



# What is Multilevel Structural Equation Modelling?

Nick Shryane

*Social Statistics Discipline Area  
University of Manchester*

*[nick.shryane@manchester.ac.uk](mailto:nick.shryane@manchester.ac.uk)*

# What is multilevel SEM?

## 1. What is SEM?

- A family of statistical models, subsuming
  - *Regression analysis*
  - *Path analysis*
  - *Factor analysis*

## 2. What is multilevel SEM?

- As above, but expanded to deal with the problem of hierarchically clustered data; children clustered in schools, schools clustered in regions, etc.

## 3. We'll review 1 before going on to 2.

# Regression analysis

- Regression uses information from a set of variables to predict the value of another variable.

# Regression analysis

'Predictors'

Deprivation  
Age  
Sex  
Social Class  
Motivation  
Aptitude

'outcome'

Grade



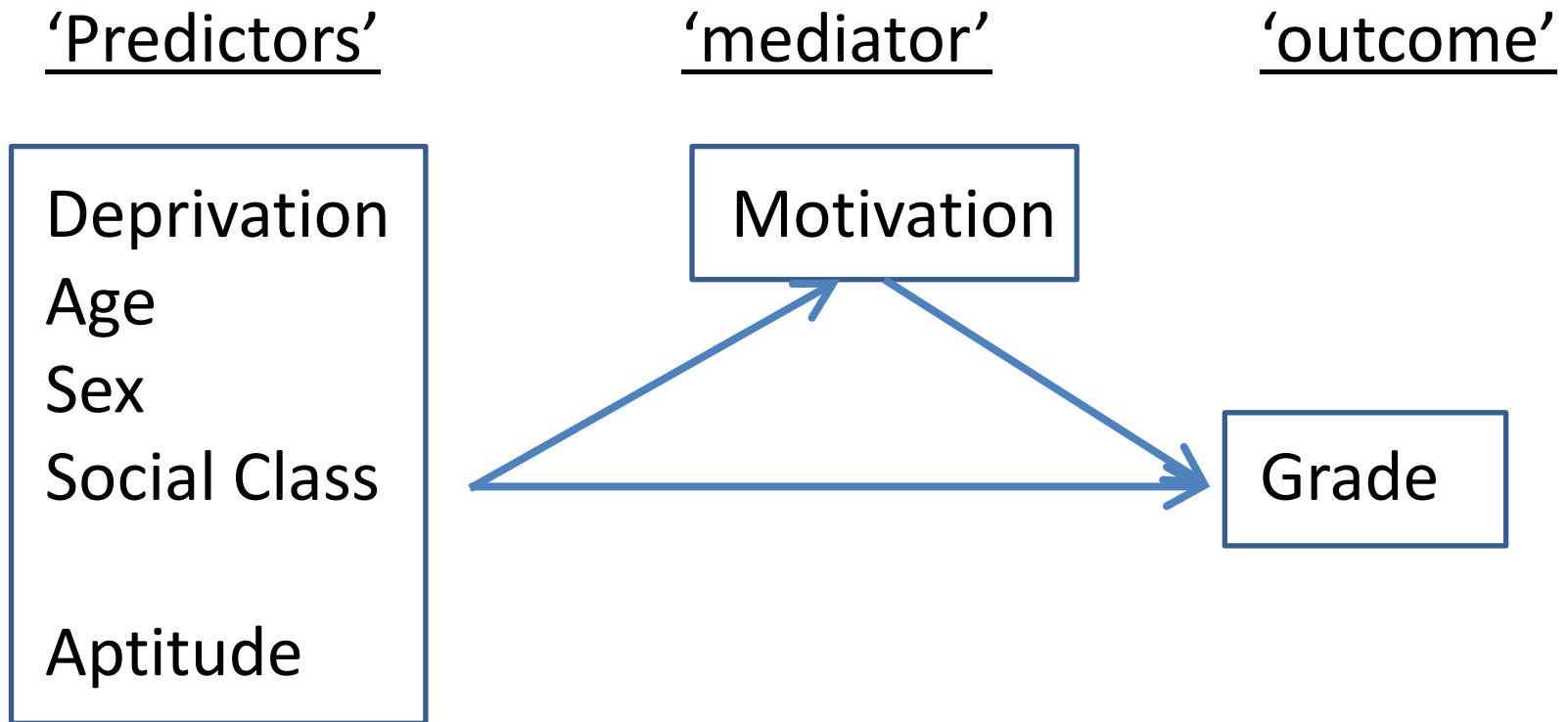
# Regression analysis

- In the extreme, this is a bit like an ‘actuarial’ model
  - The predictors provide information about *what* the expected value of the outcome is, but do not give any information about *why* or *how*.

# Path analysis

- An extension of regression analysis
- Includes several regression model 'joined together'
- Can be used to assess possible '*how*' questions.

# Path analysis - mediation

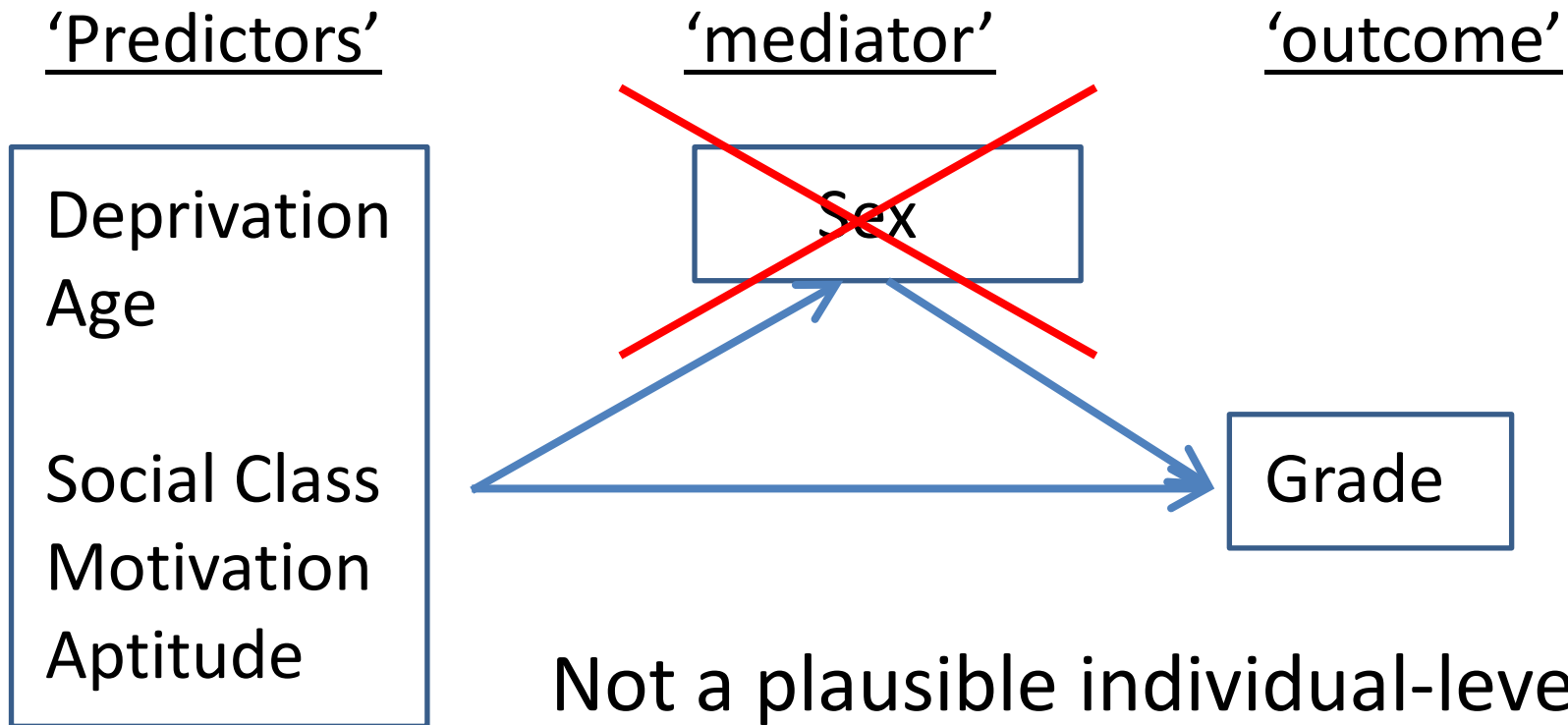


# Path analysis - mediation

- We can interpret the model as assessing what part of the variation in the outcome is 'caused' by the mediator, if the mediator is truly causal.
  - Doesn't prove that this causal interpretation is true or not
- Not an 'actuarial' model – we now care *how* one thing might affect another
- Need to think about plausible causal 'paths'

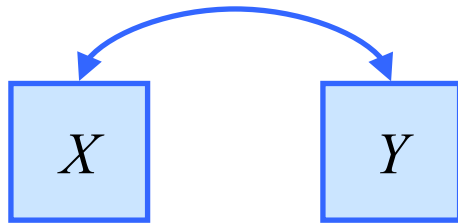


# Path analysis - mediation

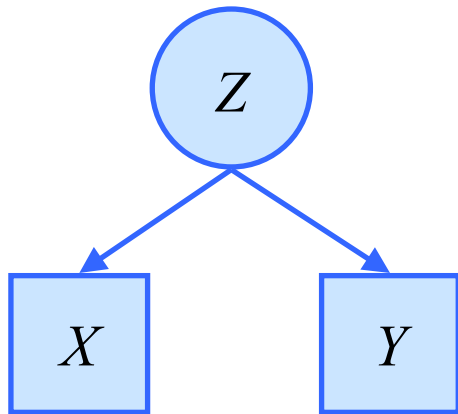


Not a plausible individual-level causal path (could be selection?)

# Factor Analysis

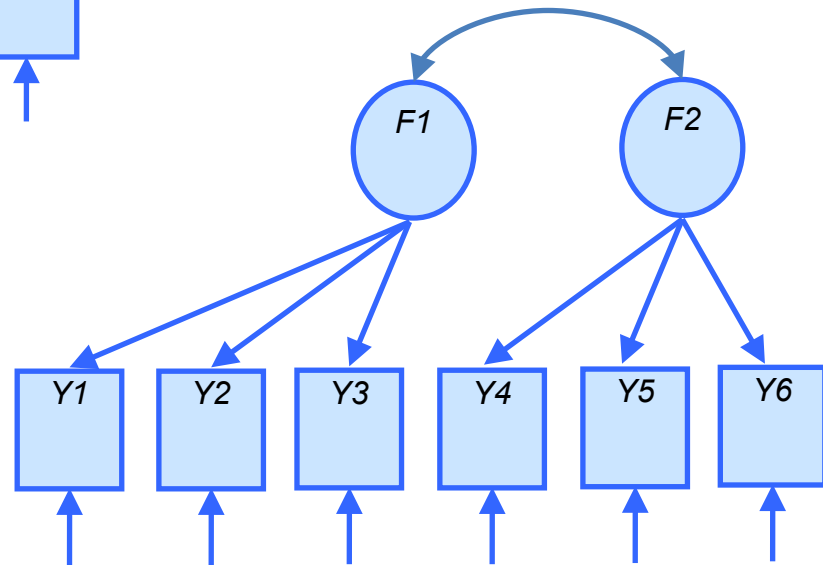
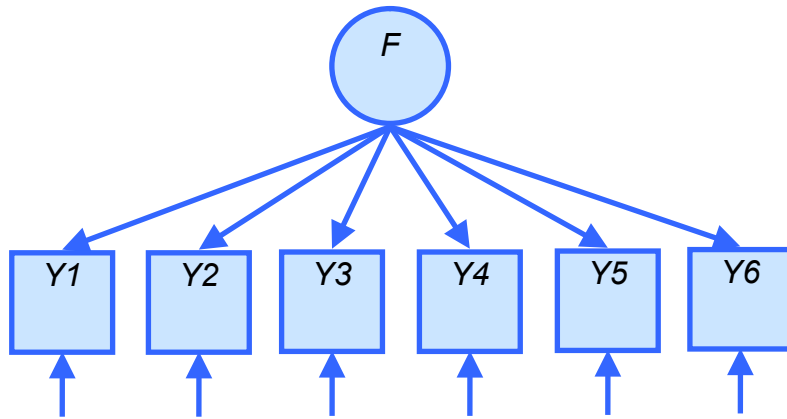


We observe a correlation between variables.



We postulate a hidden, unobserved variable that causes the correlation – a factor.

# Factor Analysis



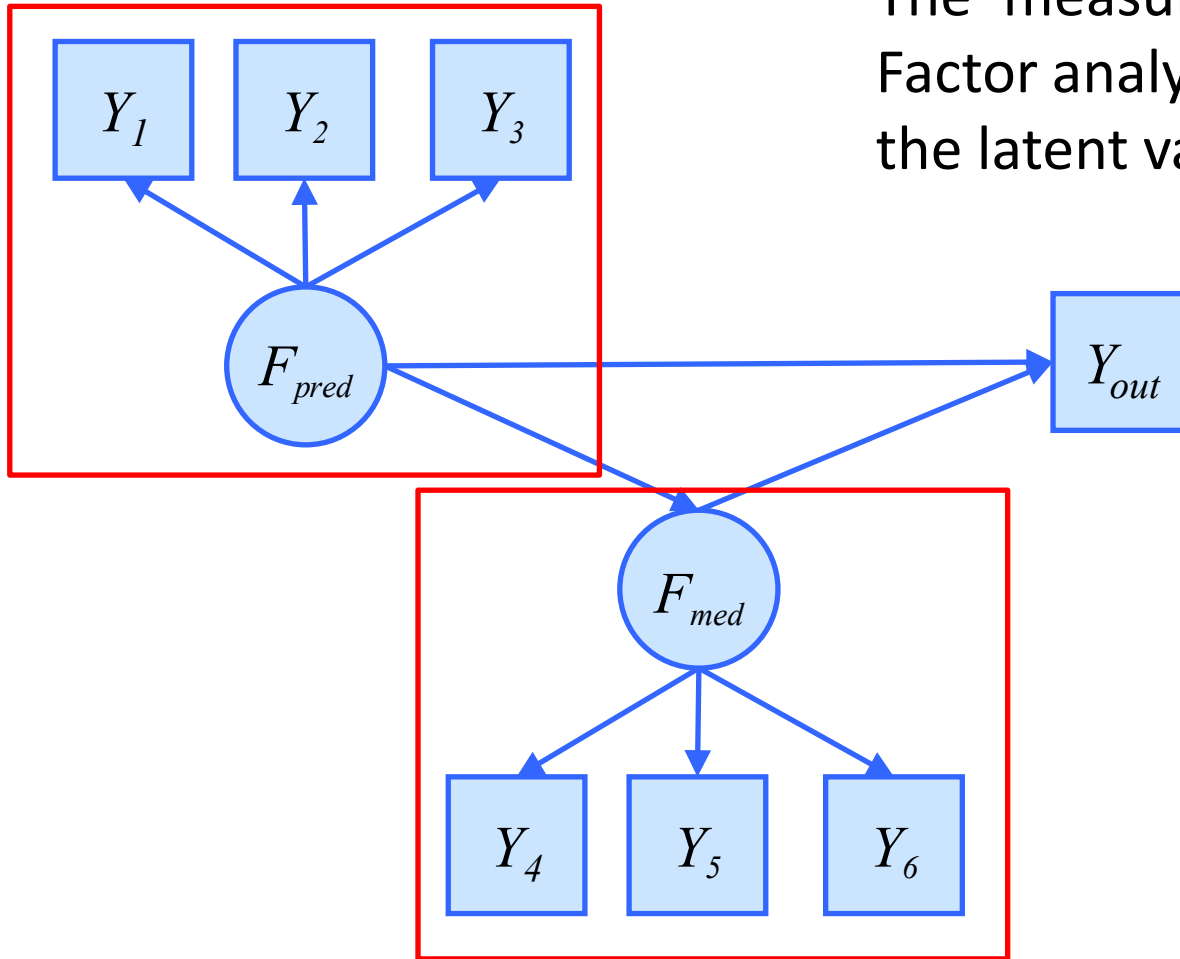
We can address questions  
Such as:  
1 factor or 2 factors?

# SEM

- Put some or all of these together and we've got a full SEM

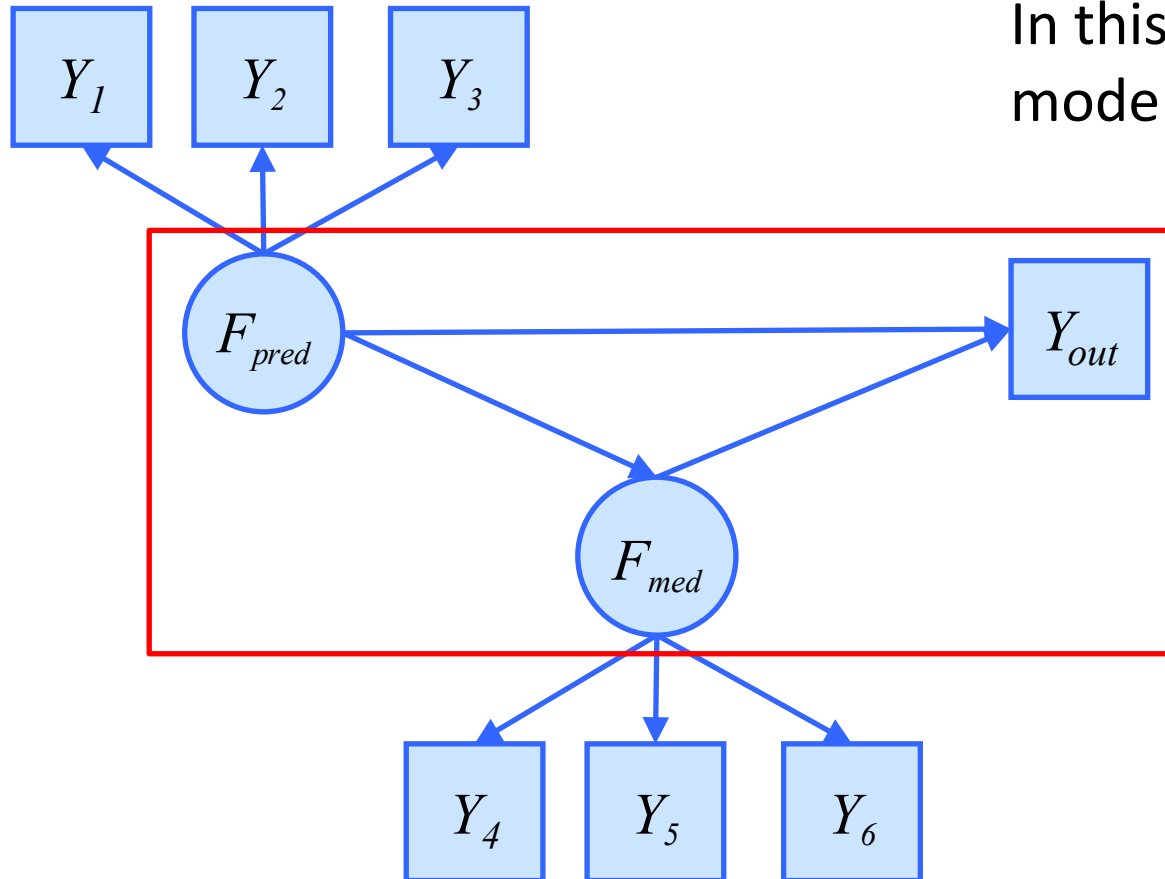
# SEM

The 'measurement' part:  
Factor analyses that define  
the latent variables.



# SEM

The structural part:  
In this case a mediation model.



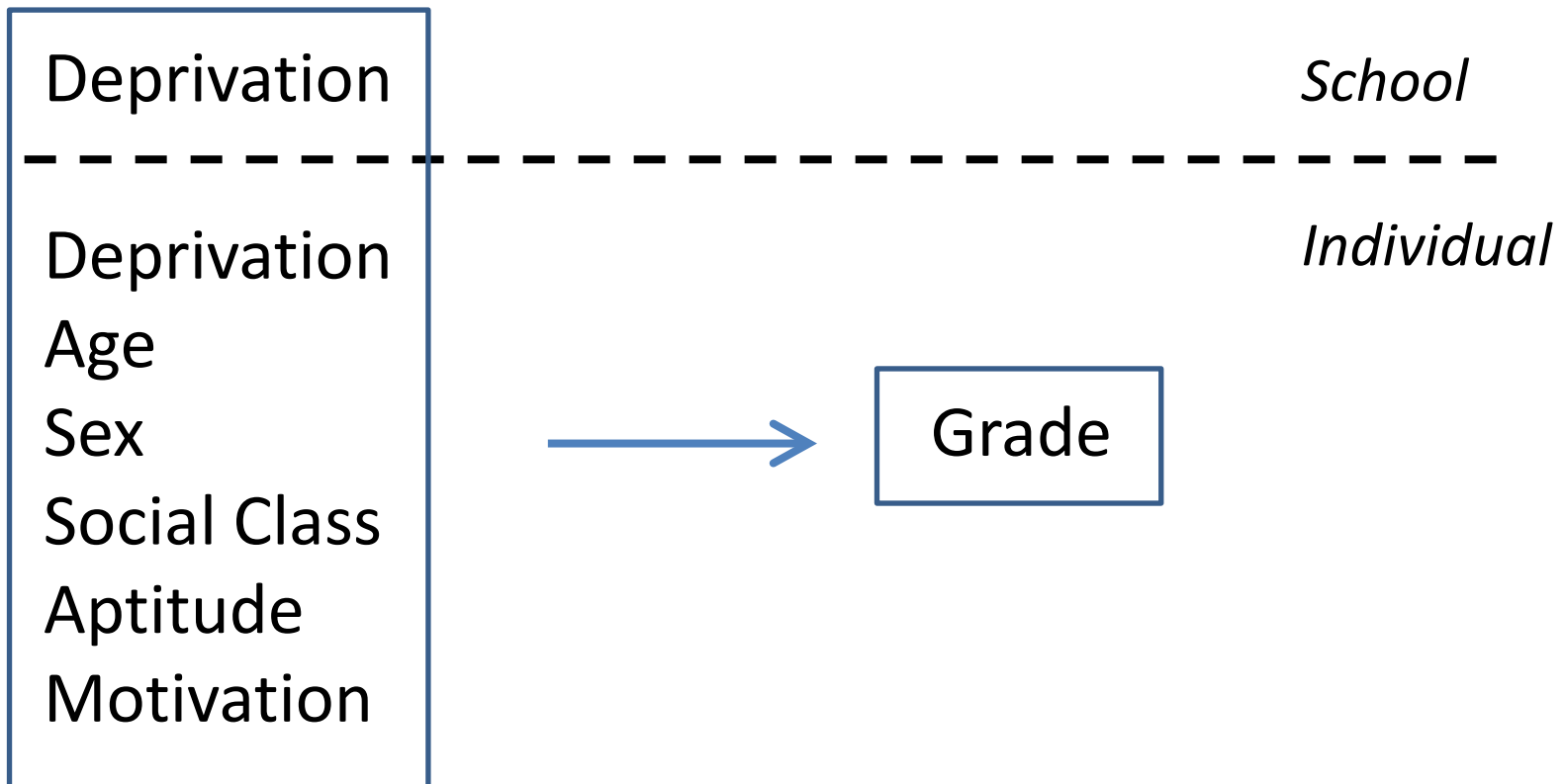
# Multilevel analysis

- Single level analysis assumes residuals are independent
- When data are sampled from a clustered, hierarchical data structure, this assumption can be violated
  - E.g. residuals may all be low/high on average from some clusters compared to others
- Multilevel analysis allows for this correlation among units, using random effects.

# Multilevel regression analysis

'Predictors'

'outcome'

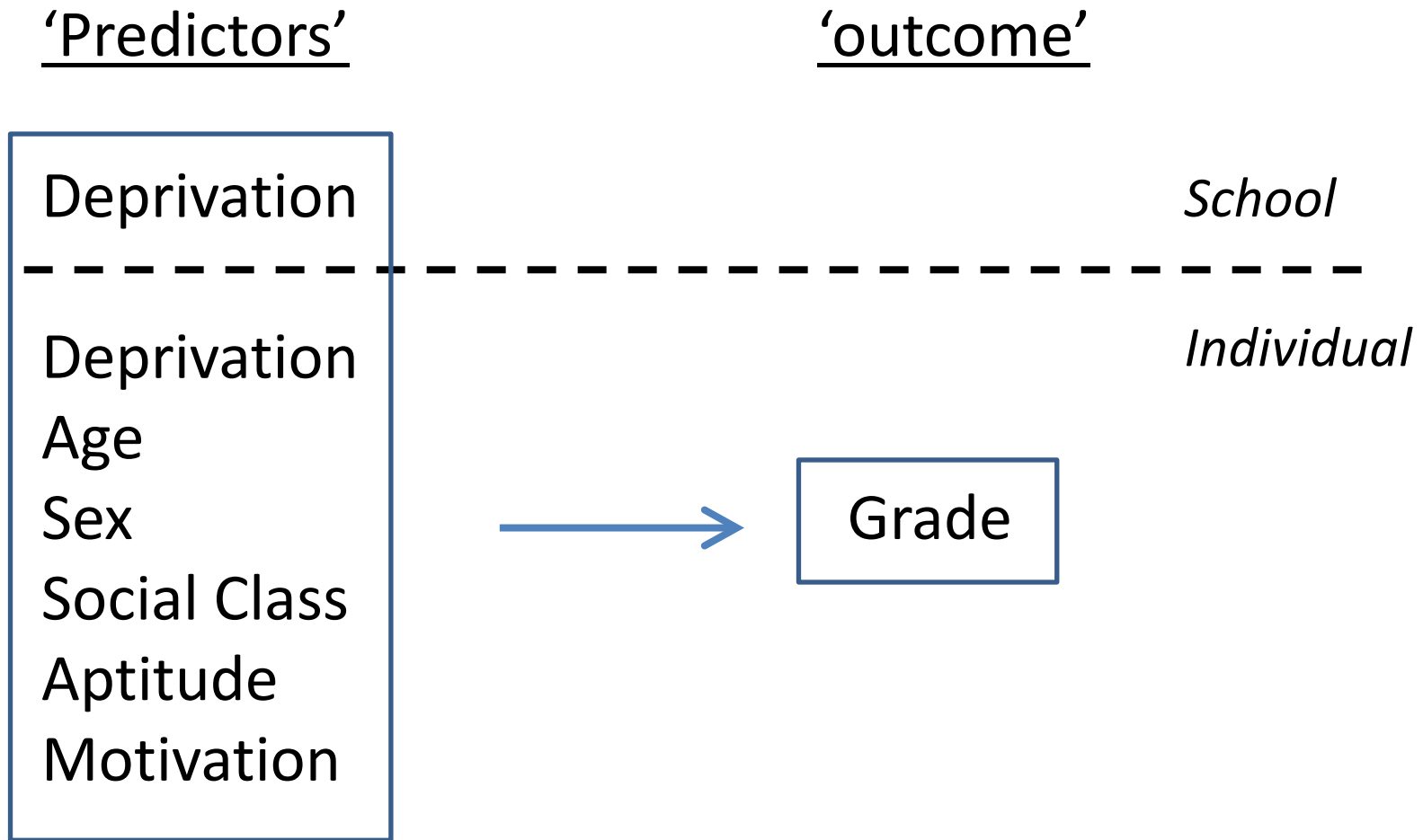




# Multilevel regression analysis

- Here, 2 levels
  - Level-1 (pupils); clustered/nested in
  - Level-2 (schools)
- Random intercepts
  - Each level-2 cluster has its own average level-1 outcome (after accounting for predictors)
- Random slopes
  - Each level-2 cluster has its own  $b$  coefficient for the effect of a predictor on the outcome.

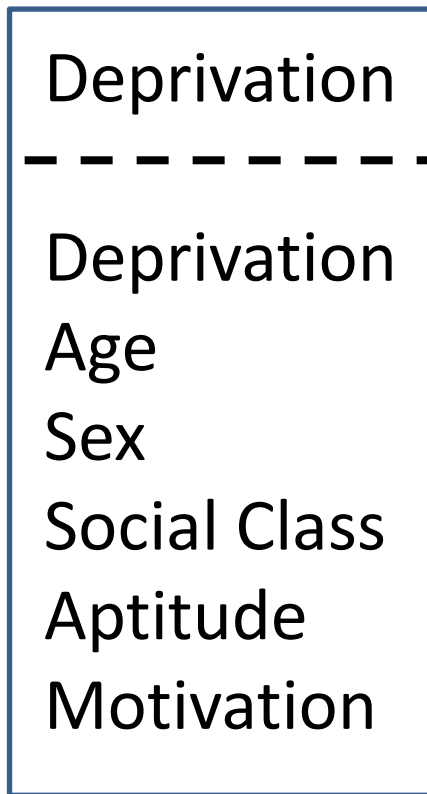
# Multilevel regression analysis



# Multilevel regression analysis

'Predictors'

'outcome'



*School*

*Individual*

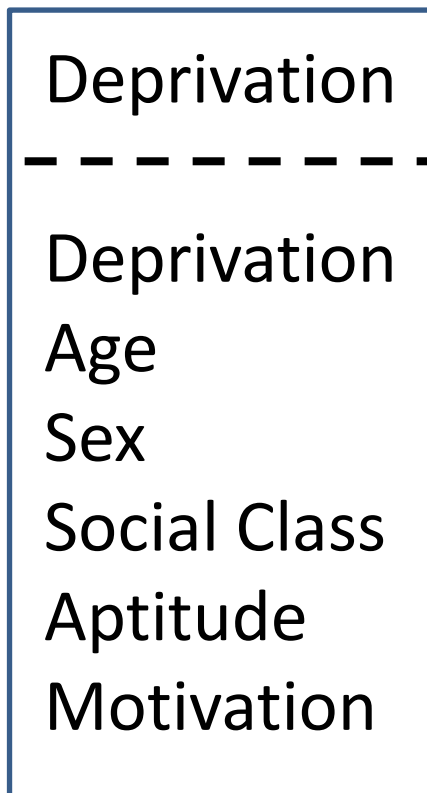


Random intercept – different  
'residual' average Grade across  
schools

# Multilevel regression analysis

'Predictors'

'outcome'



*School*

*Individual*



Random slope – different effects of predictor on Grade across schools

# Multilevel mediation analysis

- What level is the predictor?
- What level is the mediator?
- What level is the outcome?

# Multilevel mediation, 1-1-1

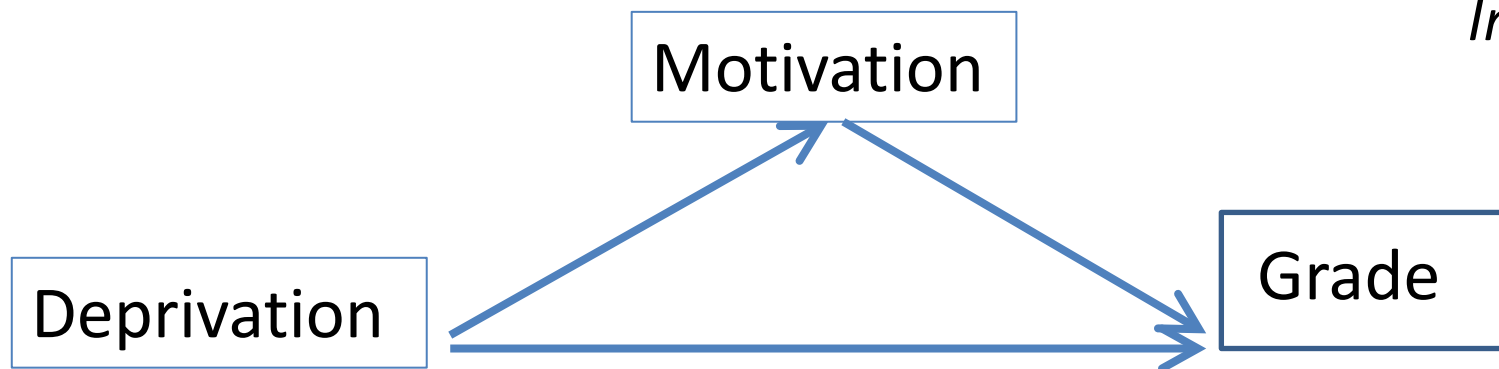
'Predictors'

'mediator'

'outcome'

*School*

*Individual*



# Multilevel mediation, 1-1-1

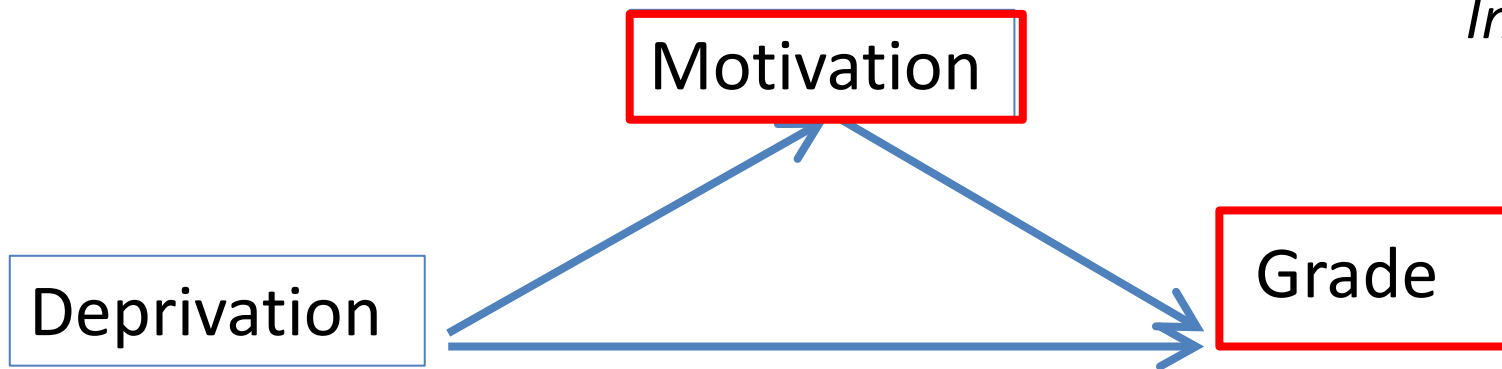
'Predictors'

'mediator'

'outcome'

*School*

*Individual*



Random intercepts

# Multilevel mediation, 1-1-1

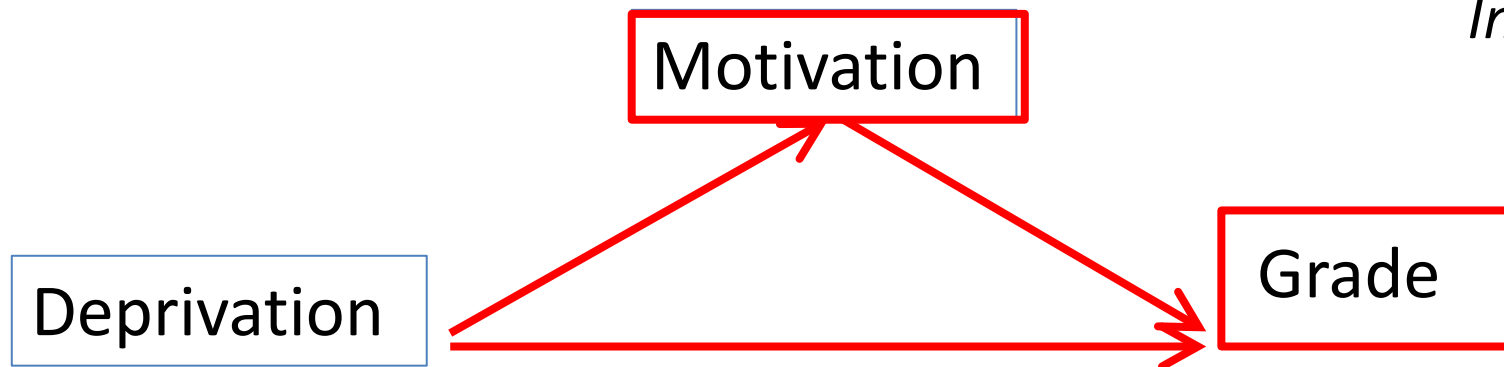
'Predictors'

'mediator'

'outcome'

*School*

*Individual*



...and possibly random slopes too

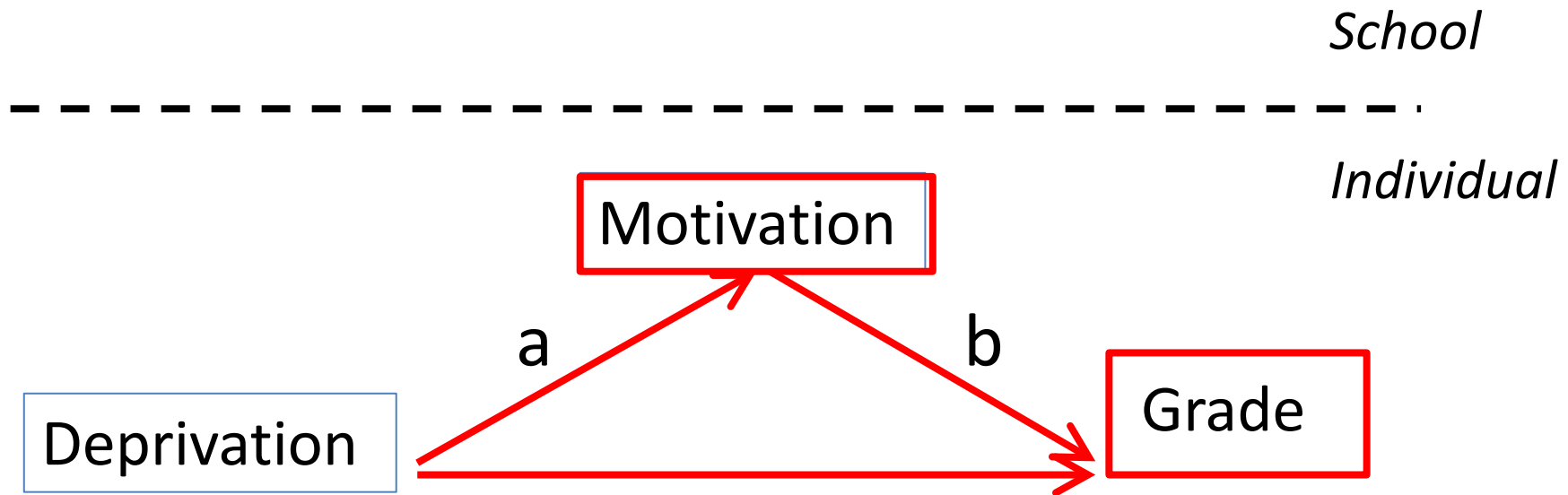


# Multilevel mediation, 1-1-1

'Predictors'

'mediator'

'outcome'



In single level mediation,  $a * b =$  'indirect effect'

# Multilevel mediation, 1-1-1

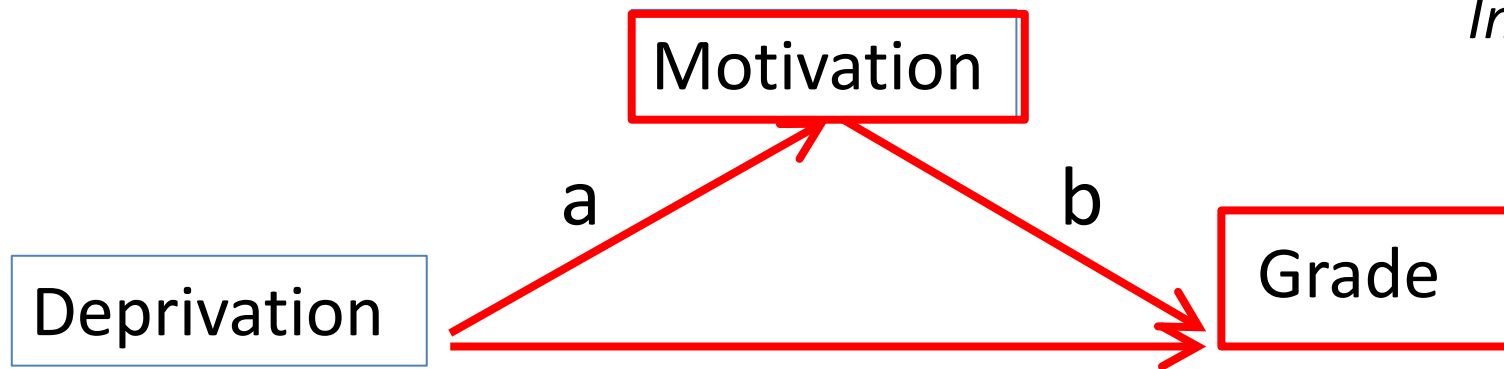
'Predictors'

'mediator'

'outcome'

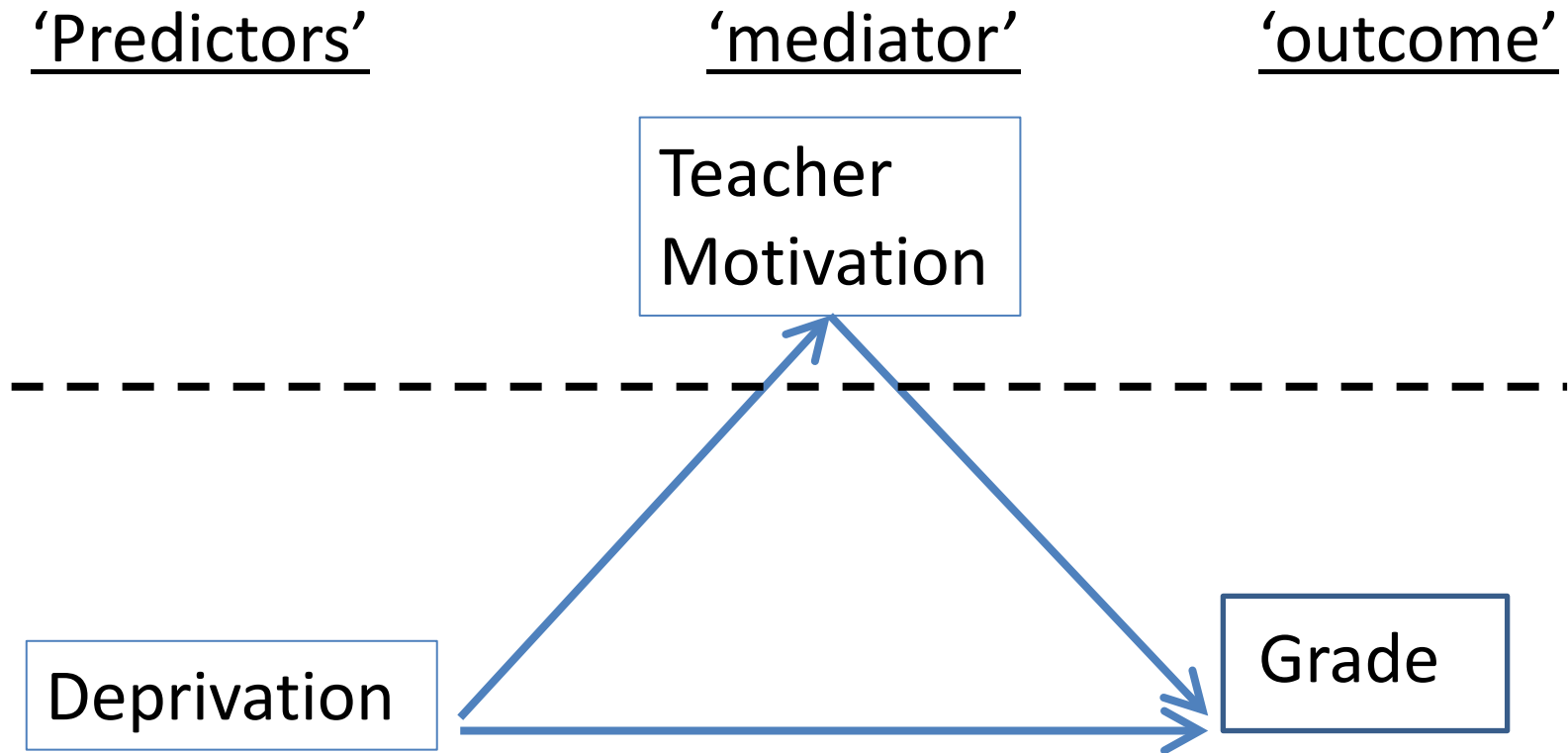
*School*

*Individual*



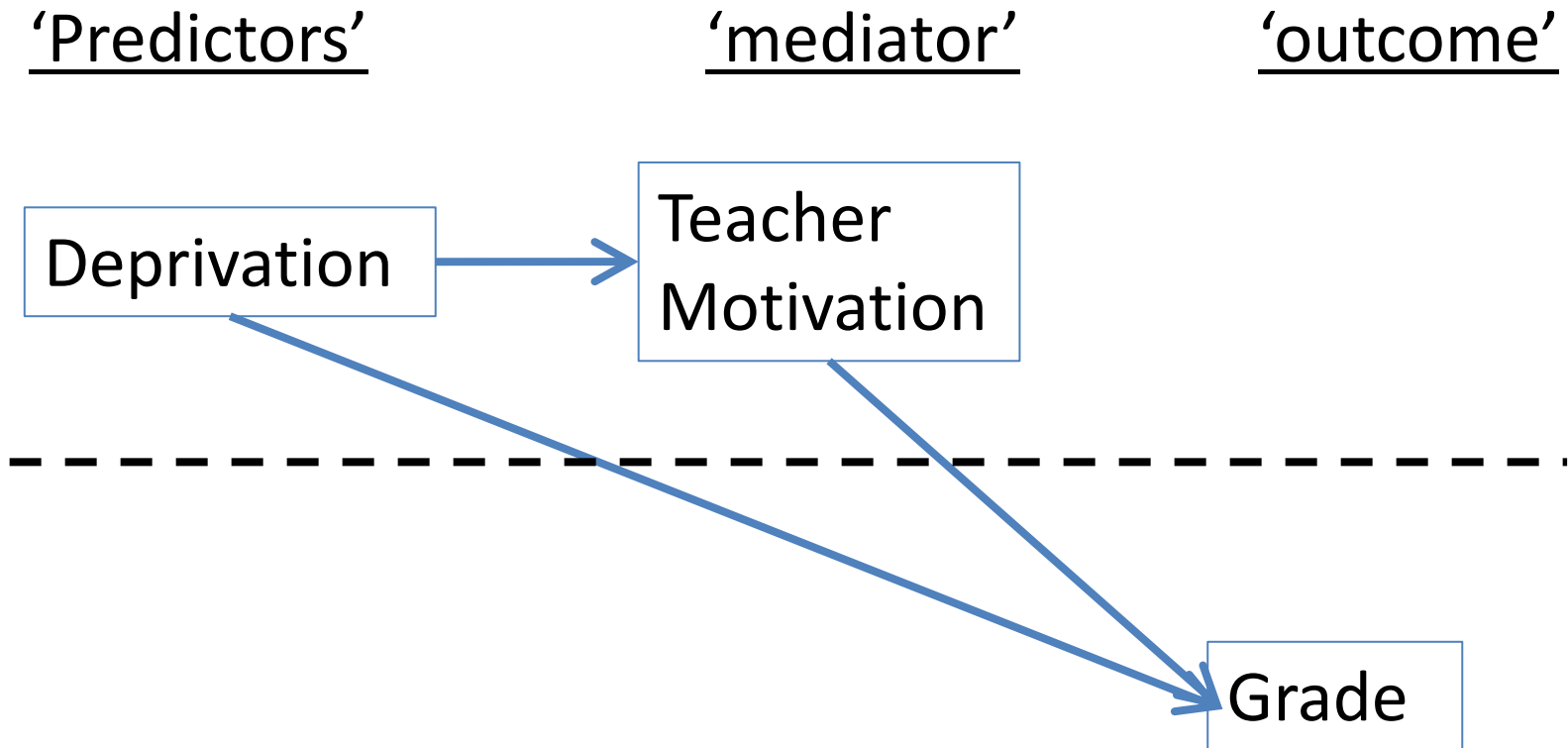
RSs may be correlated, so in ML model,  
indirect effect =  $a * b + \sigma_{ab}$  (cov ab)

# Multilevel mediation, 1-2-1



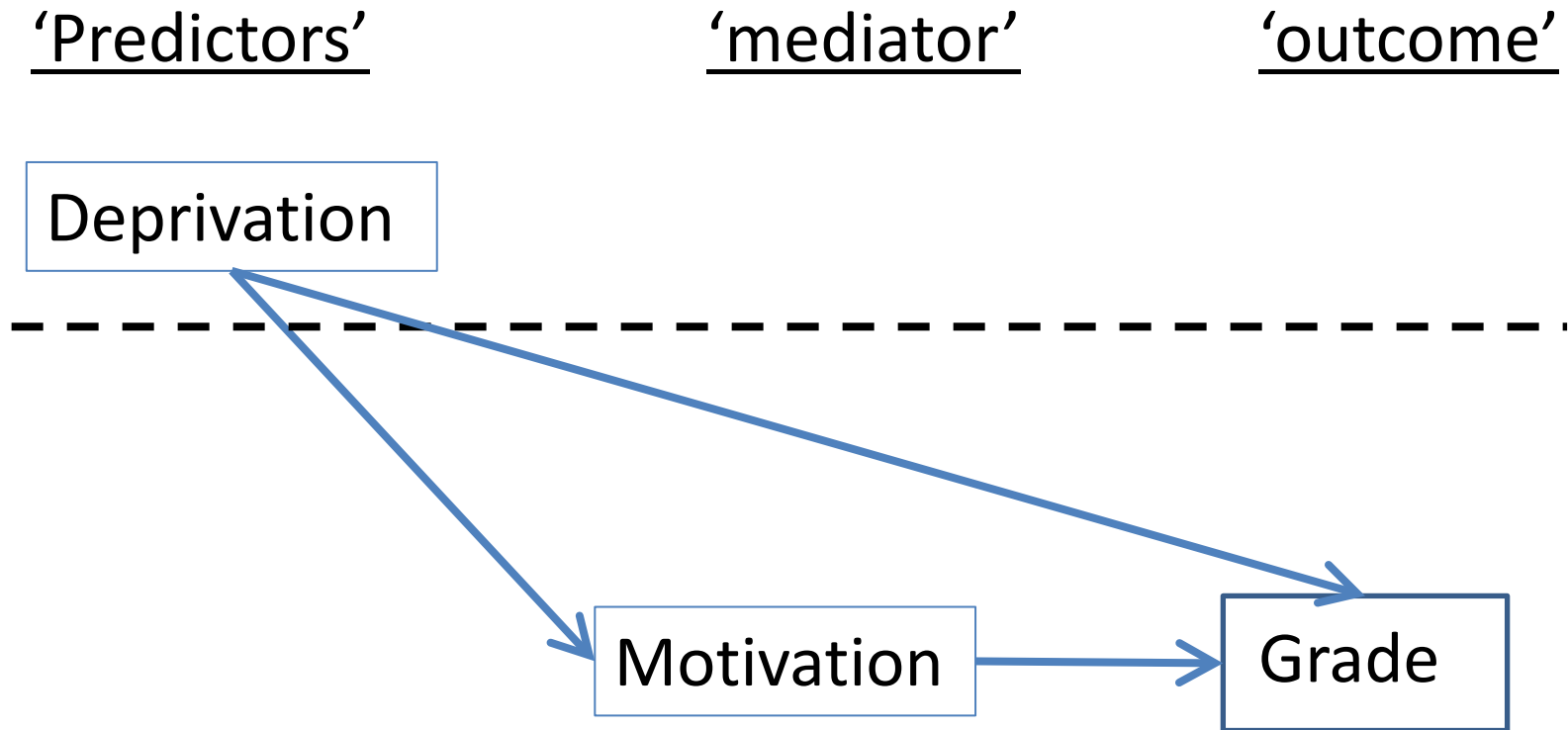
Teachers de/motivated by  
rich/poor kids?

# Multilevel mediation, 2-2-1



Could be a selection effect here?

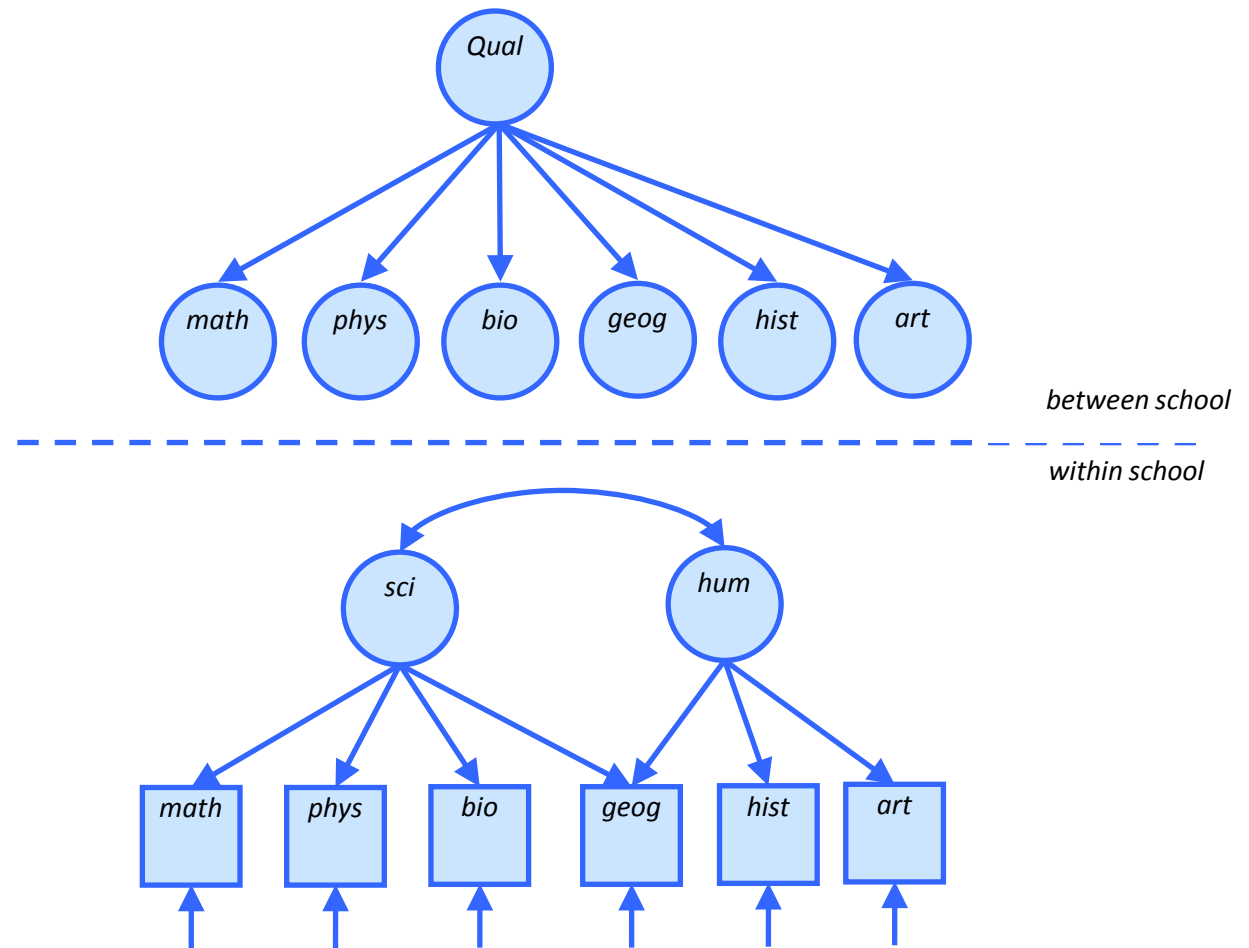
# Multilevel mediation, 2-1-1



And so on.

This demands precision from theory.

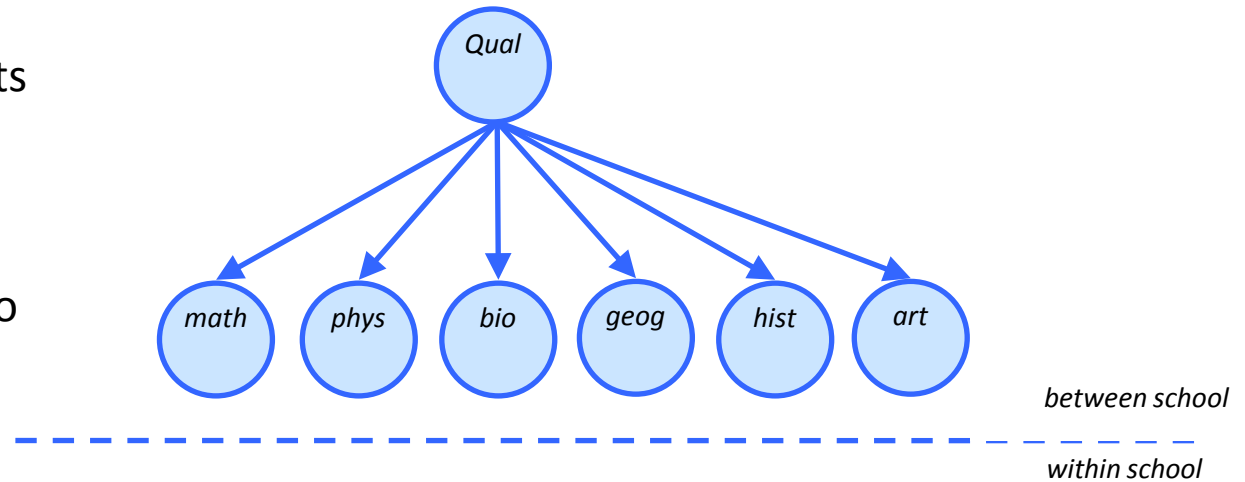
# Multilevel Factor Analysis



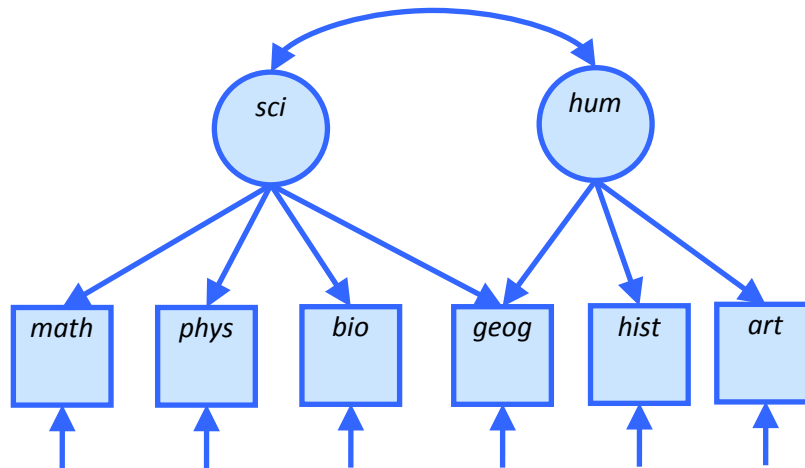
Within school model:  
Standard FA. Pupil grades  
cluster by subject: here,  
science and humanities.

# Multilevel Factor Analysis

Between school model:  
FA of the random intercepts  
for average subject grade  
across schools.  
Between model can have  
different factor structure to  
within model: here,  
'good'/'bad' schools



Within school model:  
Standard FA. Pupil grades  
cluster by subject: here,  
science and humanities.



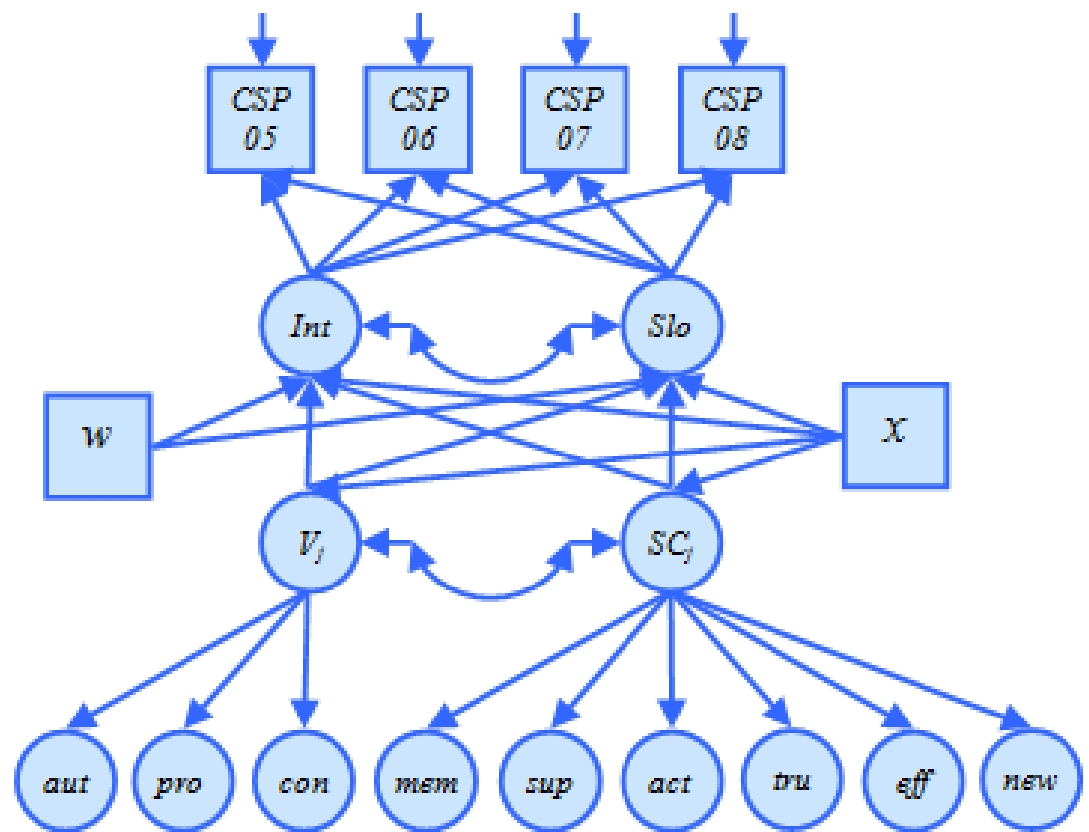
# Multilevel SEM

ESRC project:  
 “Rediscovering the Civic and Achieving Better Outcomes in Public Policy”

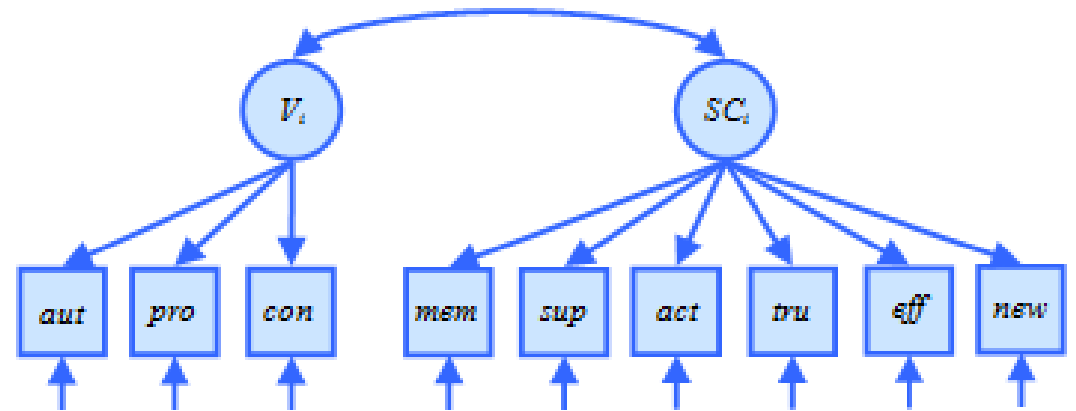
Individuals thought to have differing levels of political “voice” and “civic” behaviours / social capital.

Is this true of areas too?

Voice thought to drive performance of public bodies (i.e. “feedback”).



-----  
 between area  
 within area





# Multilevel SEM

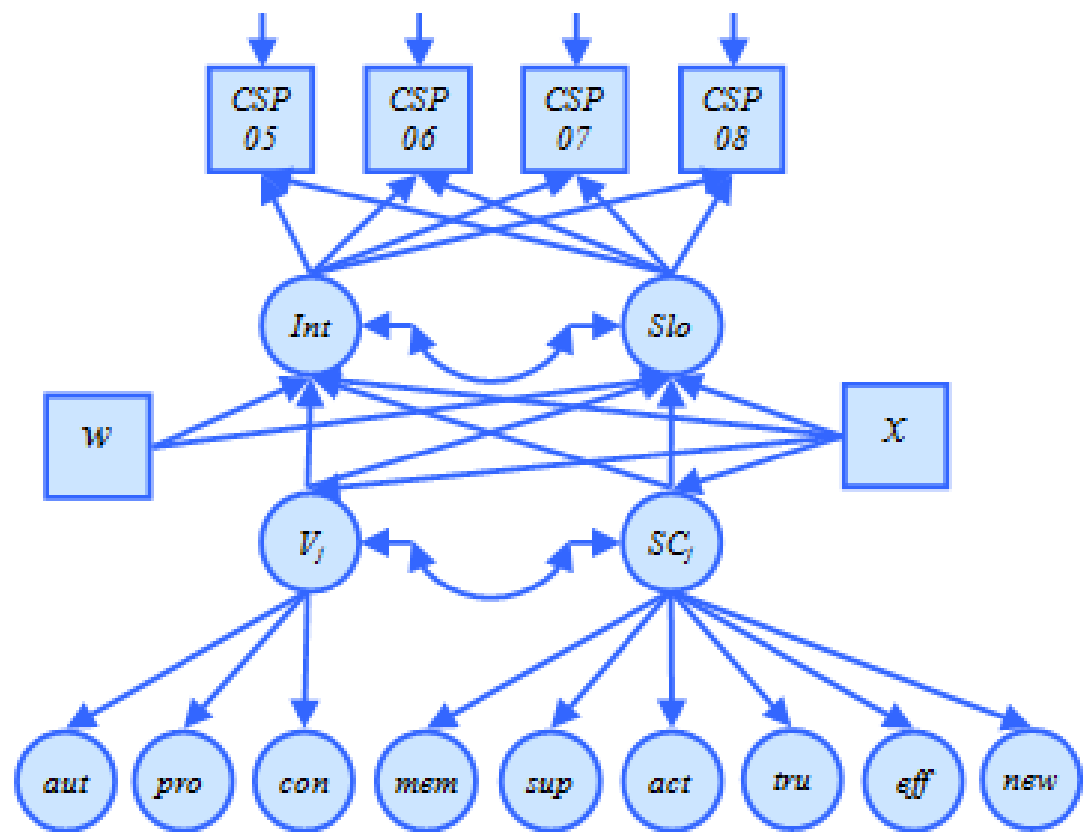
ESRC project:  
 “Rediscovering the Civic and Achieving Better Outcomes in Public Policy”

Individuals thought to have differing levels of political “voice” and “civic” behaviours / social capital.

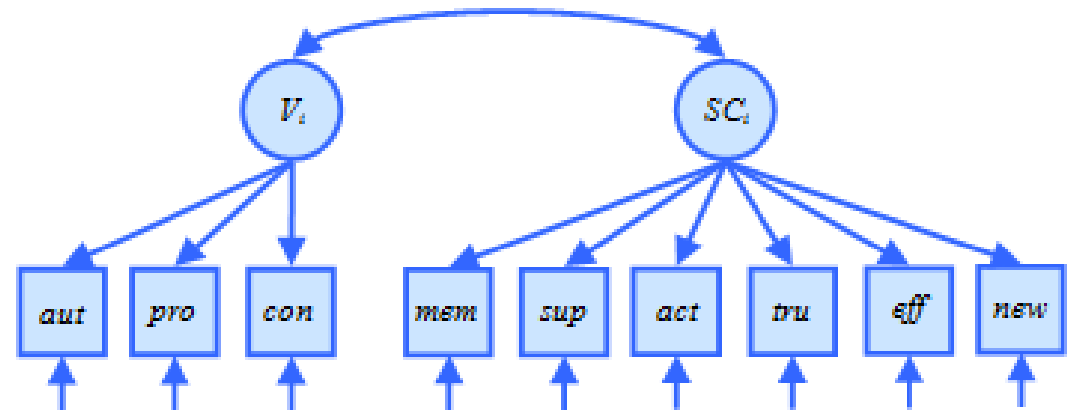
Is this true of areas too?

Voice thought to drive performance of public bodies (i.e. “feedback”).

It didn't. Neither social capital.



-----  
 between area  
 within area



# Approaches to estimation

- ‘Multilevel’ approach (e.g. GLLAMM procedure in Stata)
  - All outcomes stacked into a single response vector.
  - Can have as many levels as required
  - Very computationally intensive
- Multivariate approach (e.g. Mplus, Mx)
  - Multivariate response vector
  - Only two levels (or three, if bottom level is occasion)
    - Correlation matrix decomposed into ‘within’ and ‘between’ levels.
  - Can be very computationally efficient

# Approaches to estimation

- ML and MV approaches are equivalent for certain type of models, e.g. random intercepts and ML-CFA:
  - ML = “person  $i$  in cluster  $j$ ”, MV = “variable  $i$  in person  $j$ ”
    - Mehta & Neale (2005)

# Guides and tutorials

- Stata
  - The **gllamm** command can fit Multilevel SEMs.
  - Download the manual and lots of worked examples from
    - [www.gllamm.org](http://www.gllamm.org)
- Mplus
  - Can fit 2(3 if longitudinal)-level Multilevel Structural Equation Models, both confirmatory and exploratory.
  - Download free demo version of Mplus from:
    - [www.statmodel.com](http://www.statmodel.com)

# References

**A talk** that gives a more technical introduction to MLSEM:

- [http://www.ccsr.ac.uk/qmss/summer/Leuven11/documents/Lecture4\\_Multilevel.pdf](http://www.ccsr.ac.uk/qmss/summer/Leuven11/documents/Lecture4_Multilevel.pdf)

## Papers

Kenny, D. A., Korchmaros, J. D. and Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8(2): 115-128.

Mehta, P. and Neale, M. (2005). People Are Variables Too: Multilevel Structural Equations Modeling. *Psychological Methods*, 10(3): 259–284.

Marsh, H. W., Ludtke, O., Robitzsch, A., Trautwein, U, Asparouhov, T., Muthen, B and Nagengast, B. (2009). Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research*, 44(6): 764-802