Comparability in poverty measurement

Hector Najera

- Making comparisons as a reason for measuring
- Two perspectives on comparability
- Challenges in comparability over time and comparison between groups
- Solutions and methods for comparability
- Limits for comparability analysis

Analysis of groups: Populations, territorial units and between societies

Is poverty higher in unit A than in unit B?

Analysis over time: Trends and changes

Did poverty increase, decrease, or remain the same?

Combinations: Changes in time by group.

- Comparability: Exercise that consists of contrasting equivalent scores between different units

-

- Disaggregation: Action or operation of separation of scores -equivalent or not equivalent- given another variable.

- Being able to disaggregate an index does not make it comparable (Axiomatic framework?)

-

   What allows you to have equivalent scores?
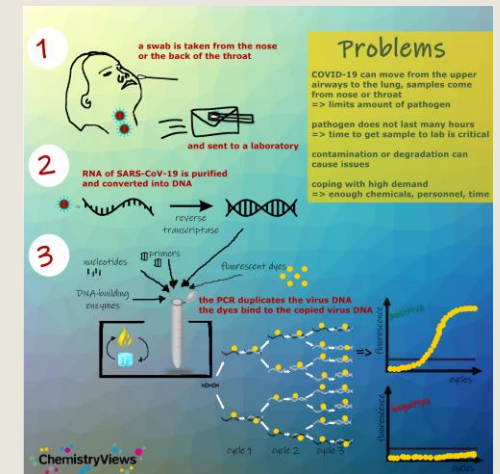
Equivalent scores?

- They are those scores that, regardless of group membership, reflect differences in poverty

-

  - Admission test scores should reflect mathematical skills not if the person comes from a certain social stratum.
  - It is one thing that the social stratum can be a cause of certain educational advantages and another that the exam is biased and favours the syllabuses of certain privileged schools.
  - Effective differences in poverty, not differences by place of residence: The palm roof of Mayan houses in Mexico could penalize rural areas at the expense of urban criteria –such kind of roof is really efficient in hot climate-

Two visions: Equivalent scores

Formative measurement
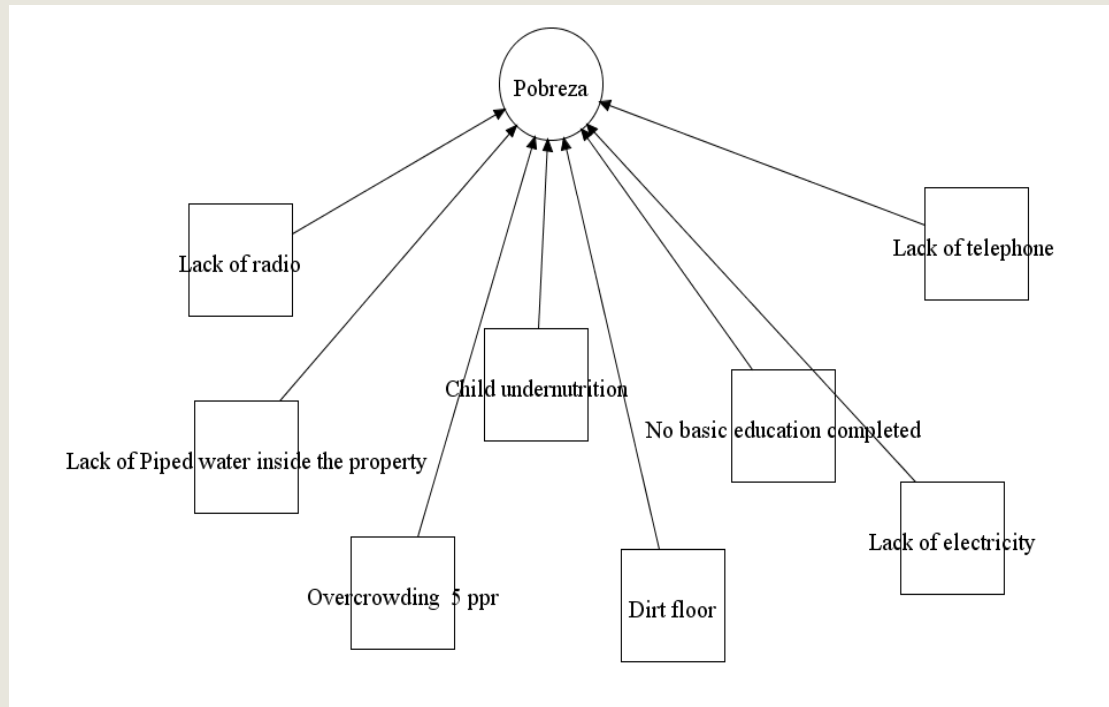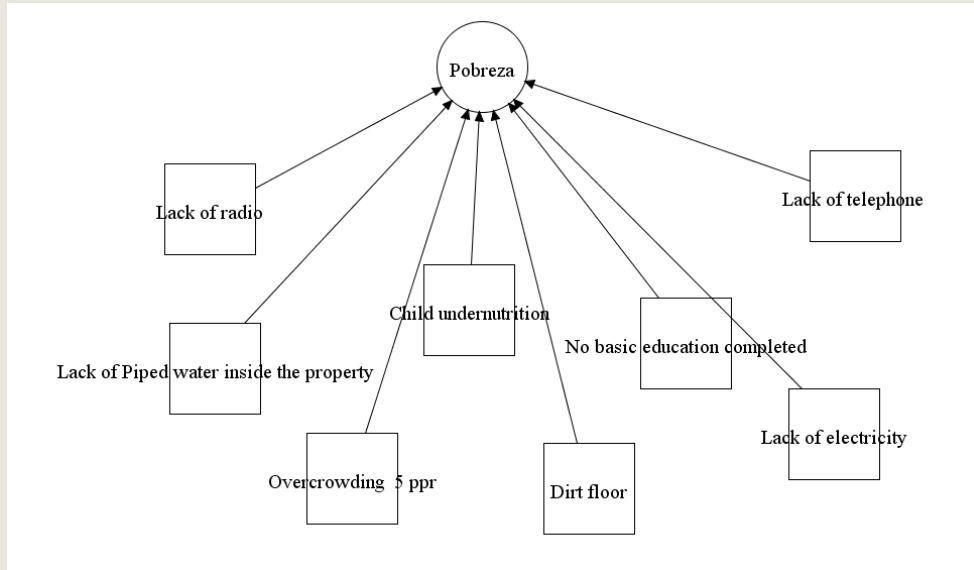

Reflective measurement
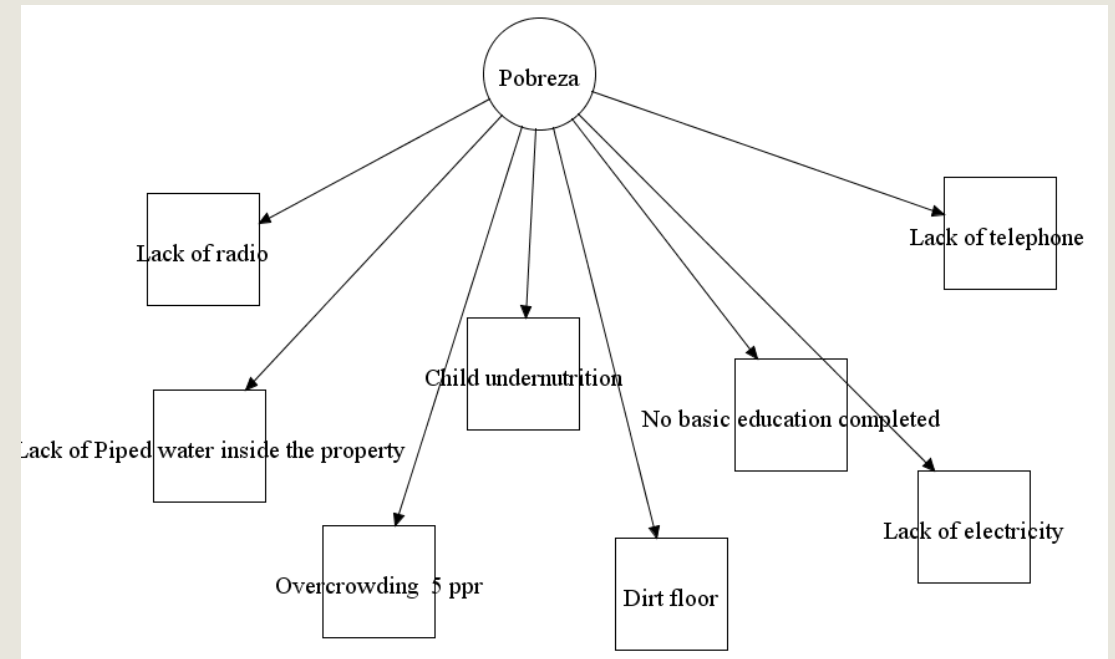
The indicators construct the concept, poverty is what is there in those indicators. What happens if I drop one? Does the definition change?

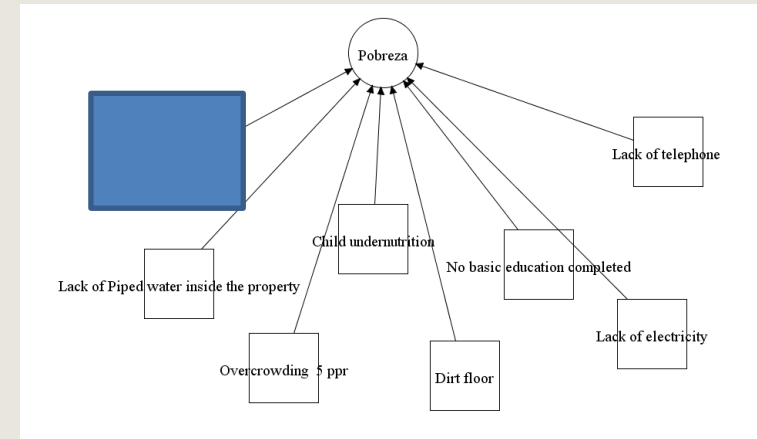The latent phenomenon is the cause of the manifestations of poverty

Comparability of formative measurement

Not associated with group/time poverty 2
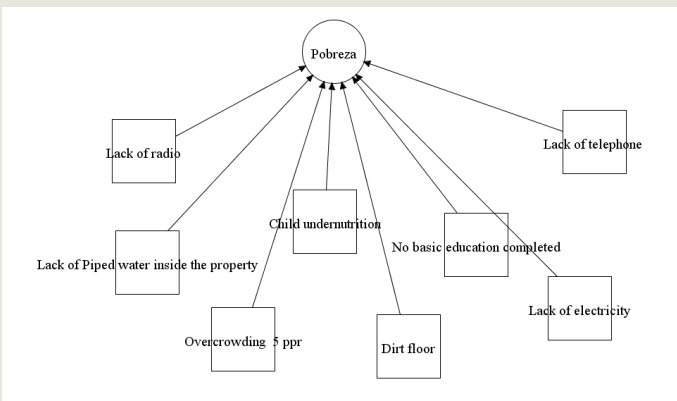
Set I_f
Scores S_{If} {0-8}

Set I_{f2}
Scores S_{If2} {0-7}

Under this formative framework, comparability is easily broken. The Score is no longer equivalent
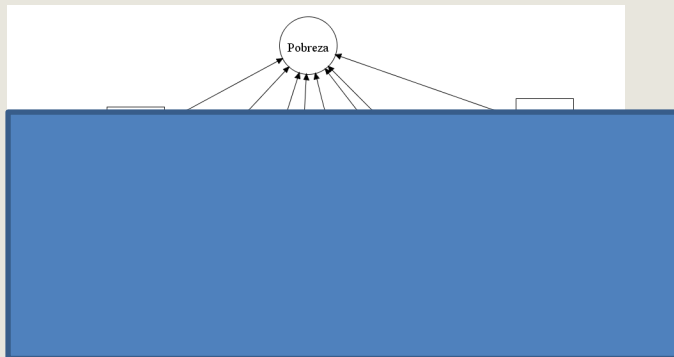
Comparability in formative measurement

$$Set\ I_f$$
$$Scores\ S_{If}\ \{0\text{-}8\}$$

We know that $S_{If} \geq S_{If2}$

This would work if we had a census of all dimensions and all indicators of poverty but we don't have it and we'll never have it!

Indicators most likely do not serve to identify who is poor

$$Set\ I_{f2}$$
$$Scores\ S_{If2}\ \{0\}$$

This means that those needs improved.

¿La pobreza es cero?

Typical case where the politician thinks that poverty is solved thinking that the indicators must be solved!
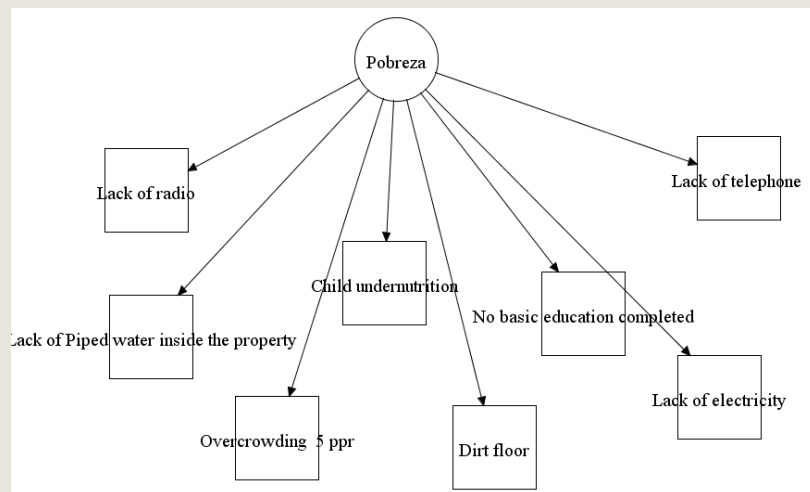
Eventually I need new indicators.

The formative approach is a dead-end trap

Formative measurement is an especial case – not very notable of the reflective one.

Seeking comparability at the expense of the same indicators will not take you very far
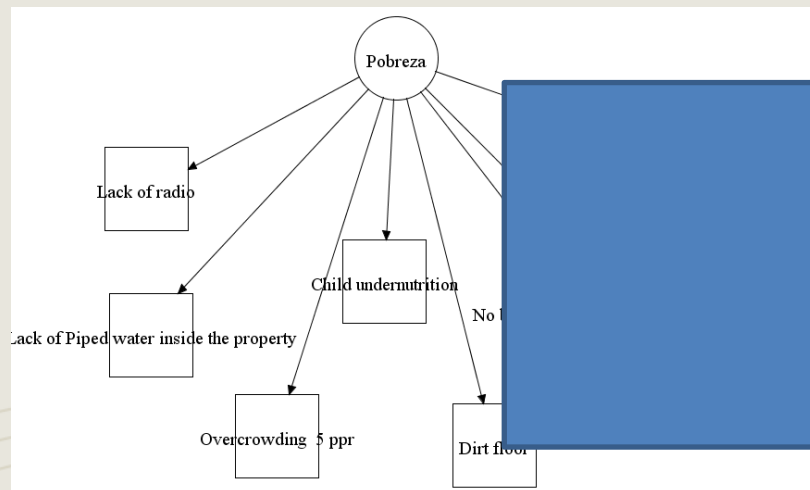
Set I_r

Scores: S_r

Set I_{r2}

Scores: S_{r2}

Remember that what we have samples of D, I, U, P and L

The reflective approach does have an inference framework for dealing with the comparability problem.

Is it S_r = S_r2? –They are equivalent-

Is the model equivalent and its scores too?

If it is not equivalent? What could I do?

Set I_r

Scores: S_r
{0,8}

Set I_{r2}

Scores: S_{r2}
{0,9}

Remember that what we have samples of D, I, U, P and L

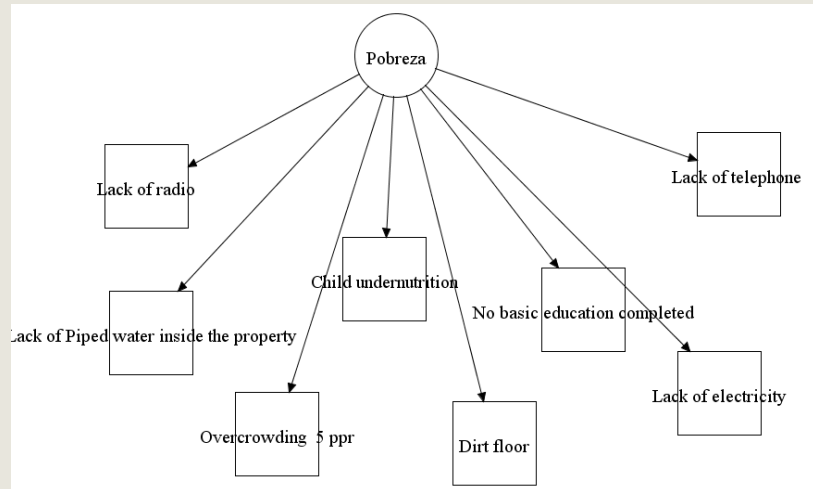The reflective approach does have an inference framework for dealing with the comparability problem.

Is it S_r = S_r2? –They are equivalent-

What does S_r=3 mean in terms of S_r2=4?

Is the model equivalent and its scores too?

If it is not equivalent? What could I do?

Comparability as an invariant measurement – measurement invariance-

- Measurement invariance/equivalence

  – If reliability and validity do not hold, there is no point on undertaking comparisons

  – Measurement invariance is a necessary condition to be able to make valid comparisons between groups.

  – Occurs when a number of conditions are met

Published: December 1993

Measurement invariance, factor analysis and factorial invariance

William Meredith

Psychometrika 58, 525–543 (1993) | Cite this article

5764 Accesses | 2606 Citations | 6 Altmetric | Metrics

PSYCHOMETRIKA—VOL. 57, NO. 2, 289–311
JUNE 1992

ON THE MISUSE OF MANIFEST VARIABLES IN THE DETECTION OF MEASUREMENT BIAS
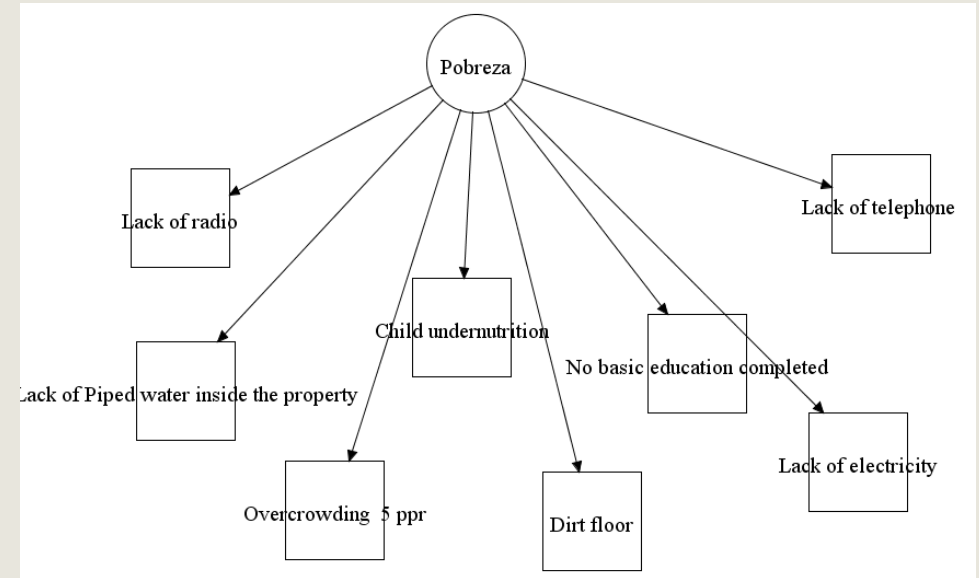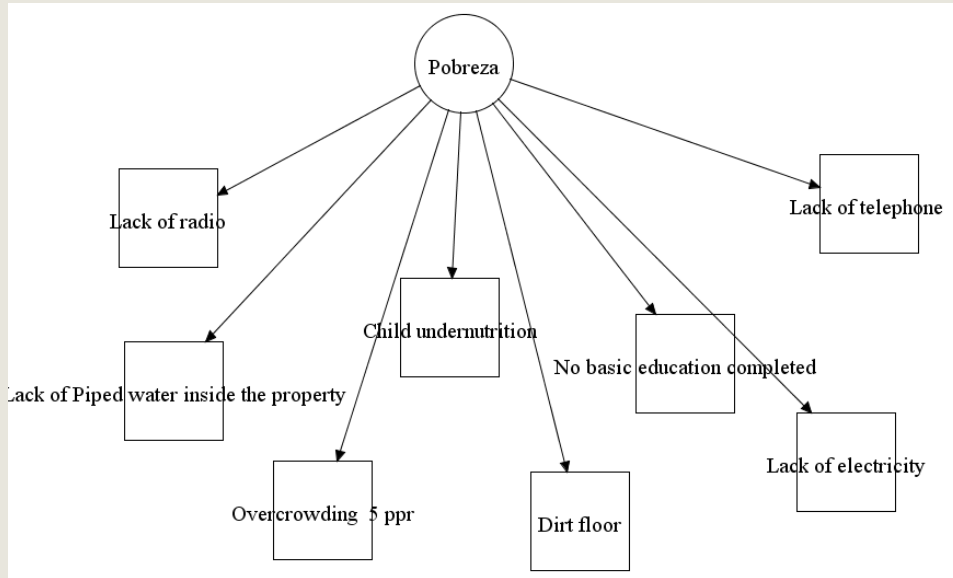
WILLIAM MEREDITH

UNIVERSITY OF CALIFORNIA, BERKELEY

ROGER E. MILLSAP

BARUCH COLLEGE, THE CITY UNIVERSITY OF NEW YORK

Case 1: Same indicators

Are the models equivalent for two units? –i.e. the model works in both units.
Are factorial loads equivalent?
Are those ordered at origin equivalent?
Are the variances of the residuals equivalent?

→ Weak invariance

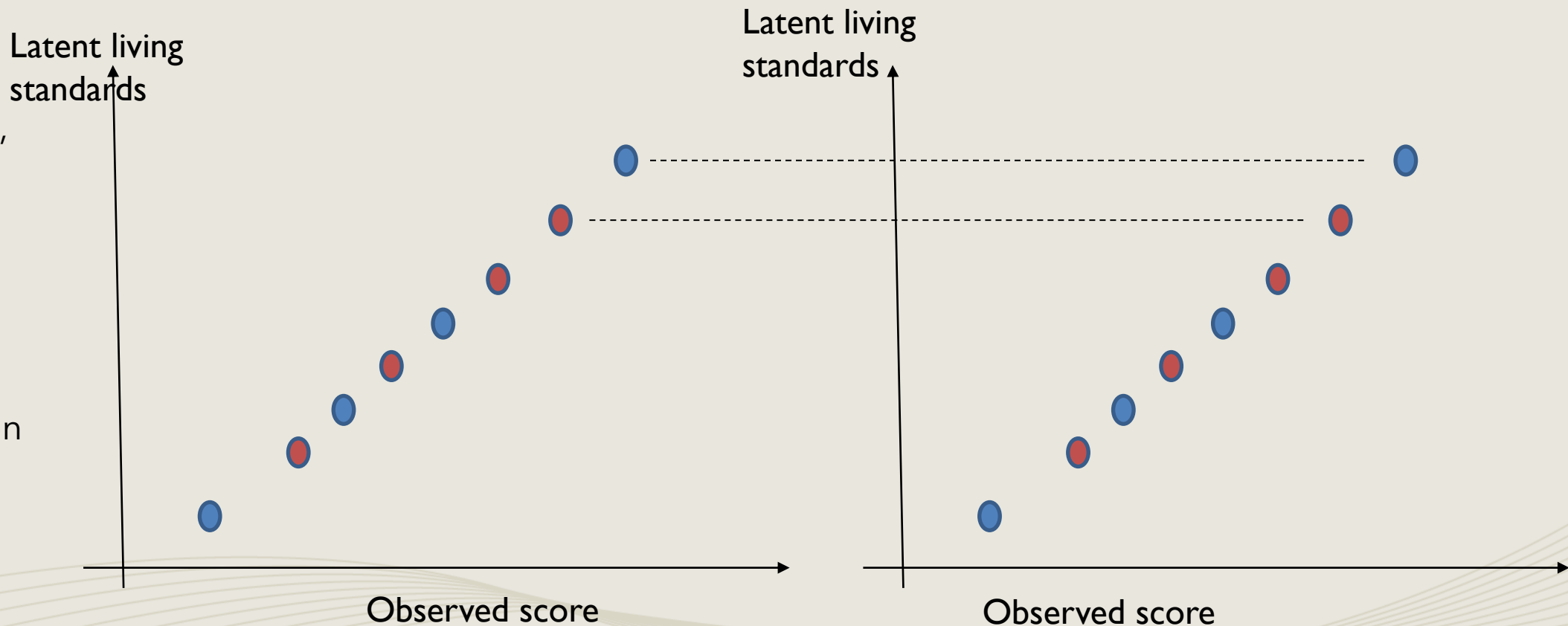→ Strong invariance

→ Strict invariance

Strict invariance

Strict invariance implies that both the residuals, intercepts and the slopes are the same between groups.

When it holds, the scores are comparable between groups.
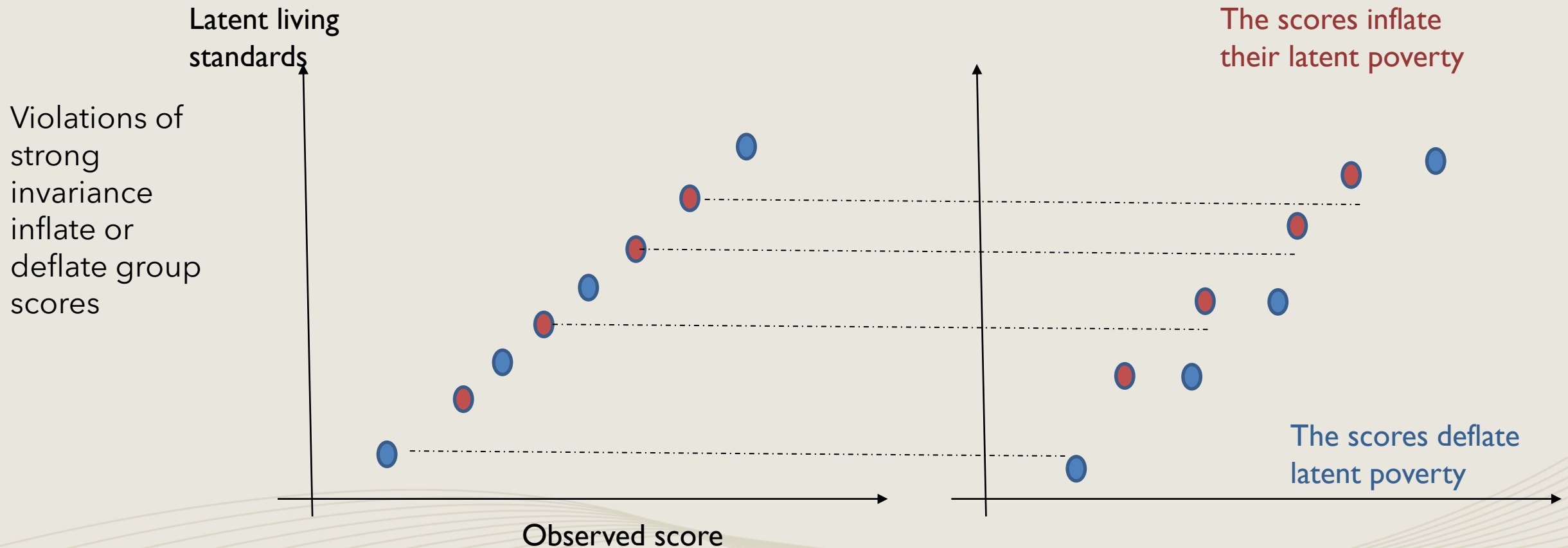
I can do disaggregations

Violations of strong invariance

Violations of strong invariance involve ordering changes in the intercepts between groups. Some are inflated while others are deflated.

Latent living standards

Violations of strong invariance inflate or deflate group scores

The scores inflate their latent poverty

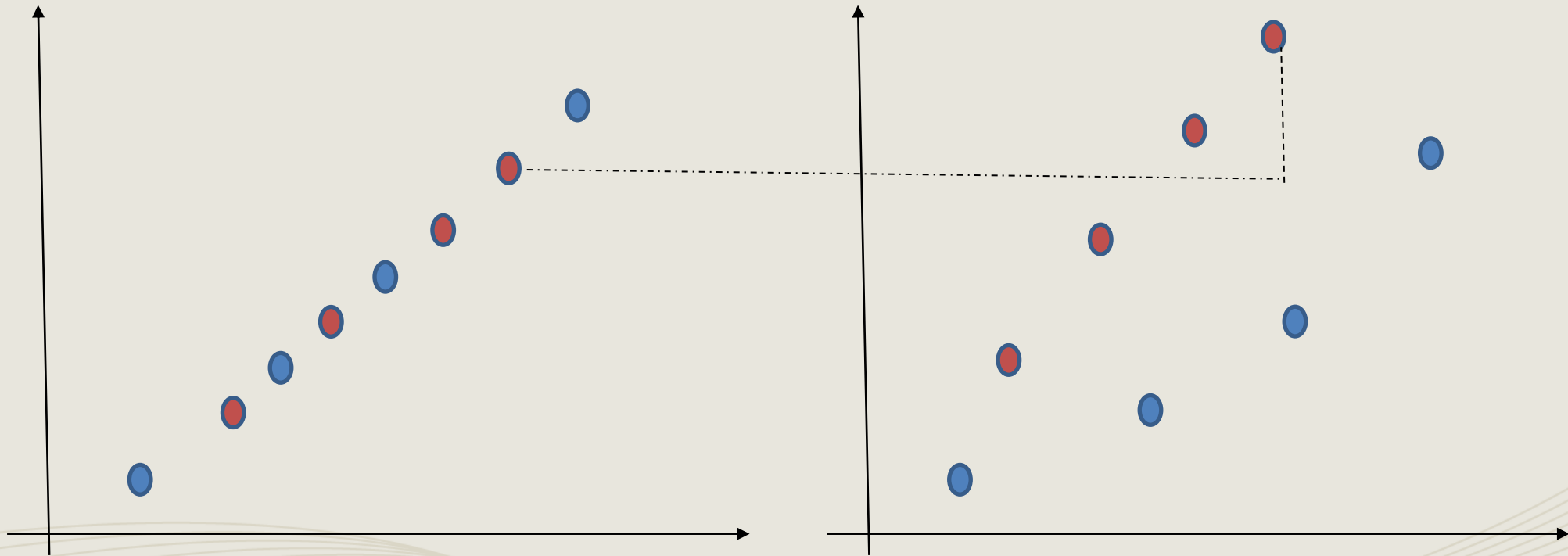The scores deflate latent poverty

Observed score

Violaciones de Invarianza débil

Weak invariance is violated when factor loadings are different between groups

The bias gets worse

Violations of weak invariance involve deviations in scores between groups

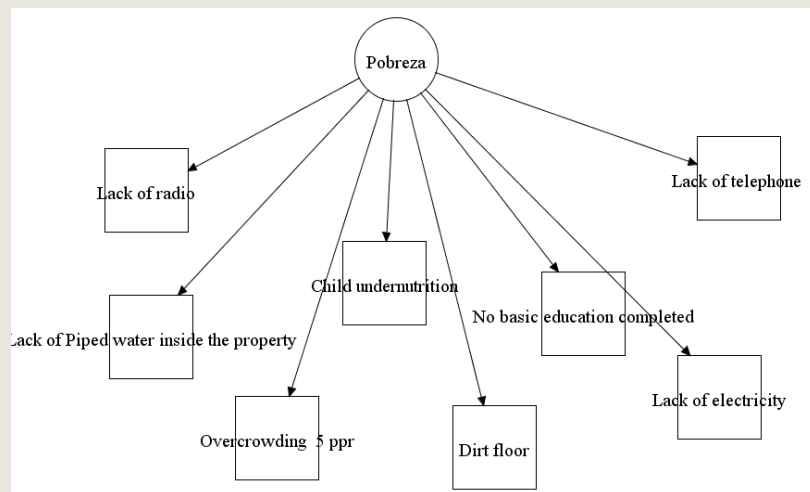Puede haber violaciones de pendientes y no de ordenadas al origen!

- Under schemes of equivalent models with the same structures and variables, the analysis of invariance allows me to know if the comparisons are fair between units

- Invariance violations will make comparisons more difficult for me because they will be biased

- If the violations of invariance are serious – i.e. weak invariance does not hold, even partially – I have to modify the scale to make comparisons over time or between groups.

- This is similar to the case of formative measurement, however, in the reflective framework there are alternatives, i.e. the definition does not change

Set I_r

Scores: S_r

Set I_{r2}

Scores: S_{r2}

If the strict invariance is met – but even the strong one partially – I can make comparisons.

Under strict invariance the subset S_r= subset S_r

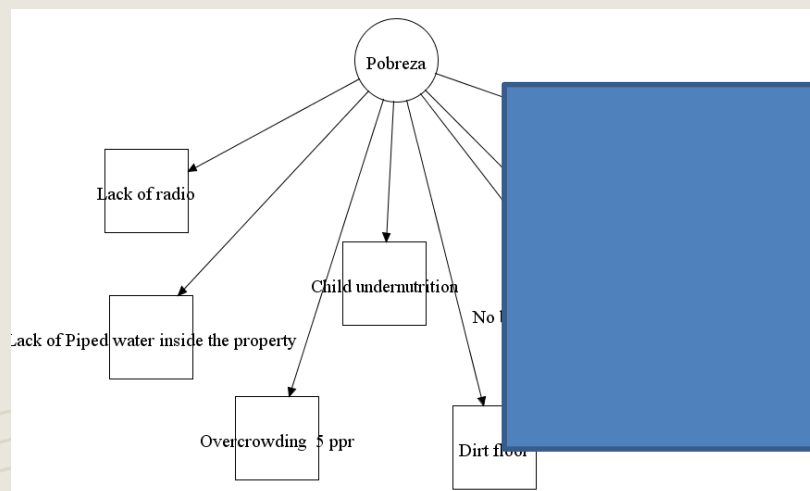If I want to preserve the original scale The question is what does S_r = 2 mean with respect to S_r = 3.
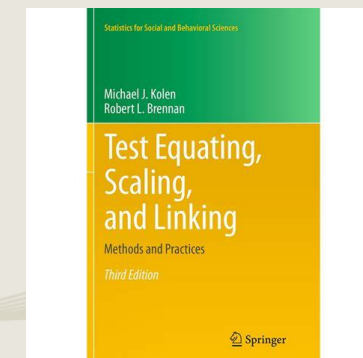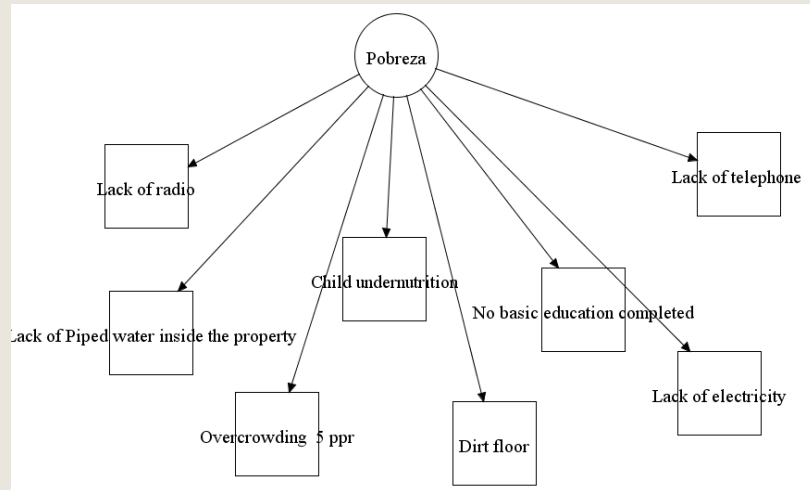
"Test-equating and linking"

Set I_r

Scores: S_r
{0,8}

Set I_{r2}

Scores: S_{r2}
{0,9}

First option:

Identify invariant items -weak and strong- and adjust the total scores from these items (anchors) -scale equating-

Second option:

1. Hypothesis: Items replace previous items – This is useful for indicators of the same dimension or for threshold updates–
2. Using invariance analysis

3. The total scores are self-adjusting – although we are still 0-8 vs 0-9

First option:

Identify invariant items –weak and strong– and adjust the total scores from these items (anchors)

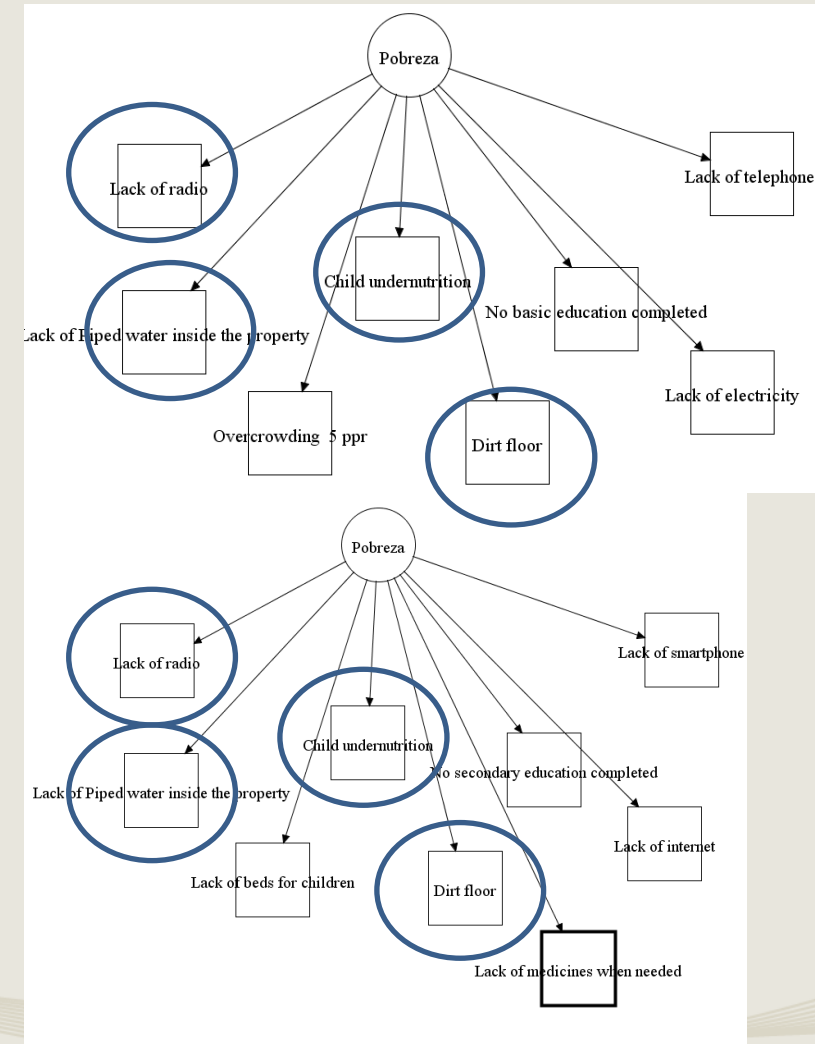I can evaluate the second model from the information I have from the first.

This evaluation is in terms of severity, which helps me to adjust the total scores

Conditions: A large number of anchors are needed. 60% of the scale, more than 10?
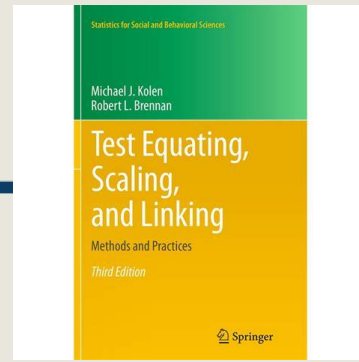
Test equating and scaling is a procedure that allows you to transform scores and put them in the same terms.

First option:

Identify invariant items –weak and strong– and adjust the total scores from these items (anchors)

In principle, a good number of anchors are needed - more than 60% of the scale.
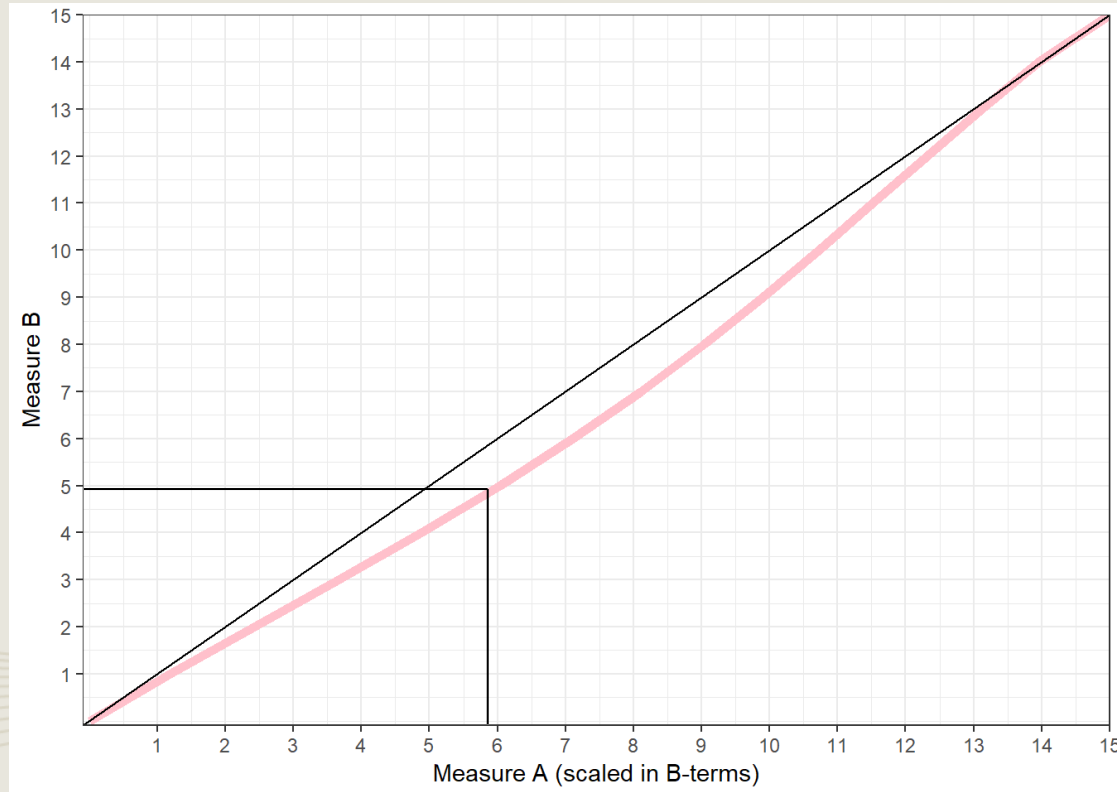
What´s a deprivation score of 6 in terms of the previous measure.

It can be done the other way around!

Second option: Replaceable items

Are they equivalent? –Same values of the parameters: severity and discrimination-

If they are, then I can make simple replacements – as in the Education tests.

The question of differences in total scores would be addressed with

¿Si quiero comparar sociedades con diferentes indicadores?
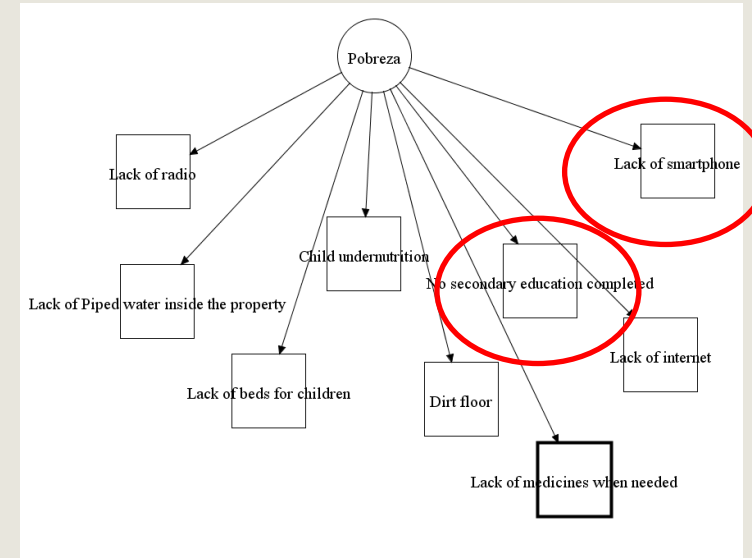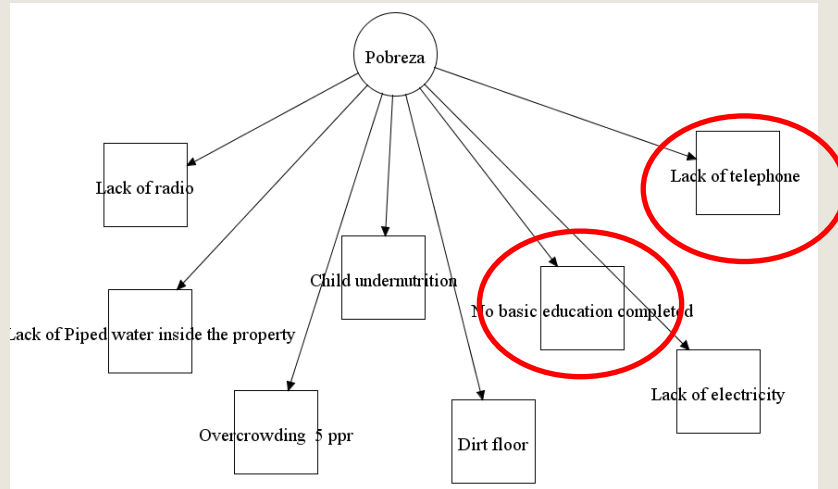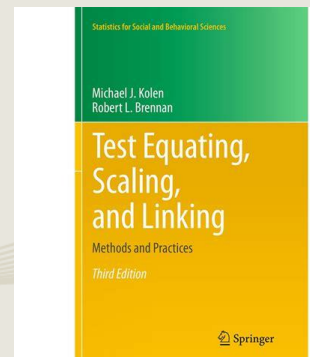
- Si la pobreza es relativa a cada sociedad, las prevalencias simples nos dirían cuál es la pobreza en dicha sociedad en sus propios términos. Esto a muchas veces es suficiente

- Si quiero hacer comparaciones con un referente, la situación cambia. Se trata de otra pregunta.
  - ¿Cuál es la severidad de la pobreza de Uganda en términos de la severidad de la pobreza en México?
  - ¿Cuál es la severidad de la pobreza en México en términos de la severidad de la pobreza en Uganda?
  - ¿Cuál es la severidad de la pobreza en determinado país condicional en un referente mundial de pobreza?

El tratamiento es similar al anterior.

Invarianza y scale-equating

- Not many!
-
- Comparison of poverty in Europe under a common European standard of needs

- Given a set of indicators we would like to compare countries using the same standard, i.e. poverty in Bulgaria is higher than in Norways under the same standard

- It is not the use of the same indicators but the use of equivalent measurement models!
-

Revising the EU material deprivation variables

ANNE-CATHERINE GUIO, DAVID GORDON, HECTOR NAJERA, MARCO POMATI

2017 edition

STATISTICAL WORKING PAPERS | eurostat
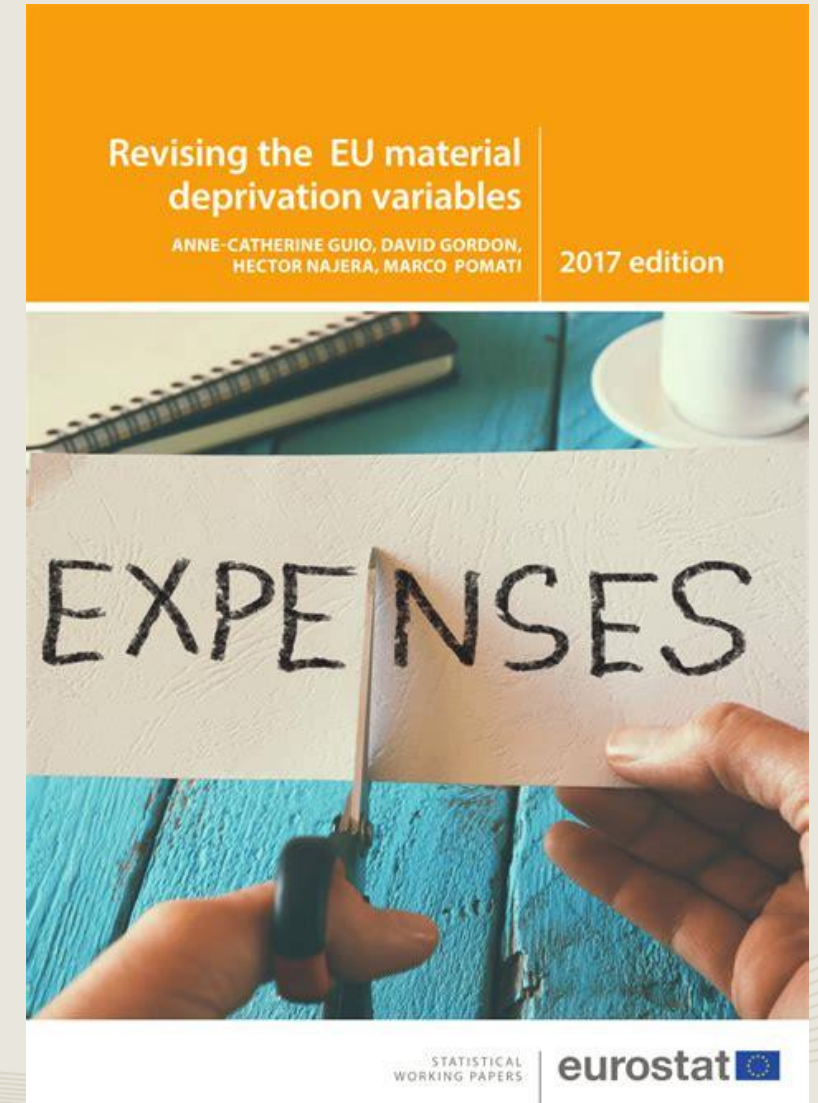
Approximate invariance

Measurement Invariance

Alignment

- Billet, J., Meeusen, C. & Abts, K. (2021). The relationship between (sub)national identity, citizenship conceptions, and perceived ethnic threat in Flanders and Wallonia for the period 1995-2020: A measurement invariance testing strategy. Forthcoming in Frontiers in Political Science. DOI: 10.3389/fpos.2021.676551

- DeMars, C.E. (2019). Alignment as an alternative to anchor purification in DIF analyses. Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2019.1617151

Special Issue: Measurement Invariance - Sociological Methods & Research

- Cieciuch, J., Davidov, E., Algesheimer, R. & Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, 47:4 665-686. DOI: 10.1177/0049124117701478

- Davidov, E., Dülmer, H., Cieciuch, J., Kuntz, A., Seddig, D. & Schmidt, P. (2018). Explaining measurement nonequivalence using multilevel structure equation modeling: The case of attitudes toward citizenship rights. Sociological Methods & Research, 47:4 729-760. DOI: 10.1177/0049124116672678

- Munck, I., Barber, C., & Torney-Purta, J. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009. Sociological Methods & Research, 47:4 687-728. DOI: 10.1177/0049124117729691

- Muthén, B. & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. Sociological Methods & Research, 47:4 637-664. DOI: 10.1177/0049124117701488 Mplus scripts.

Figure 9: Relationship between adjusted deprivation scores and the deprivation rate (proportion of people lacking at least 5 items out of 13), 2014
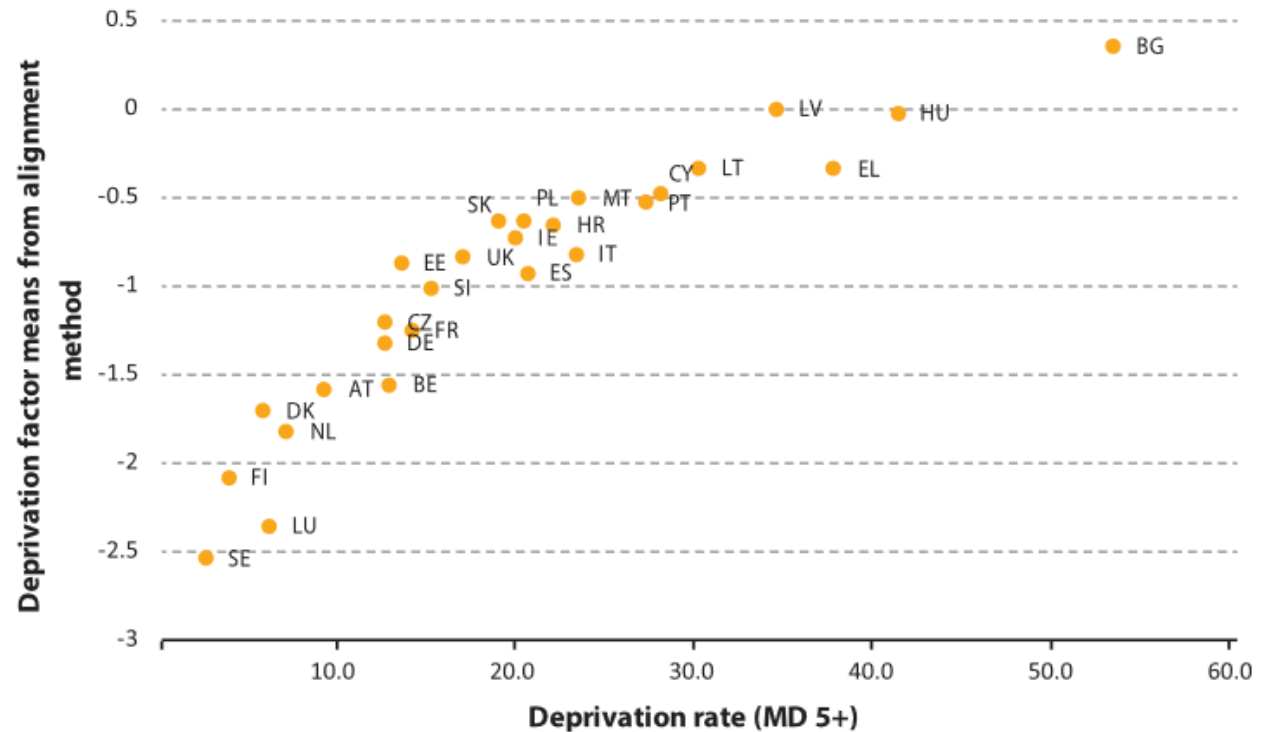
Note: see Annex 1 for country abbreviations.

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

Applications? Updating thresholds

Does changing the thresholds lead to violations of MI?

Table 2

Percentage of the population deprived in relation to the given item. Mexico, 2012

	Index A (official thresholds)	Index B (adjusted)
Education	19	19
Food deprivation	23	44
Minimum social protection floor	62	62
Flooring material	4	4
Roofing material	2	25
Walling material	2	14
Overcrowding	10	10
Access to water	9	47
Sanitation	9	40
Fuel	13	13

Source Estimations based on INEGI-CONEVAL (2012)
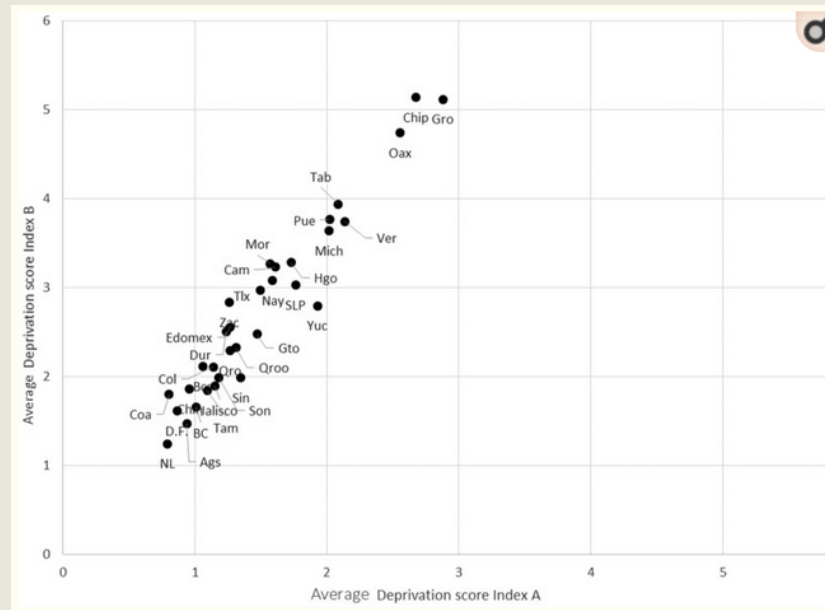
Yes, updating the thresholds improves the mexican scale.

It affects invariance but midly

Could it be posible to use scale equating?

Yes but it is posible to find partial strong invariance. I guess this Will be stronger for the most recent data.

Conclusiones

- La comparabilidad es uno de los principales objetivos en medición
- Hay dos formas de pensar comparabilidad pero una es un callejón sin salida
- Hay que hacer cambios/actualizaciones para preservar la comparabilidad no para romperla!
- Estadísticamente hablando hay un marco de referencia claro para hacerlo pero poca investigación al respecto
- Los métodos son muy nuevos y algunas de las propiedades aún se desconoce.