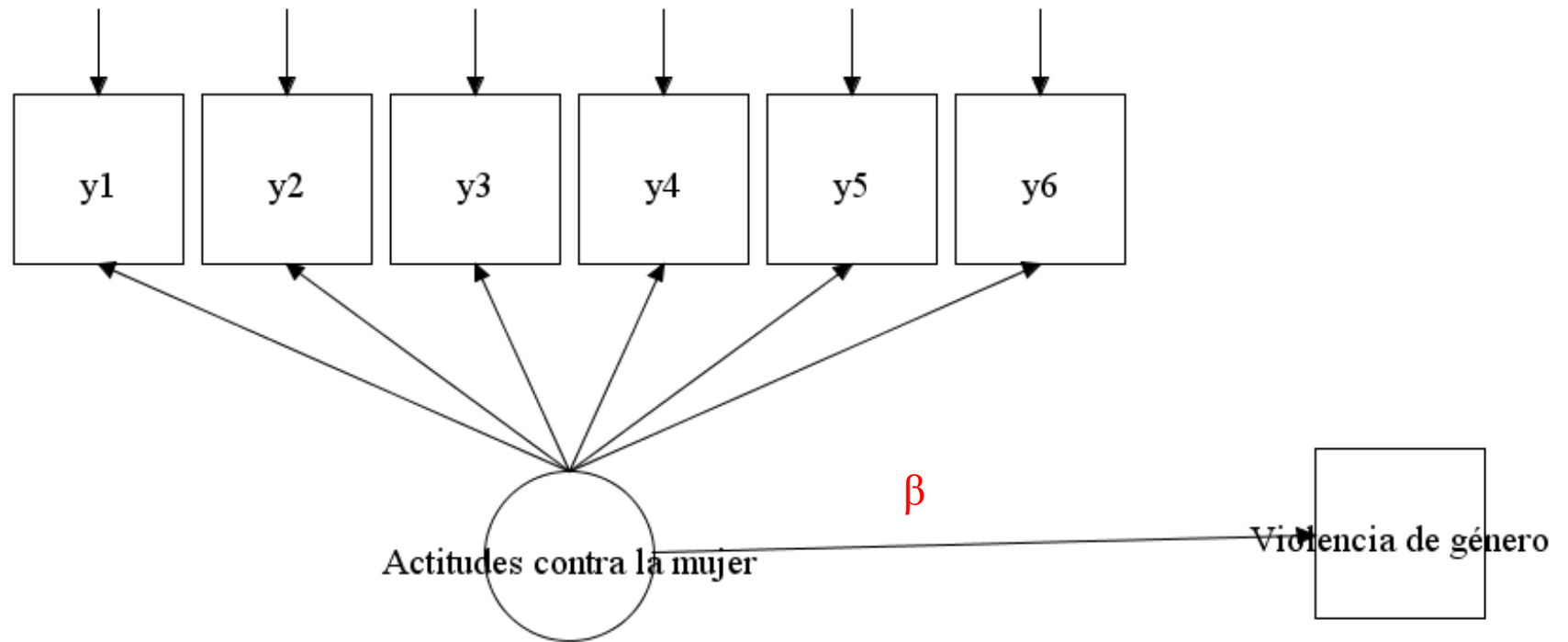
Antes era muy difícil calcular el valor de los scores latentes



# SEM: Validez de criterio o predictiva

Con SEM podemos  
usar directamente  
los scores latentes

Un coeficiente nos  
los dice todo

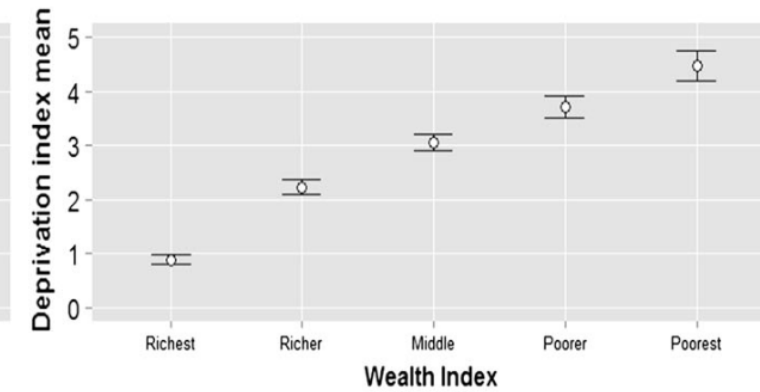
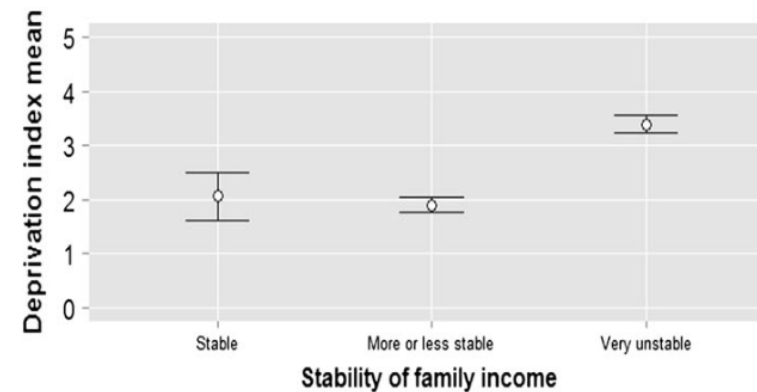
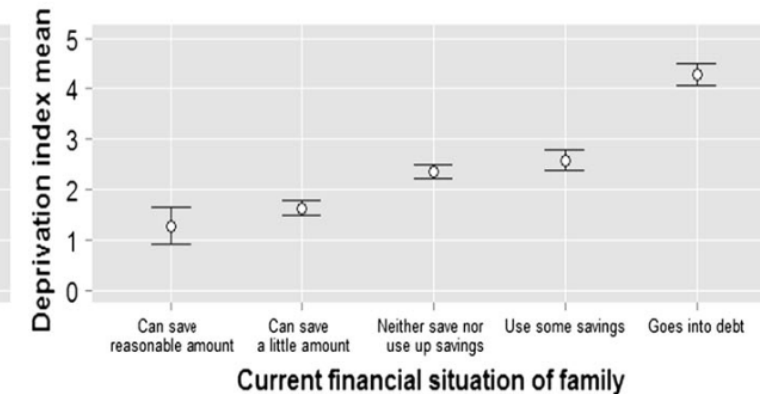
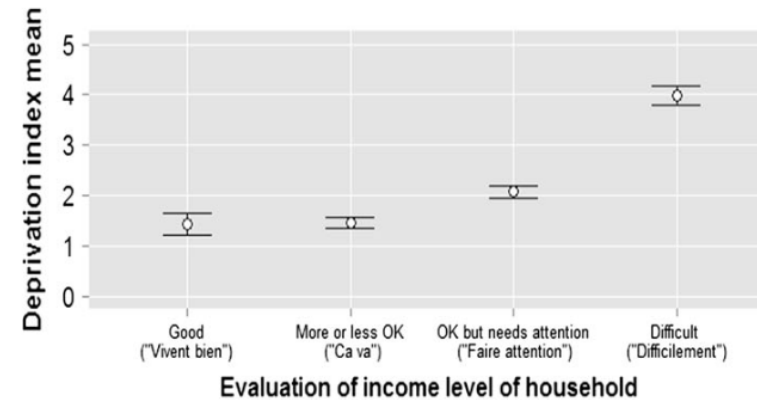


# Validez de criterio: Descriptivos

S. Nandy, M. Pomati

Siempre hay que trabajar con hipótesis. Esto no es un análisis de sensibilidad.

¿Cuáles son los cambios en gradientes esperados?



# Validez criterio

1. Head of household's education level (scored from 1 to 8 – from 'degree' to 'no formal education') – **HHEd**
2. Expenditure poverty using the official measure (1 poor, 0 not poor) – **Poor**
3. Head of household's economic activity ('working in the cash economy' vs 'subsistence/family worker or unemployed') – **HHact**
4. Subjective poverty (scored from 1 to 5 - from 'very rich' to 'very poor') – **Sub\_pov**
5. International Socio-Economic Index (ISEI) of occupational status– a widely-used measure of occupational status. Lower scores indicate higher occupational prestige – **ISEI**.

TABLE A1.8: LOGISTIC REGRESSION VALIDITY TESTS FOR CHILDREN AND HOUSEHOLD DEPRIVATION ITEMS

Children and household items + 2 adult items concerning children		HHEd	Poor	HHact	Sub_pov	ISEI
1	<b>Child:</b> Three meals a day	1.5	5.5	1.8	3.5	28.7
2	<b>Child:</b> Two pairs of properly-fitting shoes	1.6	6.9	2.3	2.9	44.6
3	<b>Child:</b> Toiletries to be able to wash every day	1.4	3.3	1.5	2.7	15.2
4	<b>Child:</b> Books at home for their age	1.3	2.1	1.3	1.6	7.1
5	<b>Child:</b> Some new clothes	1.4	4.1	1.8	2.6	18.0
6	<b>Child:</b> Educational toys and games	1.2	2.4	1.3	1.4	4.1
7	<b>Child:</b> A visit to health facility when ill and all prescribed medication	1.2	1.7	1.1	2.0	6.3
8	<b>Child:</b> Own bed	1.5	5.1	1.8	2.5	16.8
9	<b>Child:</b> Own blanket	1.5	5.1	1.8	2.5	19.1
10	<b>Child:</b> Two sets of clothing	1.4	2.9	1.6	2.6	17.5
11	<b>Child:</b> Presents for children once a year on special occasions	1.3	3.1	1.7	2.1	8.9
12	<b>Child:</b> All school fees, uniforms of correct size and equipment	1.3	2.1	1.4	1.7	7.8
13	<b>Child:</b> To be able to participate in school trips	1.2	1.8	1.2	1.5	5.1
14	<b>Child:</b> A desk and chair for homework	1.2	1.6	1.1	1.3	3.0
15	<b>Child:</b> Bus/taxi fare or other transport	1.2	1.7	1.3	1.5	3.9
16	<b>Child:</b> Own room for children over 10 of different sexes	1.1	1.3	1.0	1.3	1.2
17	<b>Household:</b> Enough money to repair or replace worn-out furniture	1.4	4.1	1.9	2.9	19.0
18	<b>Household:</b> Enough money to repair or replace electronic goods	1.2	2.0	1.3	1.6	5.7
19	<b>Household:</b> To be able to make savings for emergencies	1.3	3.4	1.8	2.6	10.5
20	<b>Household:</b> Enough money to repair a leaking roof for main living q.	1.3	2.4	1.5	2.0	9.7
21	<b>Household:</b> To have own means of transport	1.2	2.1	1.7	1.9	3.2
22	<b>Household:</b> Enough money to replace broken pots and pans	1.3	2.9	1.8	2.3	11.6
23	<b>Adult:</b> Enough money to pay school fees	1.3	1.9	1.3	1.7	5.6
24	<b>Adult:</b> Enough money to take the child to a medical facility when sick	1.3	3.0	1.6	2.4	10.7

Esto ayuda mucho, pero  
cuáles serían los  
problemas potenciales?

### Appendix 3: Item validation

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



# ¿Qué hay de esto?

Ingreso per capita	Coef.	P>t	[95% Conf.	Interval]
Pisos	-131	0.361	-412	150
Techos	43	0.092	-7	92
Paredes	202	0.238	-133	538
Hacinamiento	-735	0.000	-867	-603
Tenencia	-76	0.352	-236	84
Agua	-179	0.009	-314	-44
Saneamiento	-362	0.000	-493	-232
Combustible	-971	0.000	-1106	-837
Bienes durables	-143	0.065	-296	9
Educacion basica inf	-244	0.056	-493	6
Rezago educativo	-78	0.557	-338	182
Educacion adultos	-422	0.000	-610	-234
Desempleo	-393	0.000	-513	-272
Proteccion social	-1445	0.000	-1704	-1186
_cons	4714.443	0.000	4459.285	4969.601

Es  
manifestación o  
causa de  
pobreza.

Parecen tener la misma fuente de variación pero porque son parte de la misma!

No es un índice de pobreza sino índice de causas y consecuencias de la pobreza





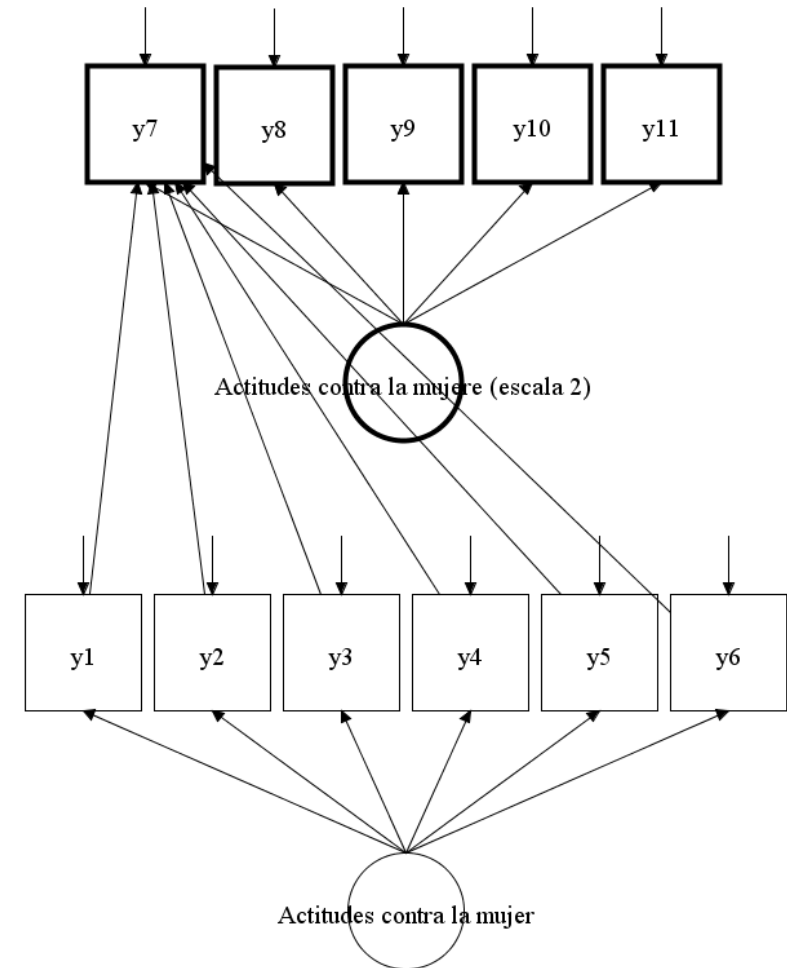
# Validez concurrente

- Spearman 1904: En aquella época la distinción entre confiabilidad y validez no era tan clara
- Spearman tenía distintas medidas de inteligencia y habilidad musical
- La idea es similar pero el foco es son los scores globales entre distintos tests

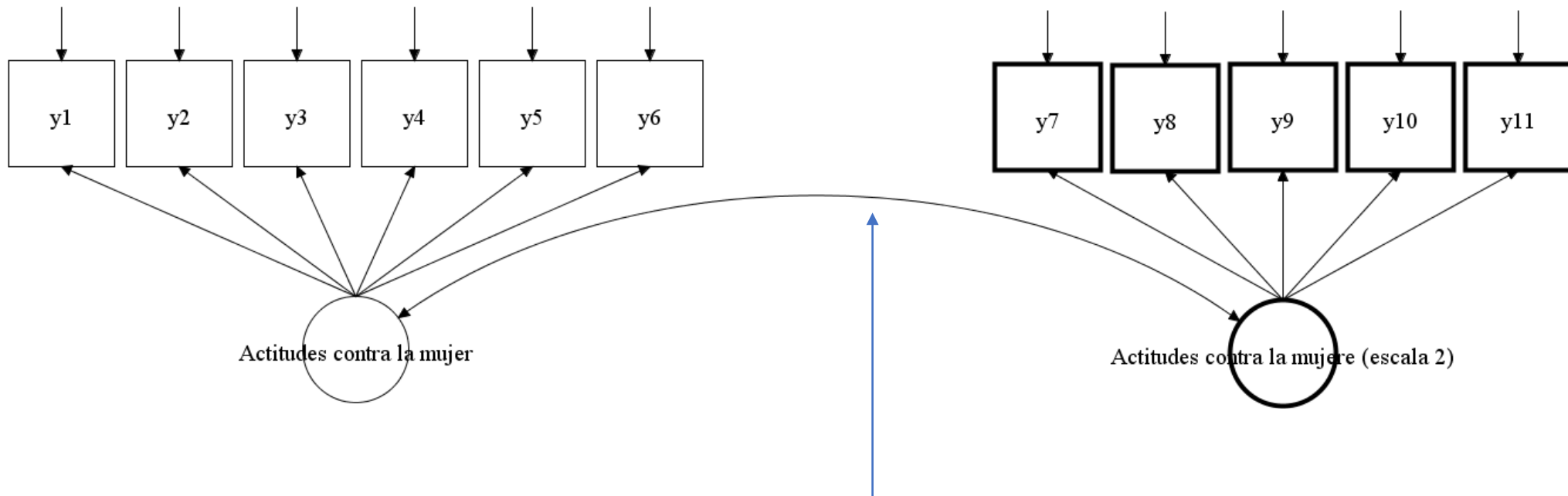
# Validación concurrente a la antigua

Antes se tomaban las correlaciones entre distintos componentes de diferentes tests

Si se dan cuenta es similar a confiabilidad porque se piensa que los componentes miden el mismo fenómeno

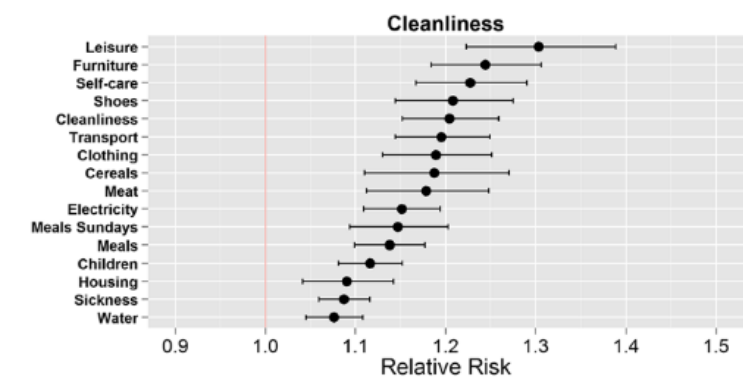
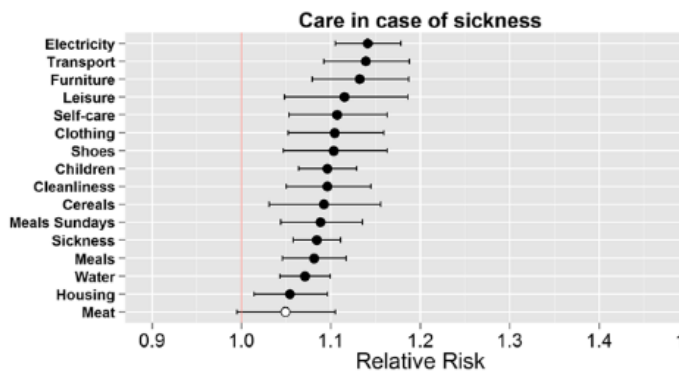
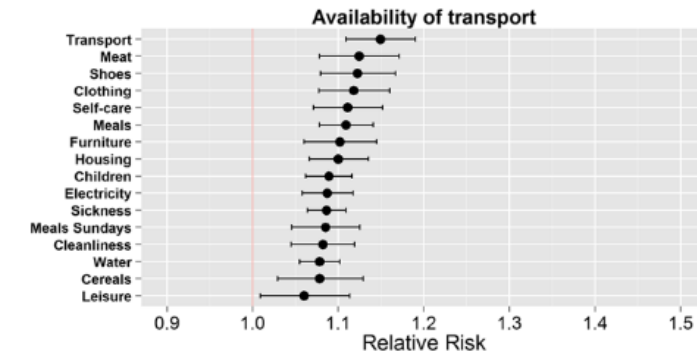
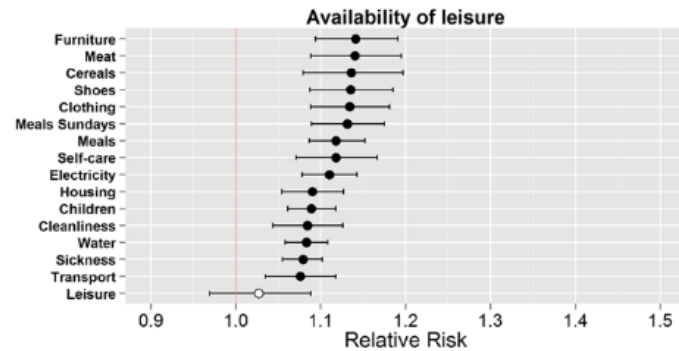
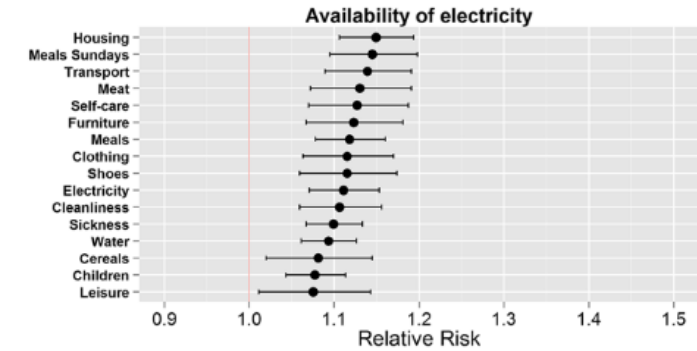
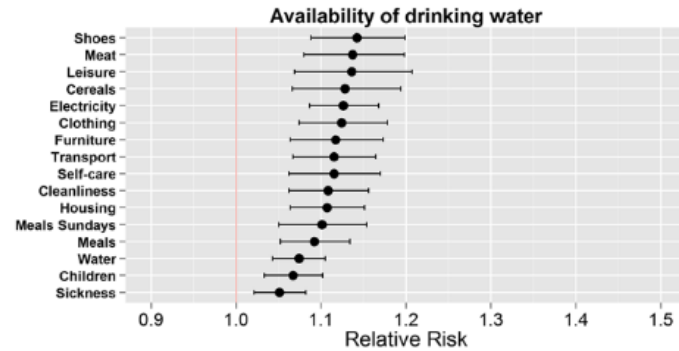


# SEM: Validación concurrente



Con un coeficiente podemos examinar la validación concurrente. PERO... Son latentes con latentes.

# Ejemplo: Validez concurrente



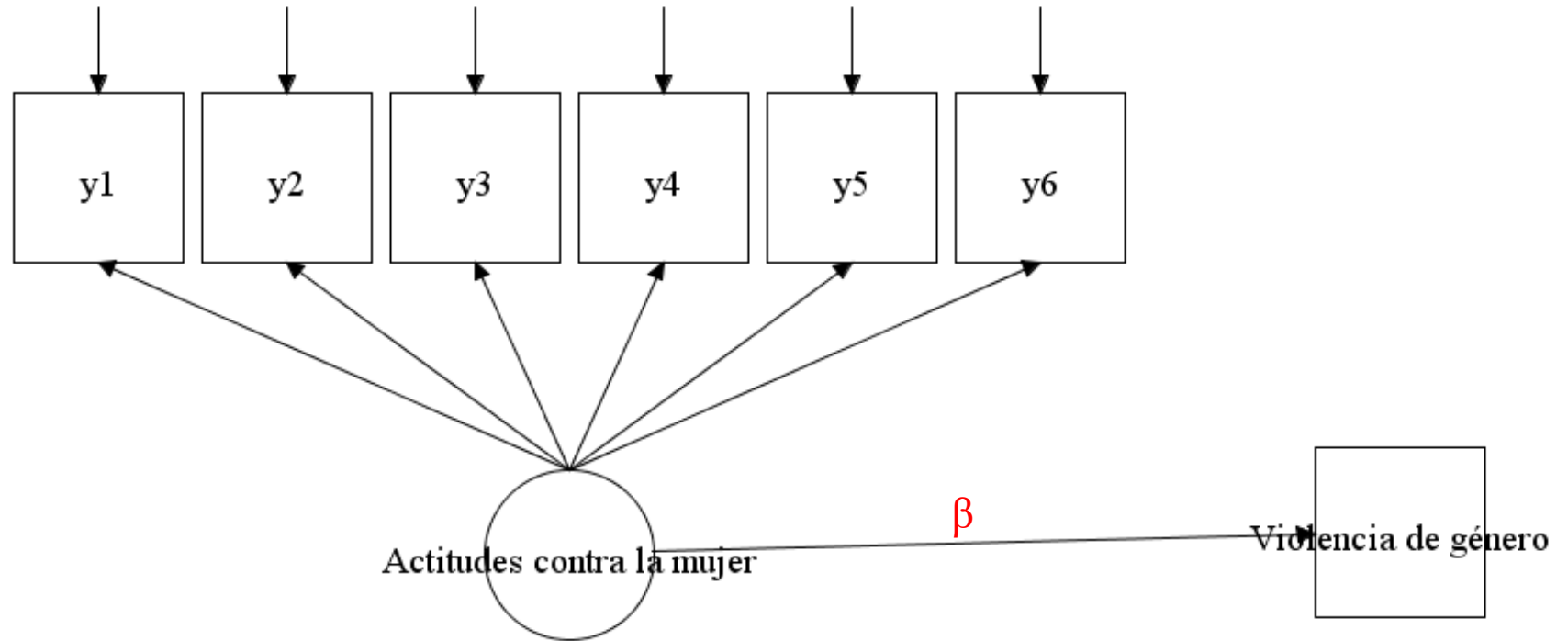
# Validez de constructo

- Durante décadas hubo varios tipos de validez
- En 1954 la sociedad americana de psicología enlistó los tipos de validez y añadió un cuarto: constructo
- En una serie de contribuciones Chronbach, Meehl y Messick anidaron todos los tipos de validez en la validez de constructo
- ¿Por qué?
- Al final todos los tipos de validez ven desde distintos ángulos si los scores pueden interpretarse como queremos

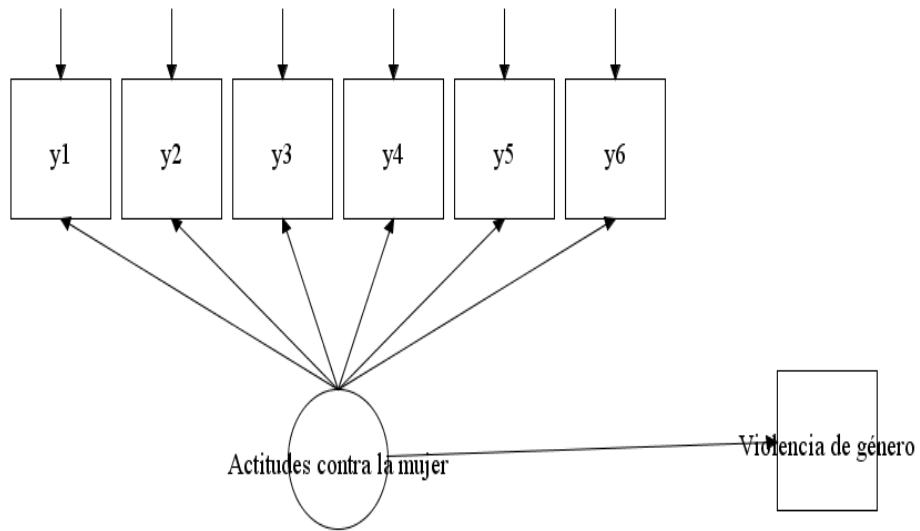


# Validez de constructo

No sólo se  
preocupa por  $\beta$   
sino por estructura  
de la medición

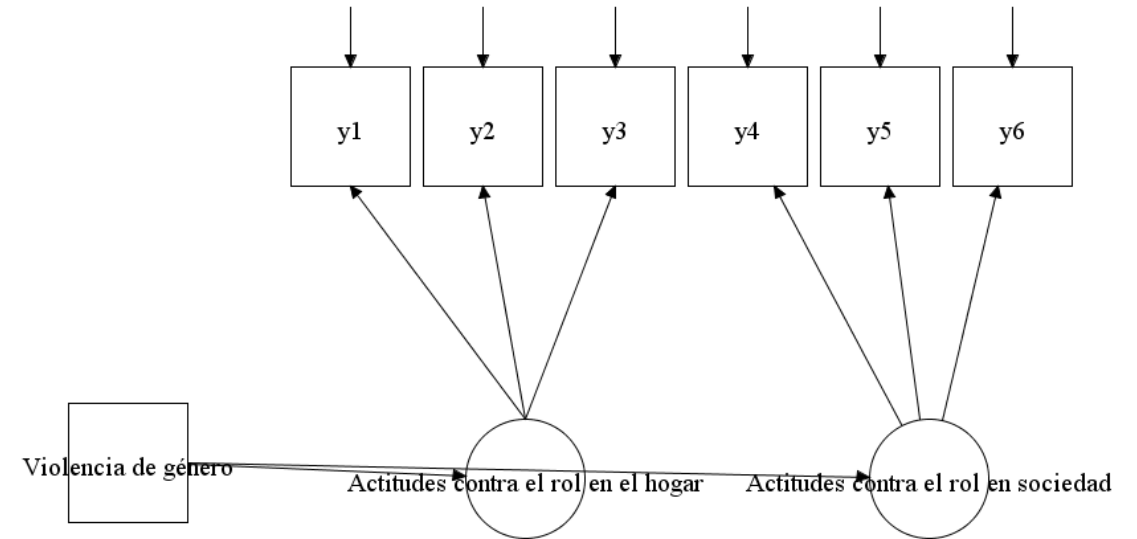


# Validez de constructo



TLI = .8  
RMSEA= .07  
CFI = .76

VS



TLI= .94  
RMSEA= .04  
CFI= .94

# Validez de constructo

Figure 2 UNDP Multidimensional Poverty Index Hierarchical Model and Weightings

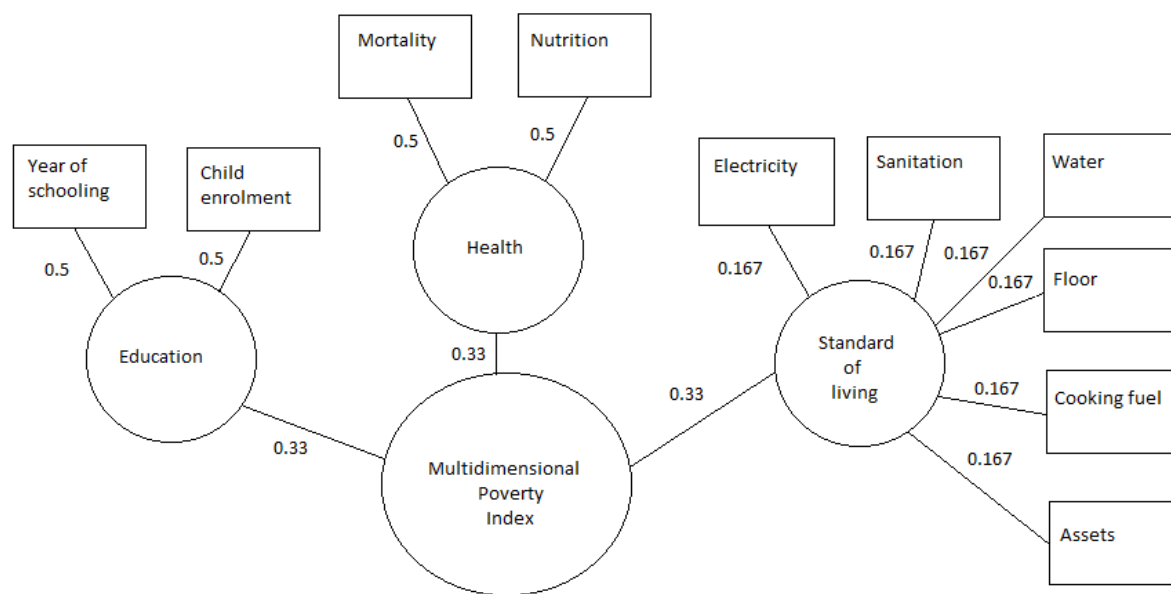


Table 3. Confirmatory Factor Analysis. Model Fit statistics

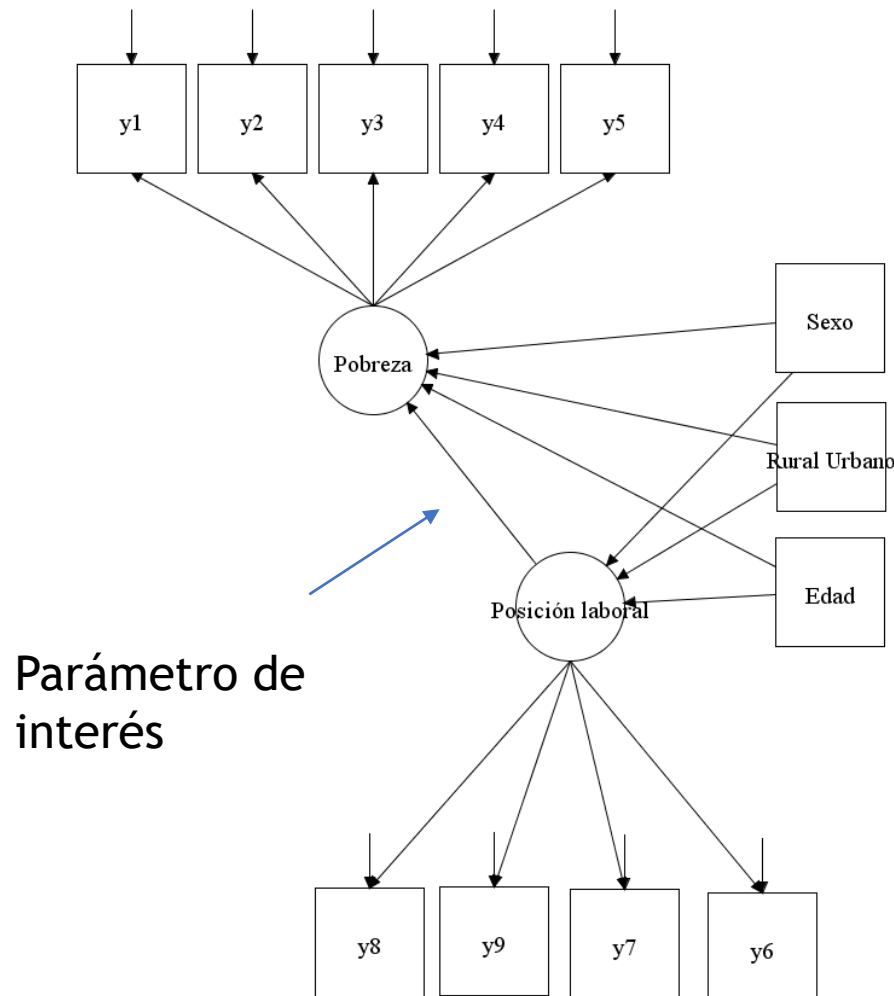
		H1	H2	H3	H4
Country	Statistic	Fixed item loadings	Fixed dimension loadings and free item loadings	Full free weights	Unidimensional
	TLI	0.73	0.94	0.85	0.94
Uganda	RMSEA	0.08	0.04	0.06	0.04
	TLI	0.78	0.83	0.68	0.97
Benin	RMSEA	0.06	0.06	0.08	0.02
	TLI	0.79	0.91	0.92	0.94
Cameroon	RMSEA	0.08	0.05	0.05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





# SEM y Validez

- La flexibilidad de SEM permite hacer en un paso modelos causales:



Si tengo un buen modelo teórico (*distinción entre causas y consecuencias*), puedo hacer ejercicios de validación más completos y unificados

MODEL:

Pobreza BY y1 y2 y3 y4 y5;

Posición laboral BY y6 y7 y8 y9;

Pobreza ON Posición laboral Sexo Edad Rural Urbano;

Posición laboral ON Sexo Edad Rural Urbano;



# Validez y complementariedad: EMSA y ELCSA

## Why identifying households by degree of food insecurity matters for policymaking

Rafael Pérez-Escamilla <sup>a</sup>  , Mireya Vilar-Compte <sup>b</sup> , Pablo Gaitan-Rossi <sup>b</sup> 


Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.jgfs.2020.100459>

[Get rights and content](#)

## A Model-Based Approach to Identify Classes and Respective Cutoffs of the Brazilian Household Food Insecurity Measurement Scale

Michael E Reichenheim, Gabriela S Interlenghi, Claudia L Moraes, Ana M Segall-Corrêa, Rafael Pérez-Escamilla, Rosana Salles-Costa 

*The Journal of Nutrition*, Volume 146, Issue 7, July 2016, Pages 1356–1364, <https://doi.org/10.3945/jn.116.231845>

Published: 08 June 2016 [Article history](#) 

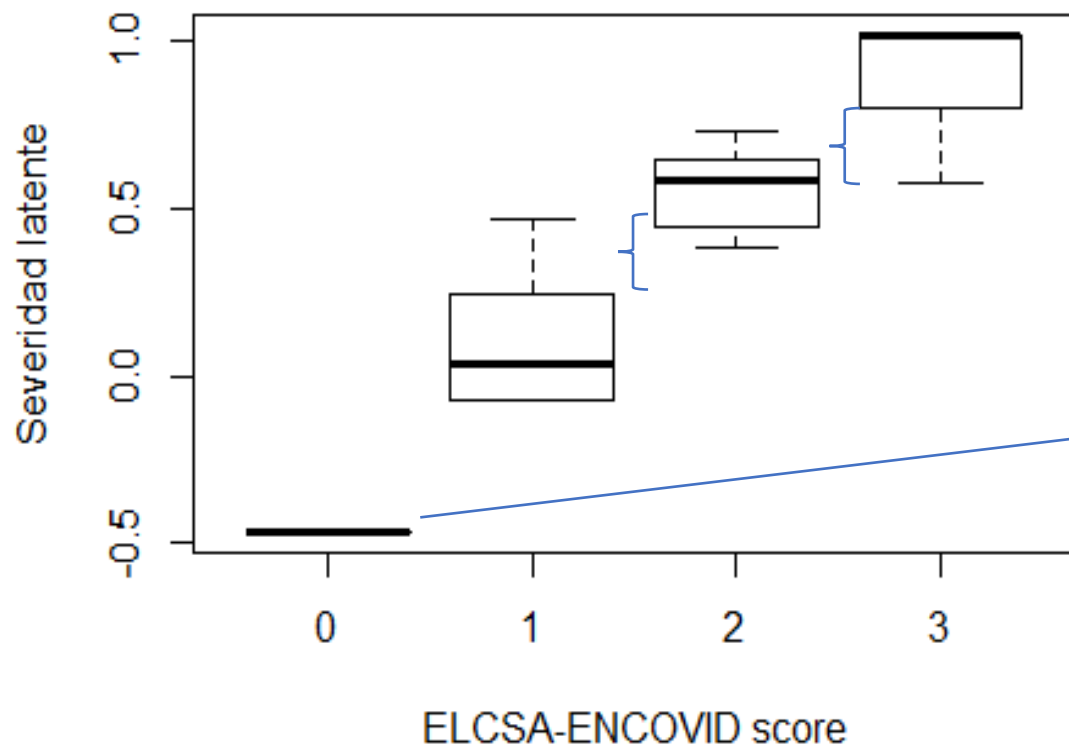
 PDF  Split View  Cite  Permissions  Share 

**Conclusion:** Although roughly classifying EBIA as in previous studies, the current approach suggests that, in terms of raw score, households endorsing only one item of the scale would be better classified by being placed in the same stratum as those remaining negative on all items.



# ¿Es grave que no se sostenga el Rasch?

Severidad latente por grupo. IRT Dos parámetros. Ad



No hay buena distinción en severidad de los grupos de inseguridad 1 (leve) y 2 (moderada)

El grupo de claramente seguros parece ser el primer grupo.

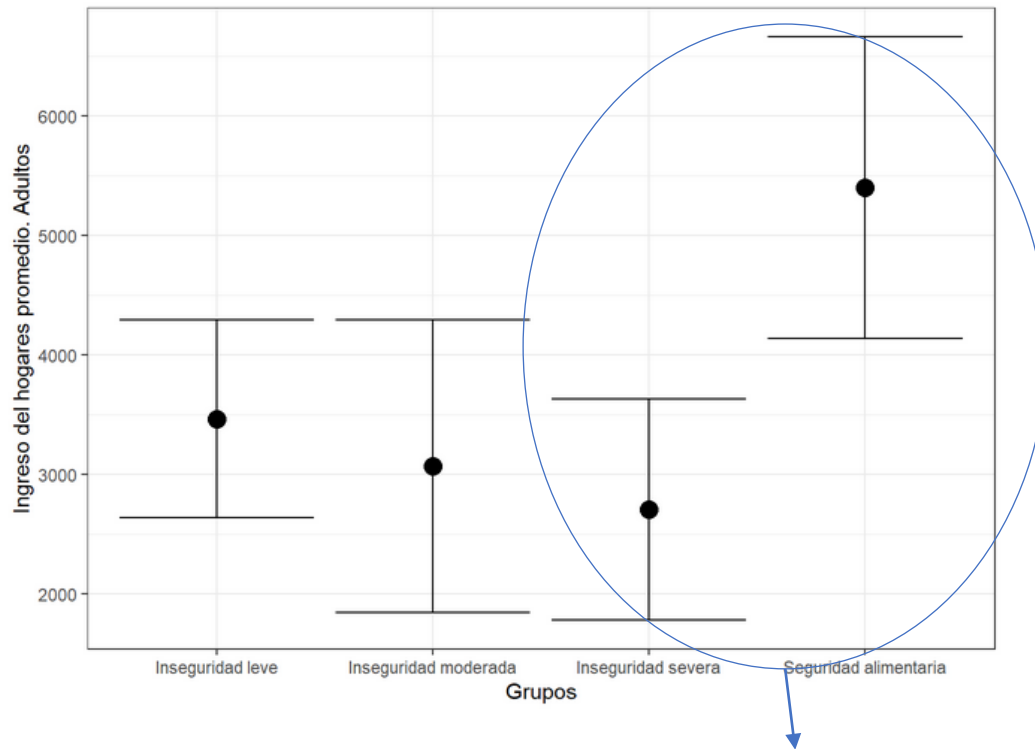
No parece que el grupo 1 sea próximo al 0 y, por tanto, que puedan agruparse!

Esto es grave si se hace la construcción manual de grupos y puntajes!



# Validez de criterio de los cuatro grupos

Validez de criterio de los 4 grupos



Las conclusiones sobre cambios en inseguridad alimentaria se basan en agrupar estos dos conjuntos.

El resultado es que nuevamente no parece haber evidencia que sustente a los 4 grupos.

En el artículo que acompaña la ENCOVID-19 hay conclusiones un tanto distintas

Sin embargo, esto ocurre con el ingresos de la ENCOVID-19 pero también (aunque de manera menos aguda) con la EMSA y la ENIGH.

# Nunca confundan validez con confiabilidad

## Validez estadística de la Escala Mexicana de Seguridad Alimentaria y la Escala Latinoamericana y Caribeña de Seguridad Alimentaria

Paloma Villagómez-Ornelas, MPD,<sup>(1)</sup> Pedro Hernández-López, L en Econ,<sup>(1)</sup> Brenda Carrasco-Enríquez, Act,<sup>(2)</sup> Karina Barrios-Sánchez, L en C de la Com,<sup>(1)</sup> Rafael Pérez-Escamilla, D en Nut,<sup>(3)</sup> Hugo Melgar-Quirón, D en C Méd.<sup>(4)</sup>

Scale (ELCSA). **Materials and methods.** Validity tests were conducted in order to verify that both scales were consistent instruments, conformed by independent, properly calibrated and adequately sorted items, arranged in a continuum of severity. The following tests were developed: sorting of items; Cronbach's alpha analysis; parallelism of prevalence curves; Rasch models; sensitivity analysis through mean differences' hypothesis test. **Results.** The tests showed that both scales



# Conclusiones

- La confiabilidad se preocupa por el tamaño del error aleatorio de los scores de una escala. Por tanto, la confiabilidad es el techo que tiene una escala, solamente puede empeorarse.
- La validez se encarga de identificar desviaciones sistemáticas en los scores -factores alternativos conocidos o por conocer-
- Validez es un tema de apilar evidencia a favor de la interpretación de los scores. No hay formas substitutas de validez sino complementarias.
- Validez me permite interpretar los scores de una escala con la intención que me propuse



# CONTACTO

Dr. Héctor Nájera  
Investigadores

Programa Universitario de Estudios del Desarrollo (PUED)  
Universidad Nacional Autónoma de México (UNAM)  
Antigua Unidad de Posgrado (costado sur de la Torre II de Humanidades), planta baja.  
Campus Central, Ciudad Universitaria, Ciudad de México, México.  
Tel. (+52) 55 5623 0222, Ext. 82613 y 82616

Tel. (+52) 55 5622 0889

Email: [hecatalan@hotmail.com](mailto:hecatalan@hotmail.com), [chuffman@colmex.mx](mailto:chuffman@colmex.mx)



# Referencias

- Bock, D. (1997). A brief history of item theory. *Educational measurement: issues and practice*, 16(4), 21-33.
- Lord, F. M., & Novick, M. R. (1968) Statistical Theories of Mental Test Scores. *Reading, Mass.: Addison-Wesley*.
- Ferguson, G. A. (1942). Item selection by the constant process. *Psychometrika*, 7(1), 19-29.
- Finney, D. J. (1944). The application of probit analysis to the results of mental tests. *Psychometrika*, 9(1), 31-39.
- Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. *British Journal. Of Mathematical and Statistical Psychology*, 42, 139-167.
- Goldstein, H. (2012). Francis Galton, measurement, psychometrics and social progress. *Assessment in Education: Principles, Policy & Practice*, 19, 147-158.
- Lawley, D. N. (1943). XXIII.—On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 61(3), 273-287.
- Markon, K. E. (2018). Reconciling information and reliability in scaling local measurement precision: A comment on O'Connor (2017).
- Nájera Catalán, H. E., & Gordon, D. (2020). 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*, 56(9), 1763-1783.
- O'Connor, B. P. (2018). An illustration of the effects of fluctuations in test information on measurement error, the attenuation of effect sizes, and diagnostic reliability. *Psychological assessment*, 30(8), 991.
- O'Connor, B. P. (2018). Clarifications regarding test information and reliability, and new methods for estimating attenuation due to measurement error: Reply to Markon (2018).
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of educational psychology*, 16(7), 433.

