EARLI 2005
JuRe preconference

Analysis of hierarchical data

Johannes Hartig

Deutsches Institut für Internationale Pädagogische Forschung
German Institute for International Educational Research
Mitglied der Leibniz-Gemeinschaft

Contents of this workshops

- Hierarchical data structures: definition and examples
- Multilevel linear regression:
  - Purpose and basic concepts
  - Regression equations on different levels
  - Multilevel regression coefficients and their meaning
  - An example using HLM 6.0
  - Decomposing effects of a lower level predictor
- Multilevel structural equation modeling
  - Basic idea of structural equation modeling
  - Decomposition of correlations on different levels
  - Separate models for each level

Contents of this workshops

- What is hierarchical / multilevel data?
- Why should I bother using special methods to analyze multilevel data?

- What is multilevel linear regression?
- What effects can be tested in multilevel linear regression models?

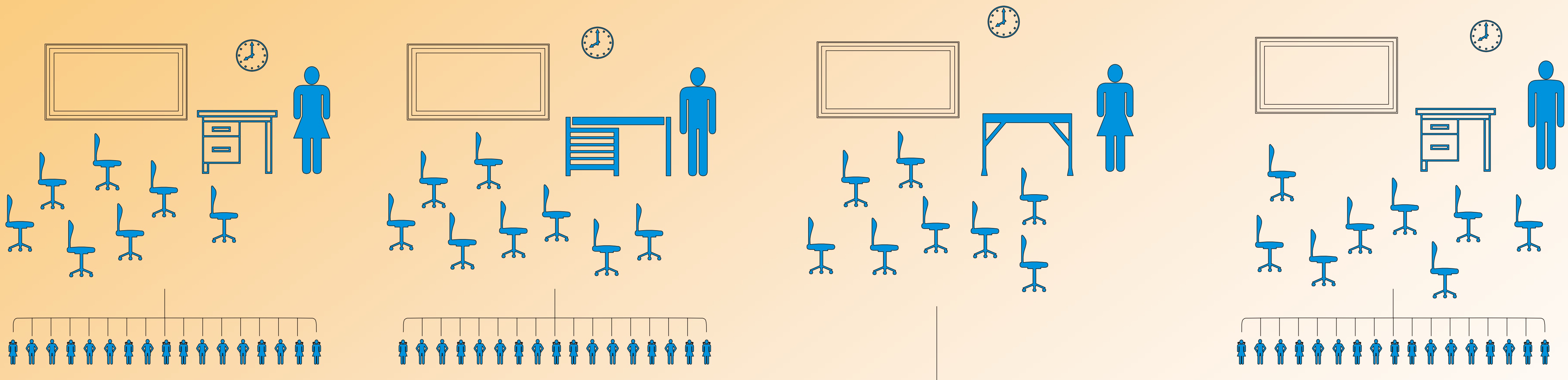
- What is the basic idea of multilevel structural equation modeling?

Definition of hierarchical data structures

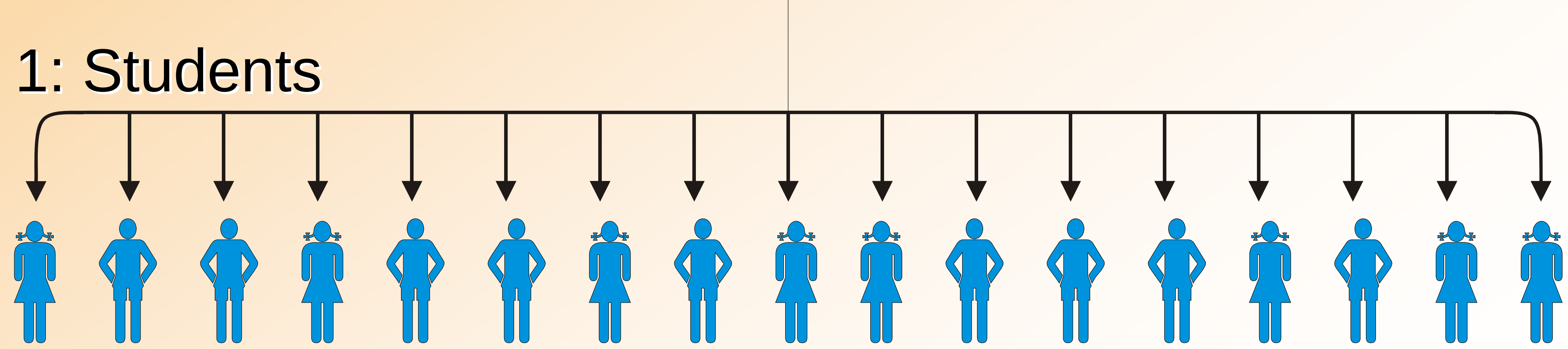
- Synonym: Multilevel structures
- Structures with several hierarchically ordered levels
- Observable units can be defined within each level (e.g. students on a lower, classrooms on a higher hierarchical level)
- Each unit on a lower level can unambiguously be assigned to one and only one unit on the higher level.

Examples of hierarchical data structures

Level 2: Classes



Level 1: Students

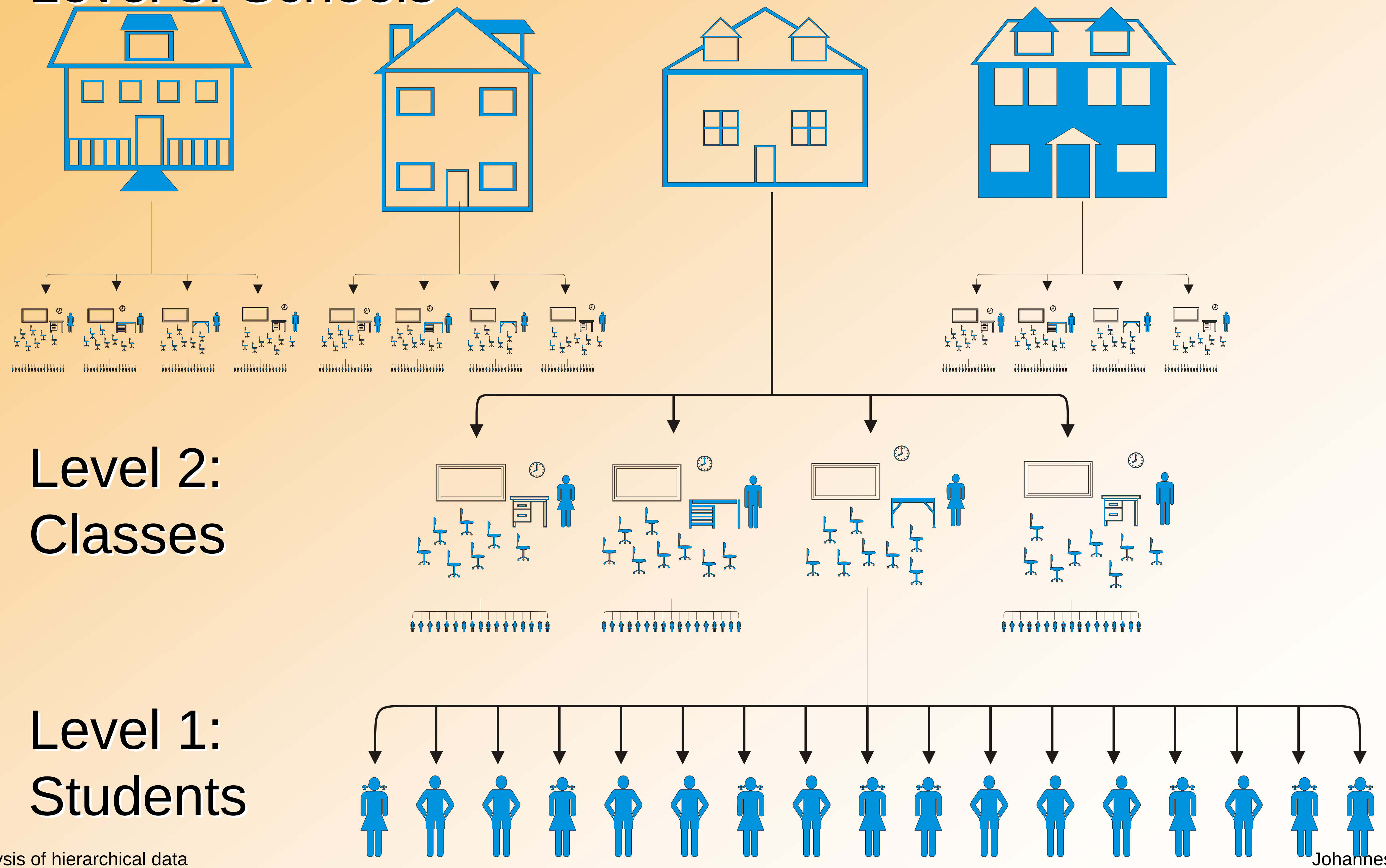


Examples of hierarchical data structures

Level 3: Schools

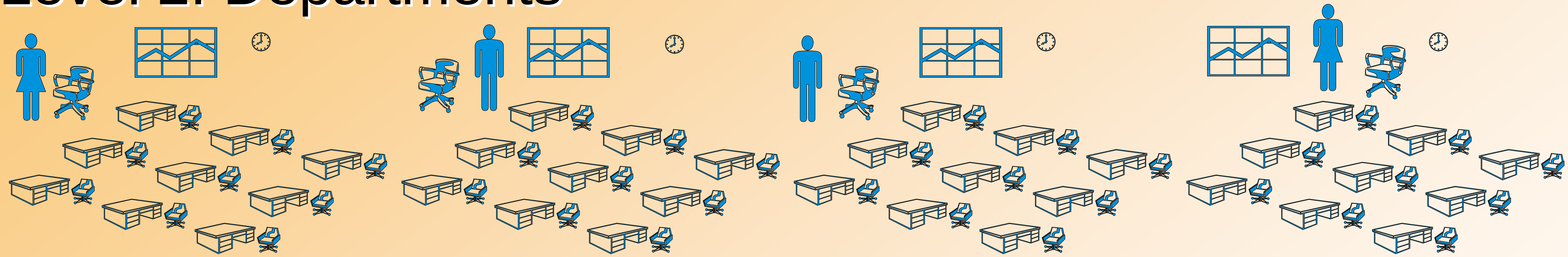
Level 2: Classes

Level 1: Students

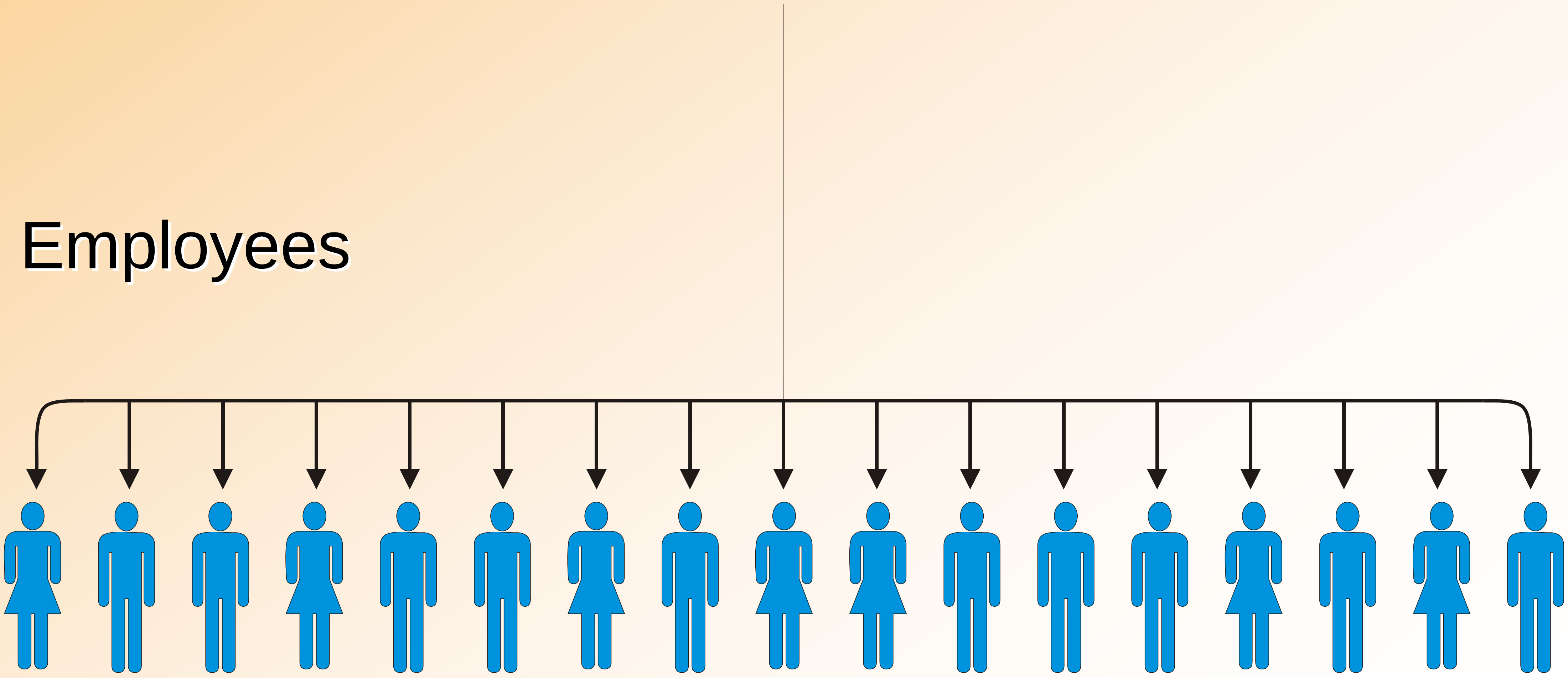


Examples of hierarchical data structures

Level 2: Departments



Level 1: Employees

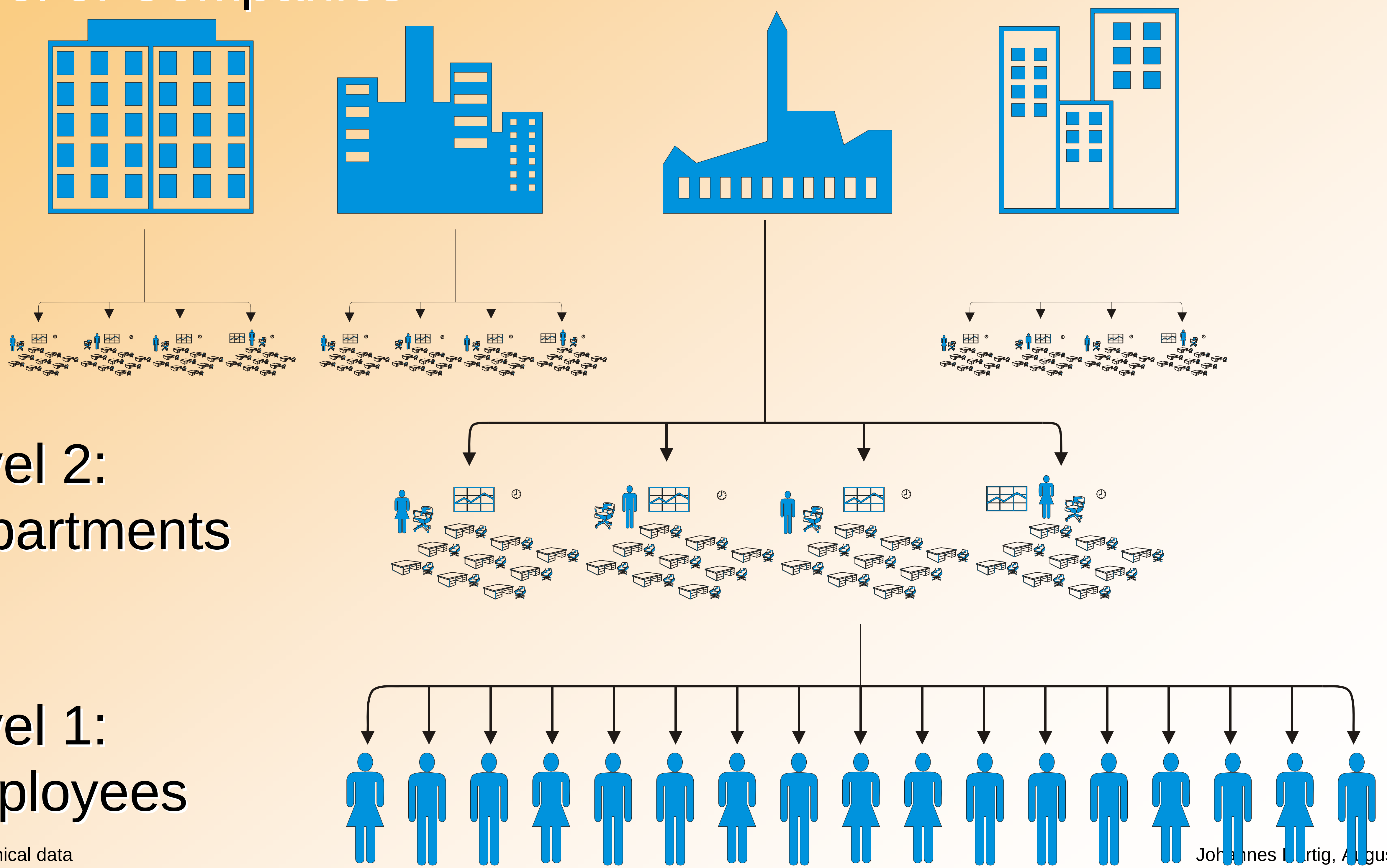


Examples of hierarchical data structures

Level 3: Companies

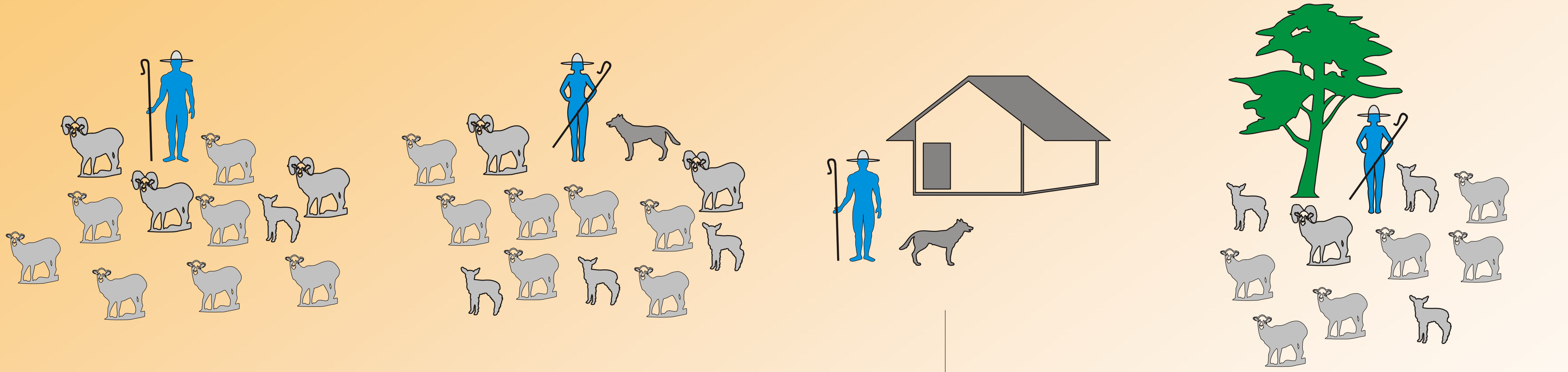
Level 2: Departments

Level 1: Employees

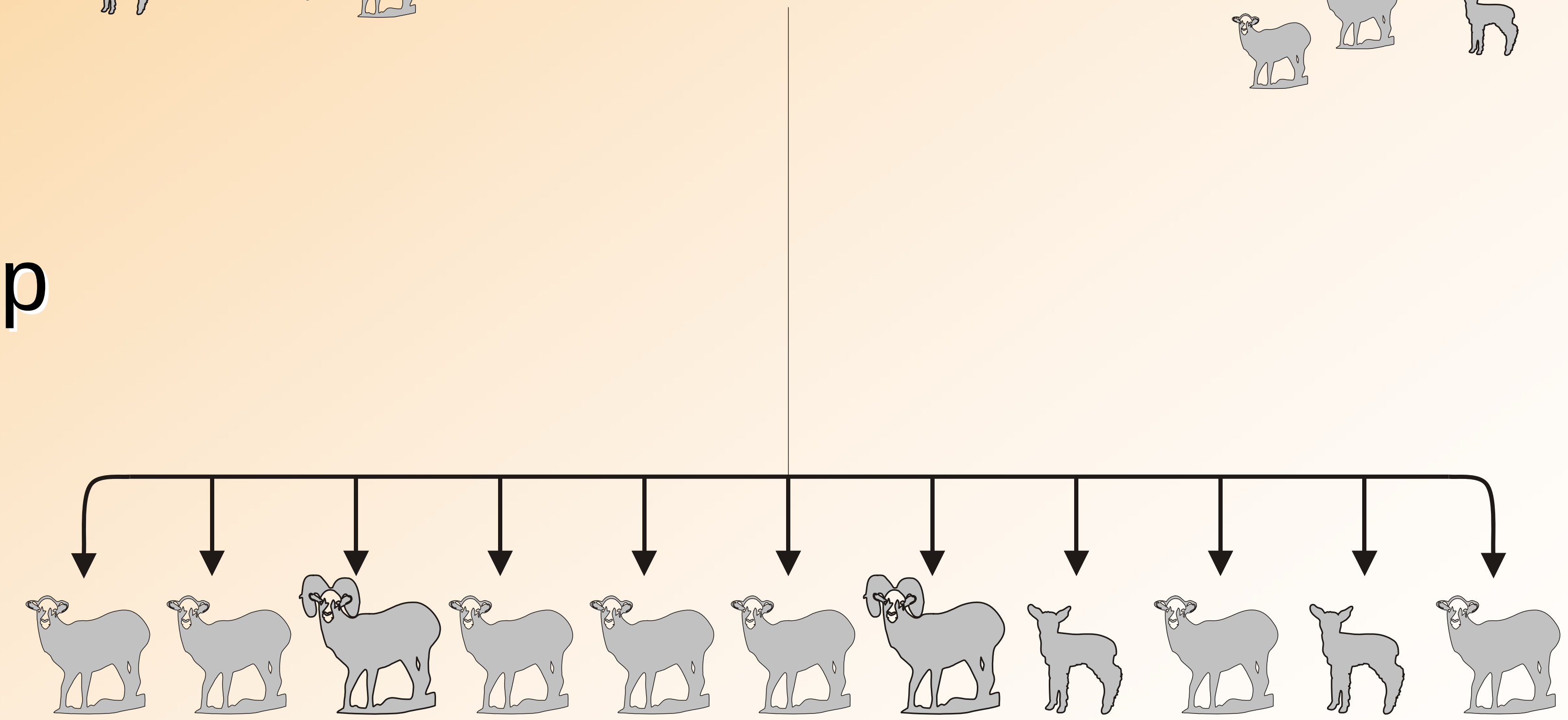


Examples of hierarchical data structures

Level 2: Flocks

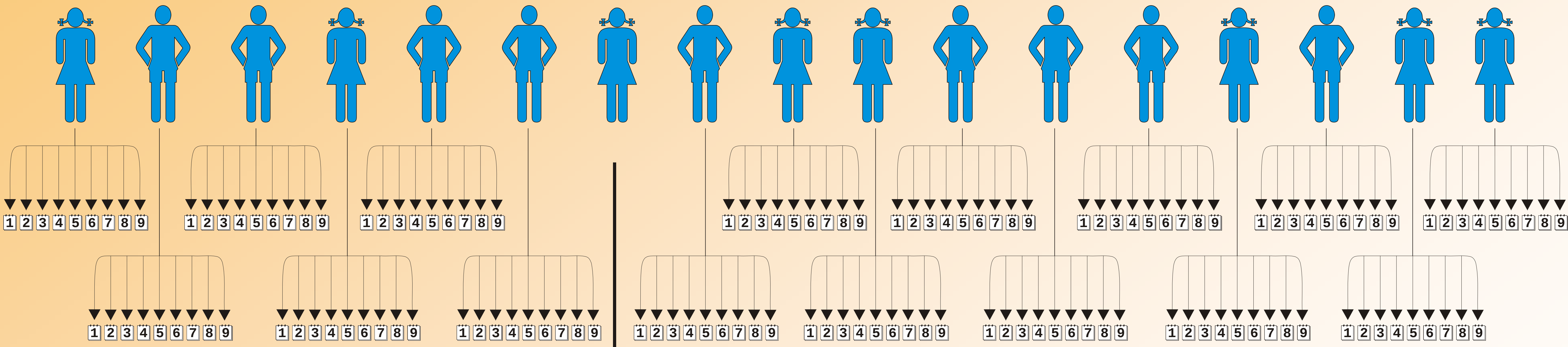


Level 1: Sheep

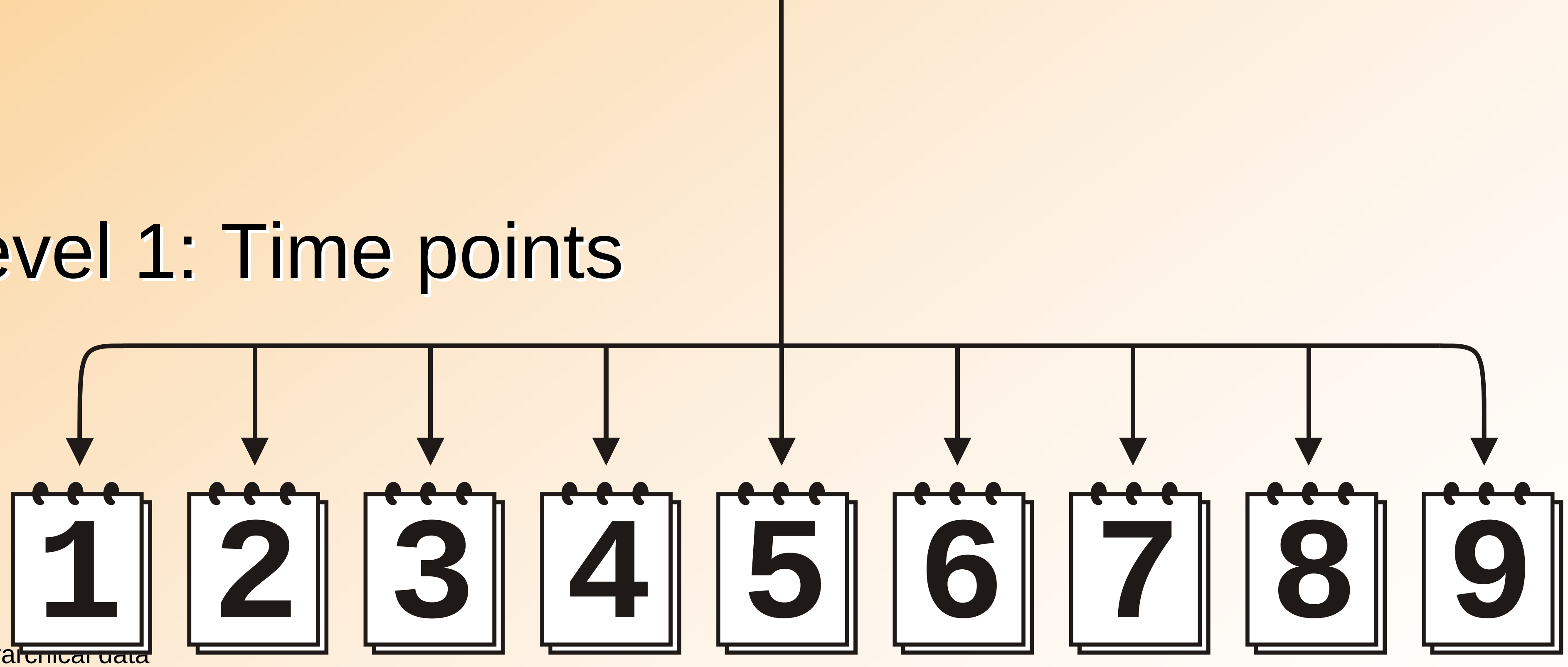


Examples of hierarchical data structures

Level 2: Persons



Level 1: Time points



Statistical problems when analyzing hierarchical data

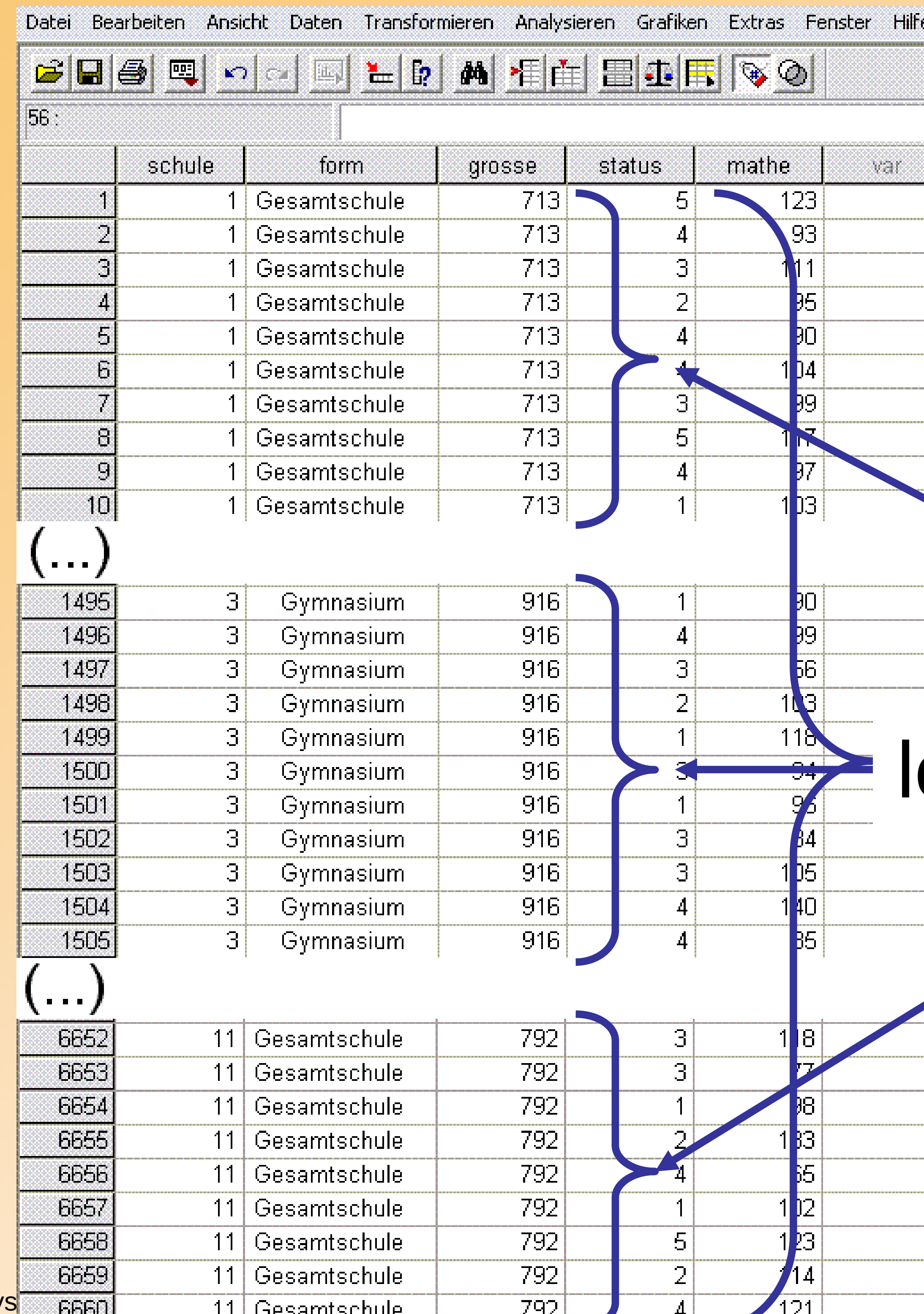
- Data of level 1 units within the same level 2 units are not independent: e.g. students within the same class are more similar among each other than to students from different classes.
- The similarity between level 1 units within the same level 2 units is expressed by the intra-class-correlation; it is a measure for the *proportion of variance between level 2 units*.
- Standard statistical analysis techniques like linear regression or analysis of variance do not take into account these dependencies, and results obtained by these methods are biased.

Dealing with hierarchical data structures: aggregation and disaggregation

Disaggregation

- level 2 data is "multiplied" by assigning each level 1 unit the properties of its level 2 unit that were measured at a higher level.
- E.g. each student is assigned classroom variables such as students per classroom, and all students in a given class have the same value on these variable.

Disaggregation of level 2 data



Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

56 :

	schule	form	grosse	status	mathe	var
1	1	Gesamtschule	713	5	123	
2	1	Gesamtschule	713	4	93	
3	1	Gesamtschule	713	3	111	
4	1	Gesamtschule	713	2	95	
5	1	Gesamtschule	713	4	90	
6	1	Gesamtschule	713	4	104	
7	1	Gesamtschule	713	3	99	
8	1	Gesamtschule	713	5	117	
9	1	Gesamtschule	713	4	97	
10	1	Gesamtschule	713	1	103	
(…)						
1495	3	Gymnasium	916	1	90	
1496	3	Gymnasium	916	4	99	
1497	3	Gymnasium	916	3	56	
1498	3	Gymnasium	916	2	103	
1499	3	Gymnasium	916	1	118	
1500	3	Gymnasium	916	3	94	
1501	3	Gymnasium	916	1	95	
1502	3	Gymnasium	916	3	84	
1503	3	Gymnasium	916	3	105	
1504	3	Gymnasium	916	4	140	
1505	3	Gymnasium	916	4	85	
(…)						
6652	11	Gesamtschule	792	3	118	
6653	11	Gesamtschule	792	3	77	
6654	11	Gesamtschule	792	1	98	
6655	11	Gesamtschule	792	2	133	
6656	11	Gesamtschule	792	4	65	
6657	11	Gesamtschule	792	1	102	
6658	11	Gesamtschule	792	5	123	
6659	11	Gesamtschule	792	2	114	
6660	11	Gesamtschule	792	4	121	

Values of level-2-variables are
level 1 variable el 1 units within
each level 2 unit

Disaggregation of level 2 data

Disaggregation

- Standard statistical methods like linear regression assume that all data is randomly drawn from one homogeneous population.
- In hierarchical data structures this is not the case. For example, schools may be sampled from the population of schools, and then students are sampled from the selected schools.
- If level 2 units are heterogeneous with respect to the dependent variable, standard statistical analysis with disaggregated data will yield wrong standard errors – giving significant results were they shouldn't.

Dealing with hierarchical data structures: aggregation and disaggregation

Aggregation

- Level 1 data is aggregated on level 2, and level 2 units are used as units of analysis.
- E.g. student performance scores are averaged to the class level and classes are used at the unit of analysis.
- The sample size is reduced to the number of level 2 units.

Aggregation of level 1 data

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster

12 :

	schule	form	groesse	am_stat	am_mathe
1	1	Gesamtschule	713	1,70	82,87
2	2	Gesamtschule	686	4,50	72,16
3	3	Gymnasium	916	1,71	96,31
4	4	Gesamtschule	677	5,64	95,25
5	5	Gesamtschule	601	1,50	95,36
6	6	Gymnasium	402	5,77	93,29
7	7	Gesamtschule	682	1,60	92,97
8	8	Gymnasiur	[illegible]	[illegible]	[illegible]
9	9	Gymnasiur	[illegible]	[illegible]	[illegible]
10	10	Gesamtscl	[illegible]	[illegible]	[illegible]
11	11	Gesamtscl	[illegible]	[illegible]	[illegible]
12	12	Gesamtscl	[illegible]	[illegible]	[illegible]
13	13	Gymnasiur	[illegible]	[illegible]	[illegible]
14	14	Gesamtsch	[illegible]	[illegible]	[illegible]
15	15	Gesamtsch	[illegible]	[illegible]	[illegible]
16	16	Gymnasium	[illegible]	[illegible]	[illegible]
17	17	Gesamtsch	[illegible]	[illegible]	[illegible]
18	18	Gymnasium	[illegible]	[illegible]	[illegible]
19	19	Gymnasium	1010	4,22	82,73
20	20	Gesamtschule	988	3,83	82,11

The sample size is reduced to the number of level 2 units

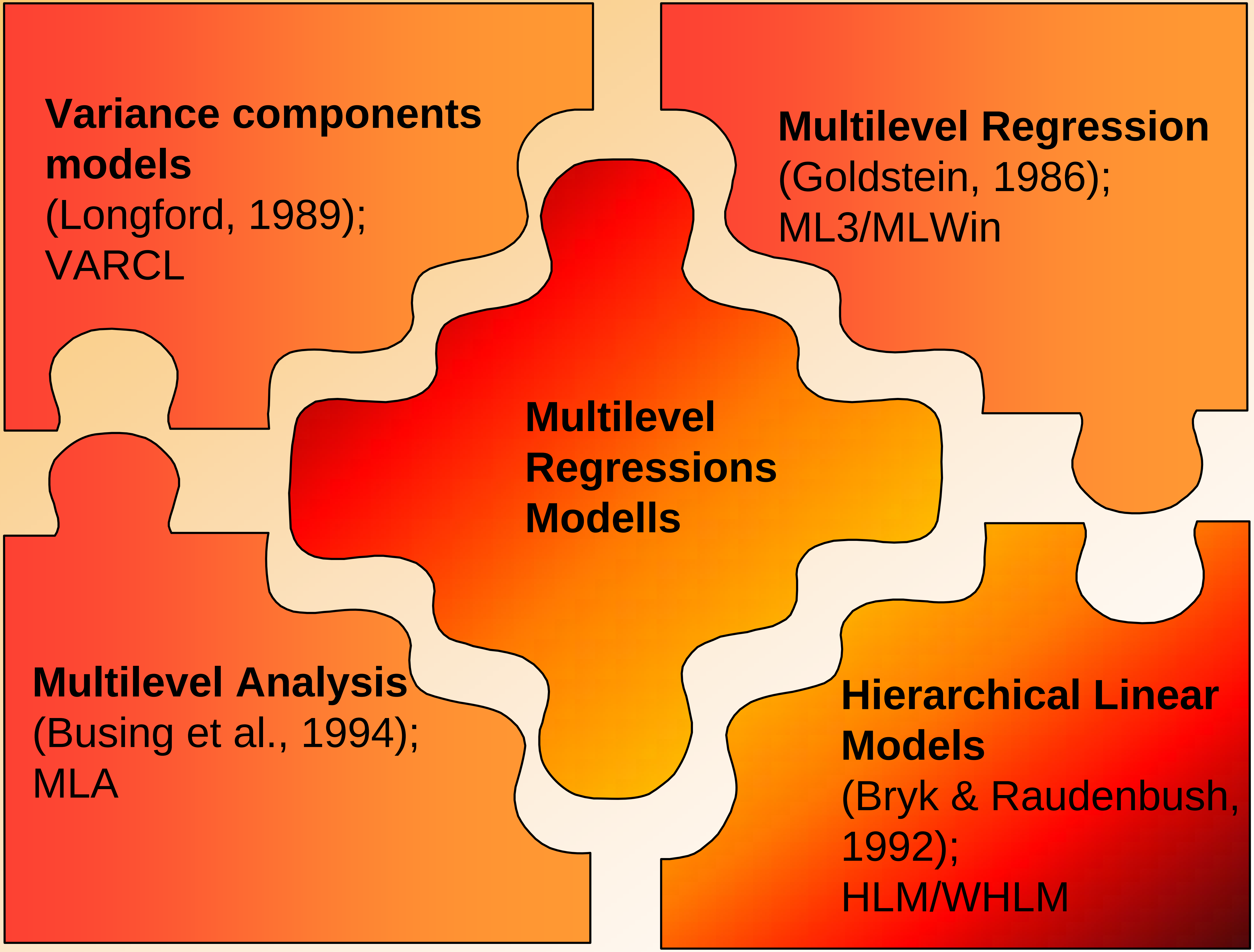
level 1 variables are aggregated (e.g. averaged) within level 2 units

Aggregation of multilevel data: “Ecological fallacy” or “Robinson-Effect”

- Results from aggregated data cannot be interpreted in terms of relations on an individual level.
- Robinson (1950) examined the relation between percentage of blacks and the level of illiteracy in different US regions in 1930.
- At an aggregated level, this correlation is .95 – at individual level, it’s just .20! (cf. Hox, 2002).
- Relations on an aggregated level (“ecological correlations”) are of little use (or even misleading) if one is interested in relations on an individual level.

Analyzing multilevel data

- Within the last decades, statistical methods to analyze relations between variables in hierarchical data structures have been developed.


Variance components models
(Longford, 1989);
VARCL
Multilevel Regression
(Goldstein, 1986);
ML3/MLWin
Multilevel Regressions Modells
Multilevel Analysis
(Busing et al., 1994);
MLA
Hierarchical Linear Models
(Bryk & Raudenbush, 1992);
HLM/WHLM

Additional terms

- Mixed models
- contextual analysis
- **random coefficients models**

Software

Program		price
HLM	Raudenbush, Bryk & Congdon (2004)	395 € (Science Plus) 470 $ (ssicentral.com)
MLA	Busing, Van der Leeden & Meijer, E. (1995)	**freeware**
MlwiN	Rasbash, Browne, Goldstein, Yang et al. (2000)	880 €
mixor / mixreg / mixno / mixpreg	Hedeker & Gibbons (1996a,b)	**freeware**
VARCL	Longford (1990)	250$
MPLUS	Muthen & Muthen (2004)	745$

Multilevel Linear Regression
(Hierarchical Linear Models)

Multilevel Linear Regression

- Multilevel linear regression:
  - Purpose and basic concepts
  - Regression equations on different levels
  - Multilevel regression coefficients and their meaning
  - An example using HLM 6.0
  - Decomposing effects of a lower level predictor

Basic concepts of multilevel linear regression

- Multilevel regression analysis ("hierarchical linear models") are used to analyze effects of independent Variables on different levels on one dependent variable on the lowest level ("level 1").
- For example, you want to predict students' math achievement by their individual socioeconomic status as well as by the number of students in the class.

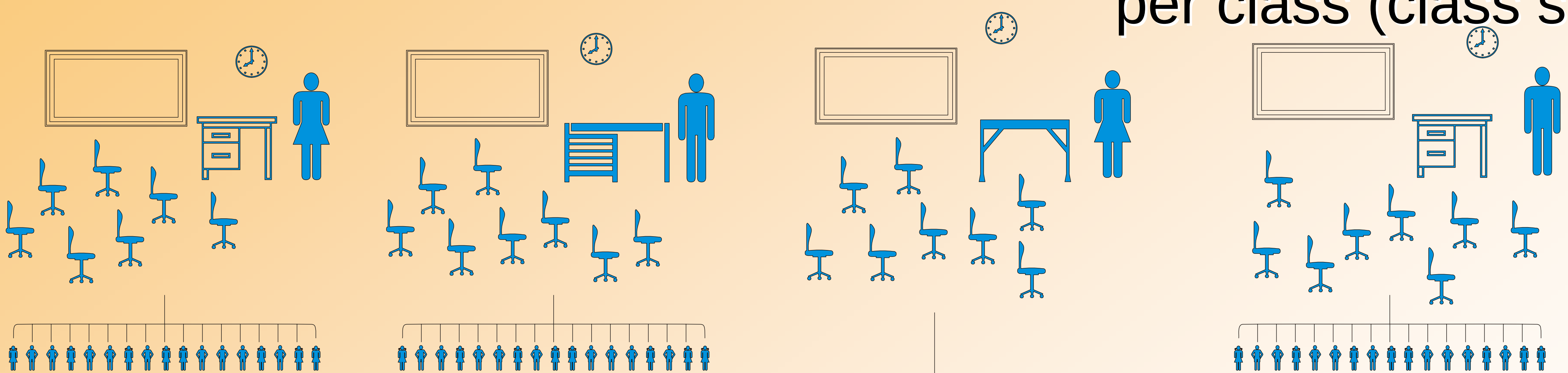
Data structure for multilevel regression analysis

- Data in the dependent variable (Y) is collected at the lowest level (level 1).
- Independent variables can be located at any level of the hierarchy.
- Units on a higher level can consist of a varying number of lower-level units.
- Statistical relations of DVs and IVs as well as relations between the hierarchical levels are represented by specific models for each level.

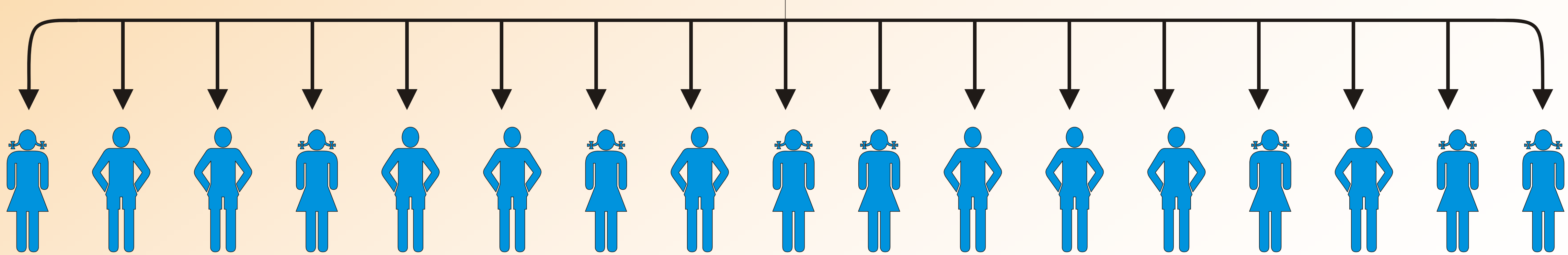
Example of a hierarchical data structure with one predictor on each level

Level 2: Classes

IV Level 2: Z = Number of students per class (class size)



Level 1: Students



DV: Y = Mathematics achievement

UV Level 1: X = socioeconomic status (SES)

Example of a hierarchical data structure with one predictor on each level

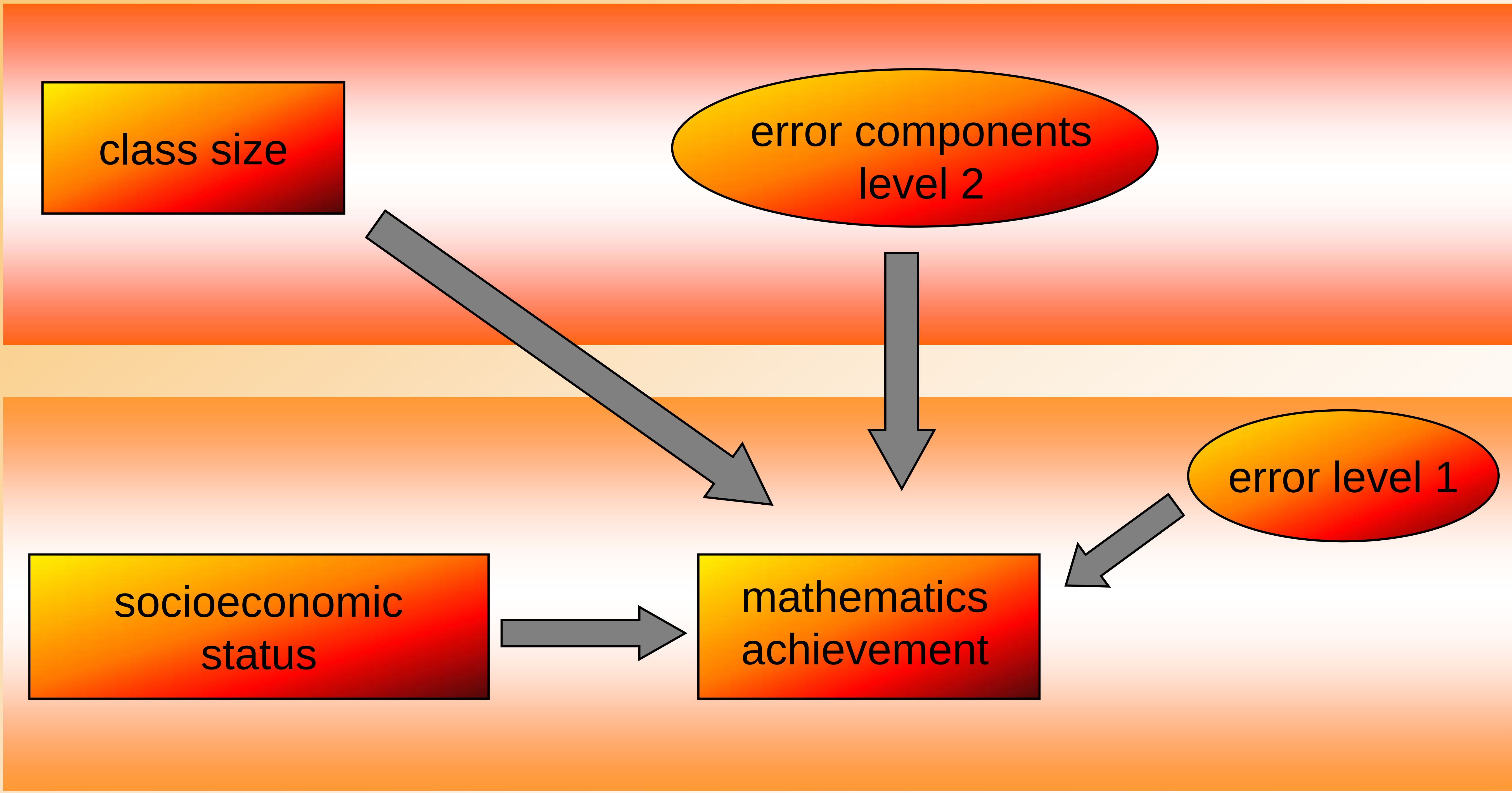
Level 1: Students

DV: Y = mathematics achievement
IV Level 1: X = socioeconomic status

Level 2: Classes

IV Level 2: Z = Class size

In multilevel regression, effects are modeled on two levels



Regression equations

- Standard linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

Y_i = dependent variable

X_i = independent variable

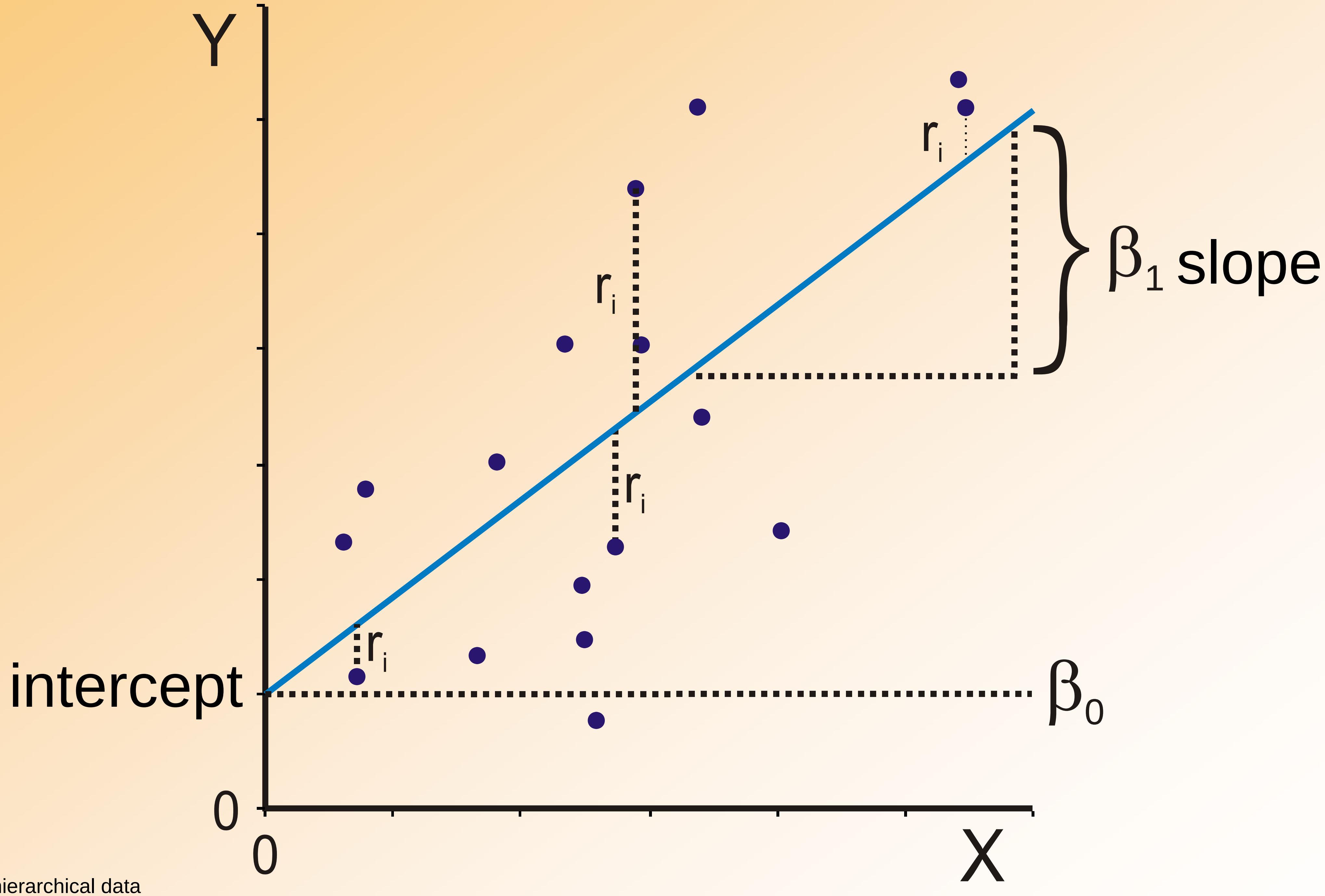
β_0 = intercept (regression constant)

β_1 = slope (regression weight of X)

r_i = residual

Standard linear regression equation

$Y_i = \beta_0 + \beta_1 X_i + r_i$



Regression equation for level 1

- Standard linear regression:

  $Y_i = \beta_0 + \beta_1 X_i + r_{ij}$

- Multilevel regression equation for level 1:

  $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$

  - β_{0j} = intercept (regression constant),
  - β_{1j} = slope,
  - r_{ij} = residual error,
  - i = subscript for level 1-unit (student),
  - j = subscript for level 2-unit (class).

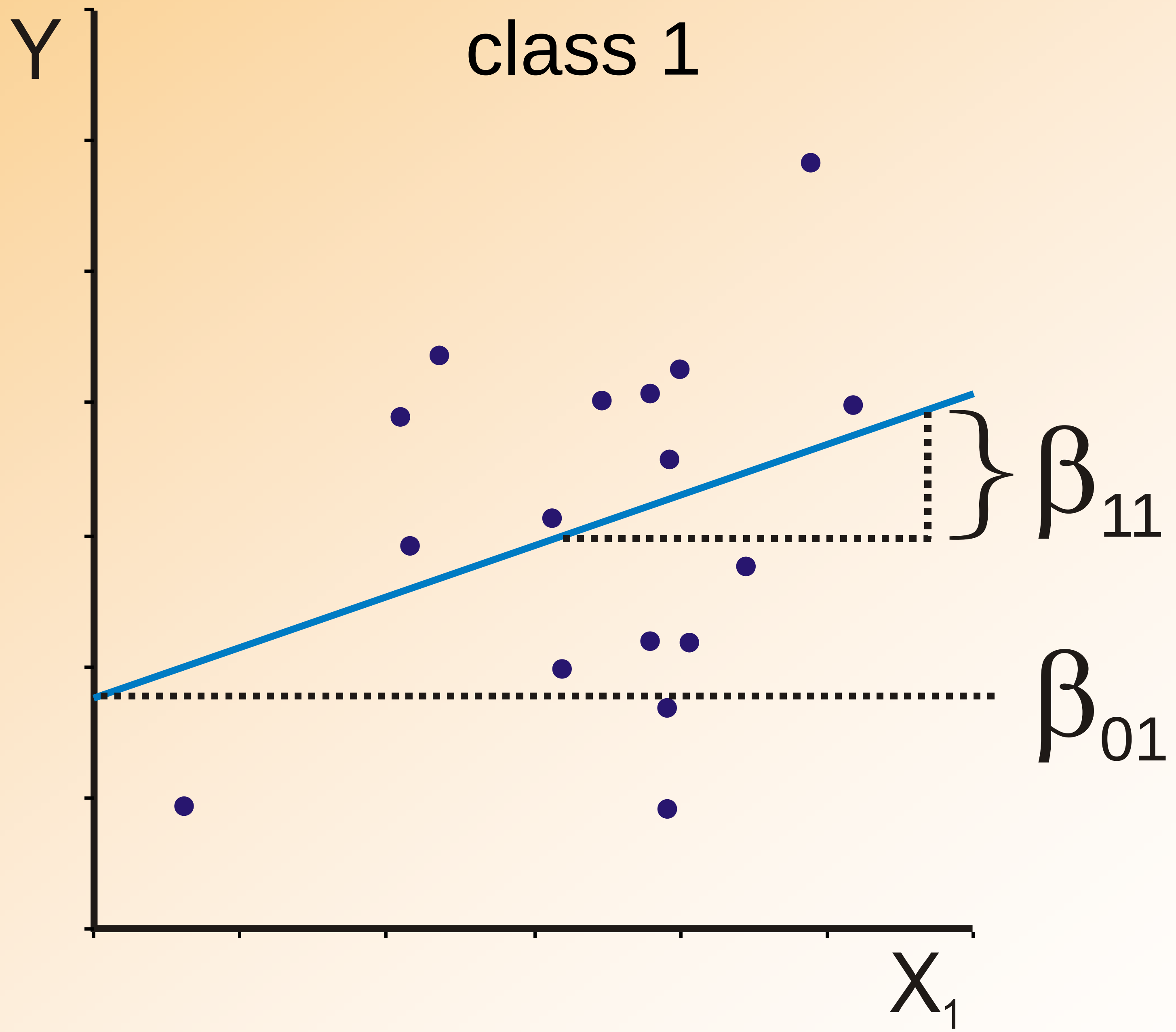
Regression equation for level 1

- Each class (level 2-unit) has its unique level 1 regression constant β_{0j};
- Each class (level 2-unit) has its unique level 1 regression slope β_{1j};

→ β_{0j} and β_{1j} vary between level 2 units.

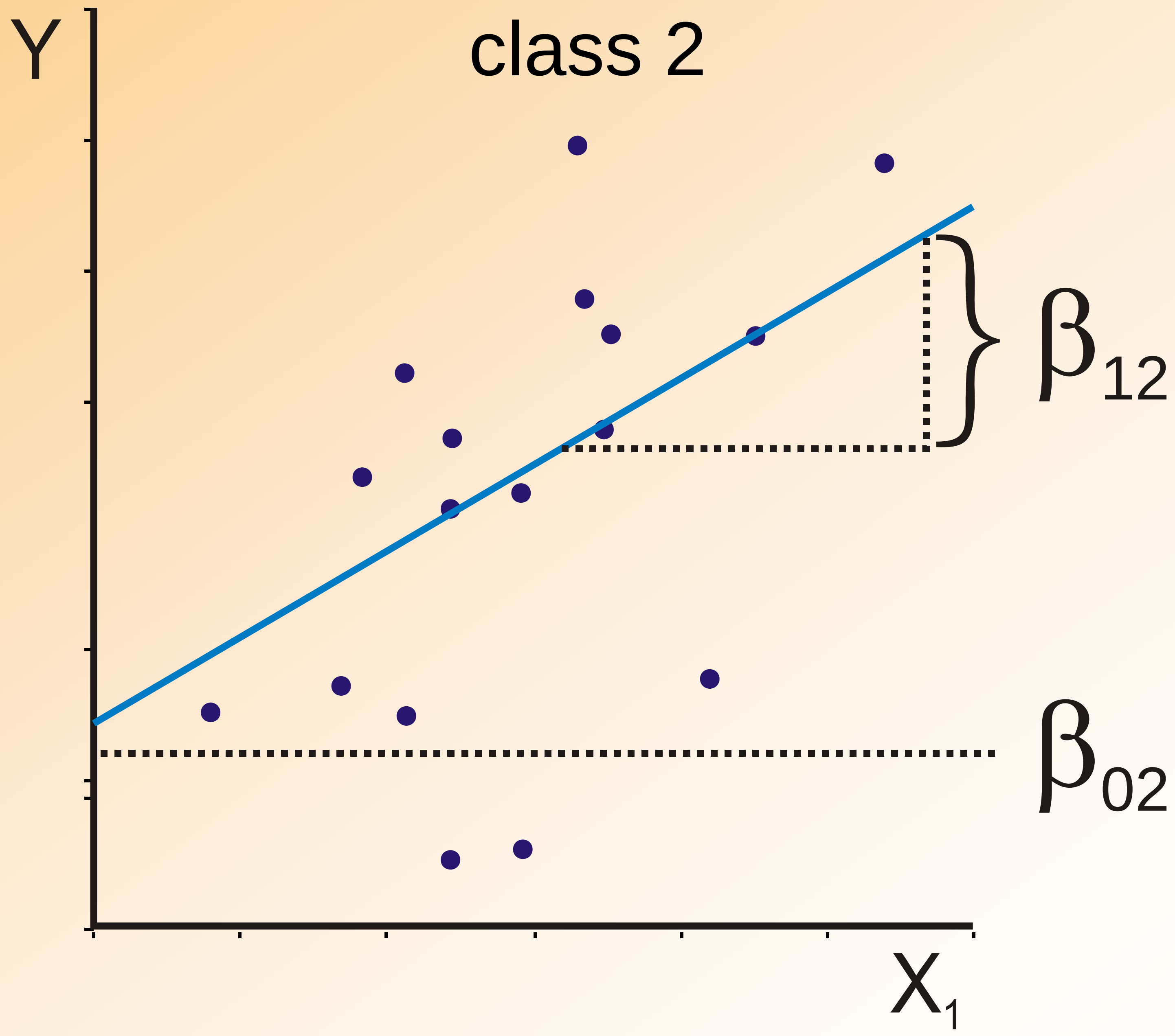
Unique regression equations for each level 2 unit

→ β_{0j} and β_{1j} vary between level 2 units



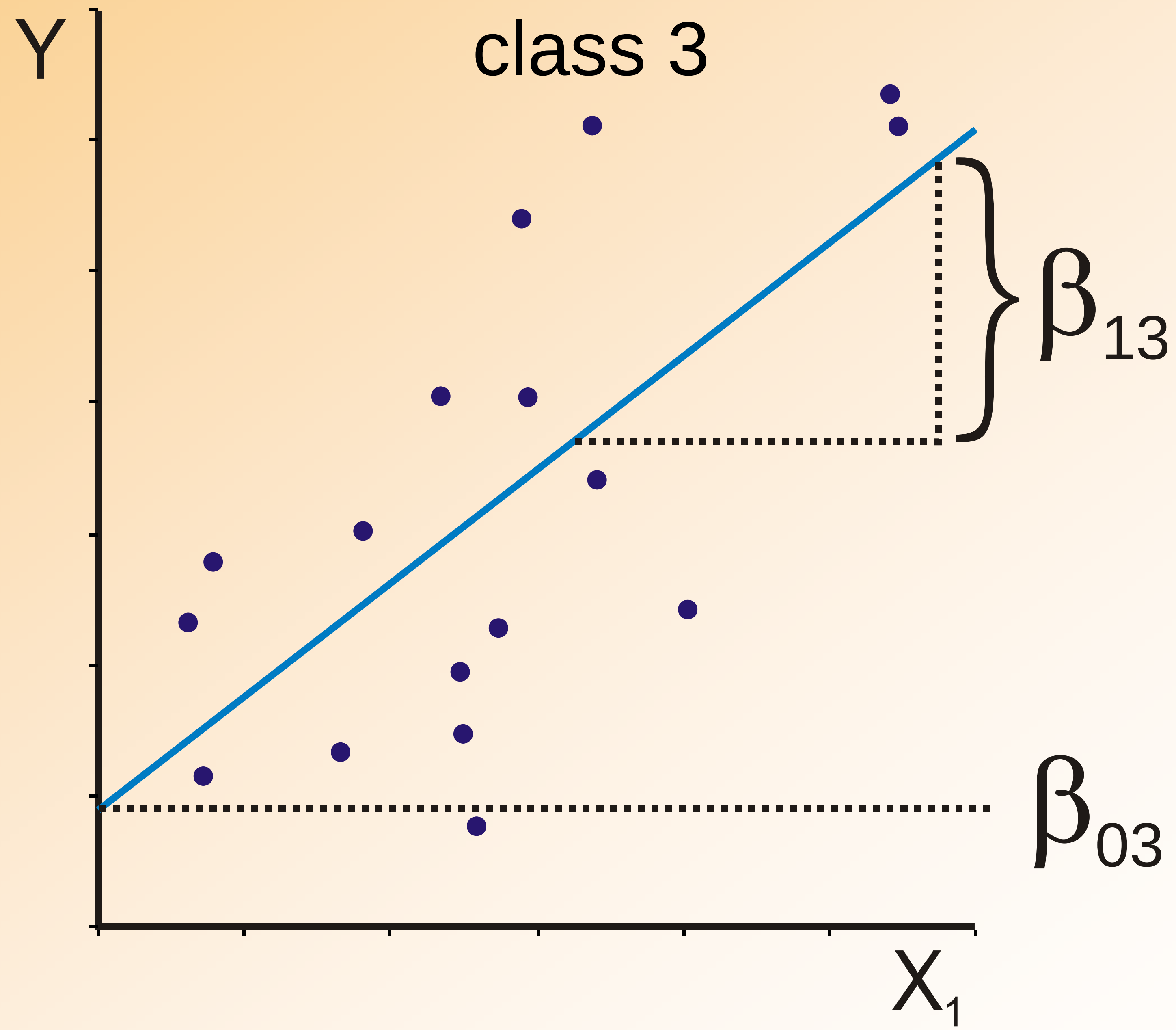
Unique regression equations for each level 2 unit

→ β_{0j} and β_{1j} vary between level 2 units



Unique regression equations for each level 2 unit

→ β_{0j} and β_{1j} vary between level 2 units



Level 2 regression equations

- Level 1 regression parameters are modeled as outcome variables in level 2 regression equations:

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$

Level 2 equations: $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$

  - γ are the level 2 regression coefficients
  - Z is a level 2 independent variable (e.g. class size as a variable measured at class level)

Level 2 regression equations

- Level 1 regression parameters are modeled as outcome variables in level 2 regression equations:

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$

Level 2 equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

- For each level 1 regression coefficient, there is one level 2 equation in a multilevel regression model.
- Level 2 regression coefficients do *not* vary across level 2 units (therefore they have no subscript j).

Fixed and random effects in multilevel regression (2 levels)

- Since level 1 regression coefficients (β_j) can vary across level 2 units, these effects are called random effects (→ “random coefficient models”)
- Level 2 regression coefficients (γ) do not vary and are referred to as fixed effects.
- In the statistical analysis of multilevel data, only fixed effects and random variances are actually estimated.

Level 2 regression equations: Level 2 regression constants γ_{k0}

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}$

Level 2 equations: $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + u_{1j}$

γ_{00} = Level 2 regression constant of β_0:
Expectation of level 1 intercept β_{0j} for Z_j being zero

γ_{10} = Level 2 regression constant of β_1:
Expectation of level 1 regression slope β_{1j} for Z_j being zero
→ the average effect of the level 1 predictor, e.g. the "overall" effect of individual SES on student performance.

Level 2 regression equations: Level 2 regression slopes γ_{k1}

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}$

Level 2 equations: $\beta_{0j} = \gamma_{00} + \boldsymbol{\gamma_{01}}Z_{1j} + u_{0j}$

$\beta_{1j} = \gamma_{10} + \boldsymbol{\gamma_{11}}Z_{1j} + u_{1j}$

In multilevel regression, variation between level 1 intercept and slopes can me predicted by level 2 independent variables Z:

- γ_{01} = effect of Z_1 on β_{0j}
- γ_{11} = effect of Z_1 on β_{1j}

Level 2 regression equations: Level 2 regression slopes γ_{k1}

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}$

Level 2 equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + u_{1j}$$

γ_{01} = effect of Z_1 on β_{0j}

The effect of Z on the regression constant is the main effect of a level 2 predictor, e.g. the effect of class size on average student performance in classes; e.g. do students in smaller classes perform higher in average?

γ_{11} = effect of Z_1 on β_{1j}

The effect of a level 2 predictor on a level 1 regression slope is called cross level interaction; e.g. is the effect of SES on achievement higher in larger classes?

Level 2 regression equations: Level 2 residuals u_{kj}

Level 1 equation: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}$

Level 2 equations: $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \mathbf{u_{0j}}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + \mathbf{u_{1j}}$

- Random variation of β_{0j} und β_{1j} between level 2 units is expressed by unique effects for each class with an expectation of zero:
  - u_{0j} = Unique effect of class j on the mean achievement β_{0j}, controlling for Z_1
  - u_{1j} = Unique effect of class j on the regression slope β_{1j}, controlling for Z_1;

Variance components in multilevel regression

$$\text{Var}(r_{ij}) = \sigma^2;\ E(r_{ij}) = 0$$

$$\text{Var}(u_0) = \tau_{00};\ E(u_{0j}) = 0$$

$$\text{Var}(u_1) = \tau_{11};\ E(u_{1j}) = 0$$

$$\text{Var}\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} = T = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}$$

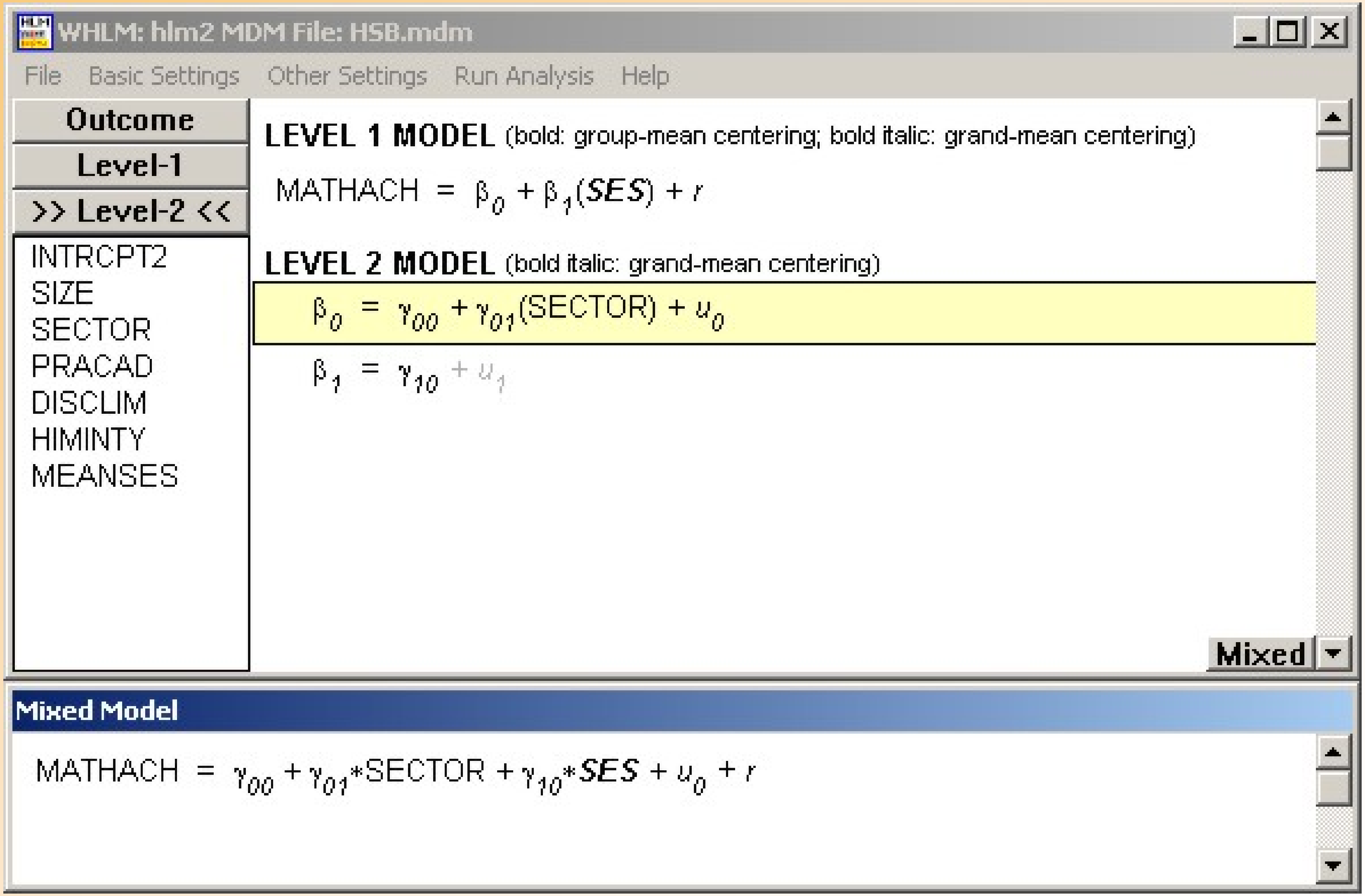
$$\text{Cov}(r_{ij}, u_j) = 0$$

Estimating variance between classes: the intercept only model

- (also null model, baseline model)
- The intercept only model contains only the level 2 regression constant γ_{00} and residuals for level 1 and 2:
- Level 1 equation: $Y_{ij} = \beta_{0j} + r_{ij}$
- Level 2 equation: $\beta_{0j} = \gamma_{00} + u_{0j}$
- → $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
- The intercept only model allows to separate variance within level 2 units from variance between level 2 units.
- Calculation of the intraclass correlation ρ:

$$\rho = \frac{\text{variance between level 2 units}}{\text{total variance}} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

HLM 6.0 examples



Data structure for analysis with HLM

- For analysis with HLM, two separate data files are needed.
- The first (level 1) contains all data collected at student level, and one ID variable indicating the belonging of each level 1 unit to a specific level 2 unit (e.g. a class ID for each student.
- The second (level 2) data set contains all data collected at class level. It consists of one “case” per class and an ID variable that is unique for each class.

Example of two level data structure for analysis with HLM

Level 2 data file

Classes.sav - SPSS Data Editor

File Edit View Data Transform Analy:

21 :

	IDclass	size
1	1224	20
2	1288	24
3	1296	26
4	1308	22
5	1317	20
6	1358	27
7	1374	23
8	1433	24
9	1436	19
10	1461	19
11	1462	20
12	1477	19
13	1499	26
14	1637	20
15	1906	27
16	1909	30
17	1942	26
18	1946	21
19	2030	27
20	2208	18
21	2277	19

Level 1 data file

Students.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities

6 :

	IDclass	gender	SES	mathach
1	1224	1	26	5.876
2	1224	1	41	19.708
3	1224	0	42	20.349
4	1224	0	40	8.781
5	1224	0	48	17.898
6	1224	0	50	4.583
7	1224	1	40	-2.832
8	1224	0	35	.523
9	1224	1	36	1.527
10	1224	0	43	21.521
11	1224	1	28	9.475
12	1224	1	40	16.057
13	1224	0	43	21.178
14	1224	1	35	20.178
15	1224	0	55	20.349
16	1224	1	40	20.508
17	1224	0	45	19.338
18	1224	0	26	4.145
19	1224	1	51	2.927
20	1224	0	49	16.405
21	1288	1	38	7.857
22	1288	0	45	10.171
23	1288	0	57	15.699

HLM 6 output: intercept only model

Example

- “High school and beyond” data (example included in the free HLM 6.0 student version).
- Data is from students drawn from schools.
- Independent variable: math achievement.
- Intercept only model (null model) without independent variables on level 1 or 2

HLM 6 output: intercept only model

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

      Y = B0 + R

Level-2 Model
      B0 = G00 + U0
```

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$PV1READ = \beta_0 + r$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

☒ Error term for currently selected level-2 equation

☒ $\beta_0 = \gamma_{00} + u_0$

HLM 6 output: intercept only model Fixed effects (γ-coefficients)

```
The outcome variable is  MATHACH

Final estimation of fixed effects
(with robust standard errors)
----------------------------------------------------------------------------
                                       Standard             Approx.
   Fixed Effect         Coefficient   Error       T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00          12.636972   0.243628    51.870       159    0.000
----------------------------------------------------------------------------
```

The only fixed effect is the level 2 regression constant γ_{00}, which is typically not very interesting…

HLM 6 output: intercept only model variance components

```
 Final estimation of variance component
 ---------------------------------------
Random Effect              Standard      Varian          df     Chi-square
P-value
                            Deviation     Compo     nt
 -----------------------------------------------------------------------
INTRCPT1,        U0          2.93501       8.61431   159    1660.23259
0.000
                                           39.14831
                                         ------------------------------
```

Level 1 residual variance

18% of the variance in math achievement is between schools, i.e. can be explained by differences between schools

Level 2 residual variance

$$ICC = \frac{\mathrm{Var}(u_0)}{\mathrm{Var}(u_0) + \mathrm{Var}(r)} = \frac{8.61}{39.15 + 8.61} = 0.18$$

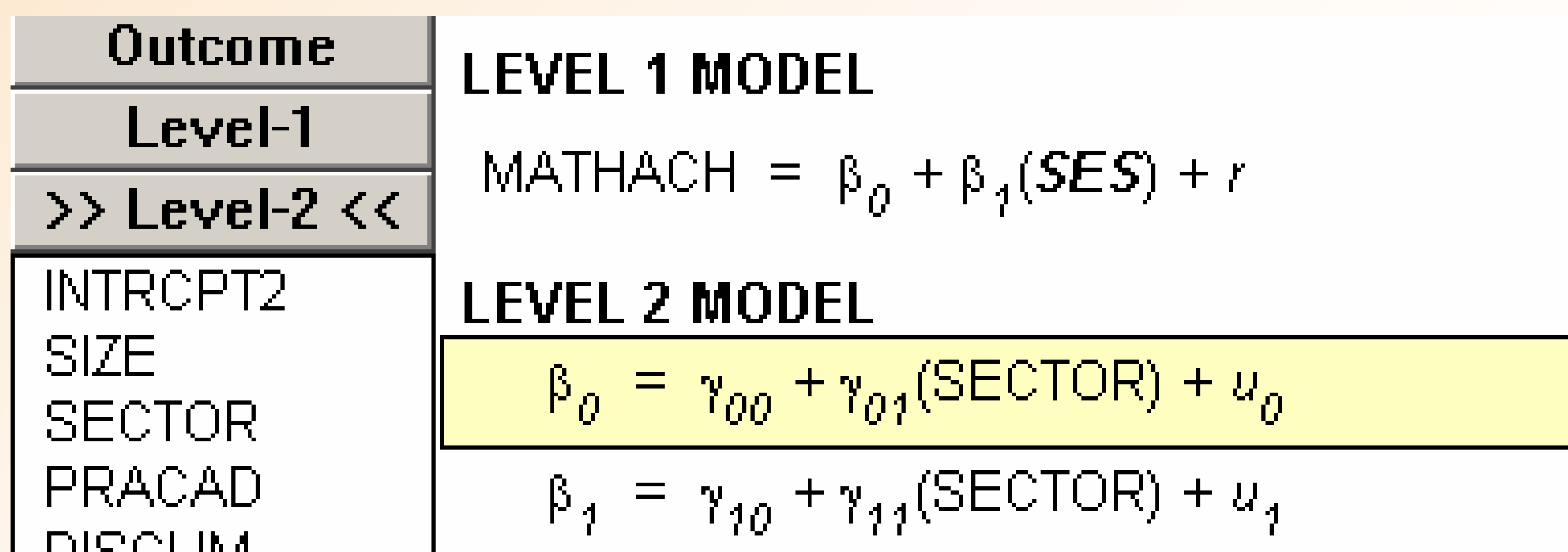
HLM 6 output: complete model

Example

- "High school and beyond" data.
- Independent variable Y: math achievement.
- One level 1 predictor:
  - Student SES (SES)
- One level 2 predictor:
  - type of school (`sector` = catholic vs. public)
- One cross level interaction
  - The type of school moderates the relation between SES and math achievement.

HLM 6 output: complete model

```
Summary of the model specified (in equation format)
---------------------------------------------------

Level-1 Model

      Y = B0 + B1*(SES) + R

Level-2 Model
      B0 = G00 + G01*(SECTOR) + U0
      B1 = G10 + G11*(SECTOR) + U1
```



HLM 6 output complete model: fixed effects (γ-coefficients)

```
 The outcome variable is  MATHACH

                      xed effects
                      errors)
                      ----------------------------------------
                                   Standard          Approx.
                      Coefficient  Error    T-ratio  d.f.    P-value
---------------------------------------------------------------------
For       INTR   T1, B0
   INTRCPT2, G00         11.750661  0.218684  53.733     158   0.000
     SECTOR, G01          2.128423  0.355700   5.984     158   0.000
For       SES slope, B1
   INTRCPT2, G10          2.9587
     SECTOR, G11
---------------------------------
```

β_0

γ_{00} average performance in public schools

γ_{01} level 2 main effect for school type (performance difference catholic and public schools)

HLM 6 output complete model: fixed effects (γ-coefficients)

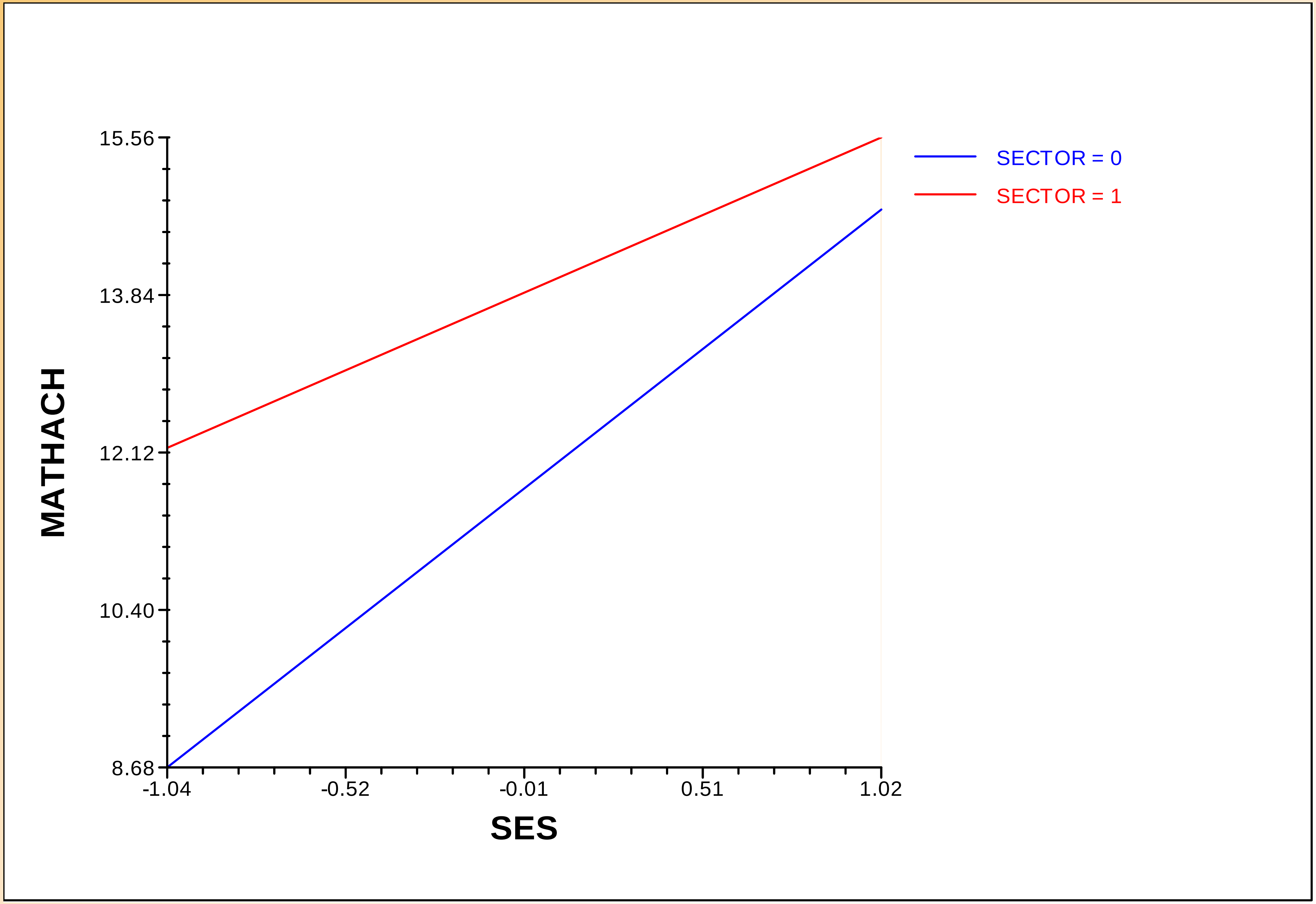
```
 The outcome variable is  MATHACH

Final estimation of fixed effects
(with robust standard errors)
------------                                        -------------------
                                                       Approx.
   Fixed Eff                                         o d.f.     P-value
------------                                        -------------------
For       INTRCPT1
   INTRCPT2, G00         11.750661   0.218684  53.733       158    0.000
     SECTOR, G01          2.128423   0.355700   5.984       158    0.000
For      SES slop , B1
   INTRCPT2, G10          2.958798   0.144092  20.534       158    0.000
     SECTOR, G11         -1.313096   0.214271  -6.128       158    0.000
----------------------   --------------------------------------------
```

β_1

γ_{10} main effect of SES in public schools (`sector`=0)

γ_{11} change in SES effect if school is catholic (`sector`=1)

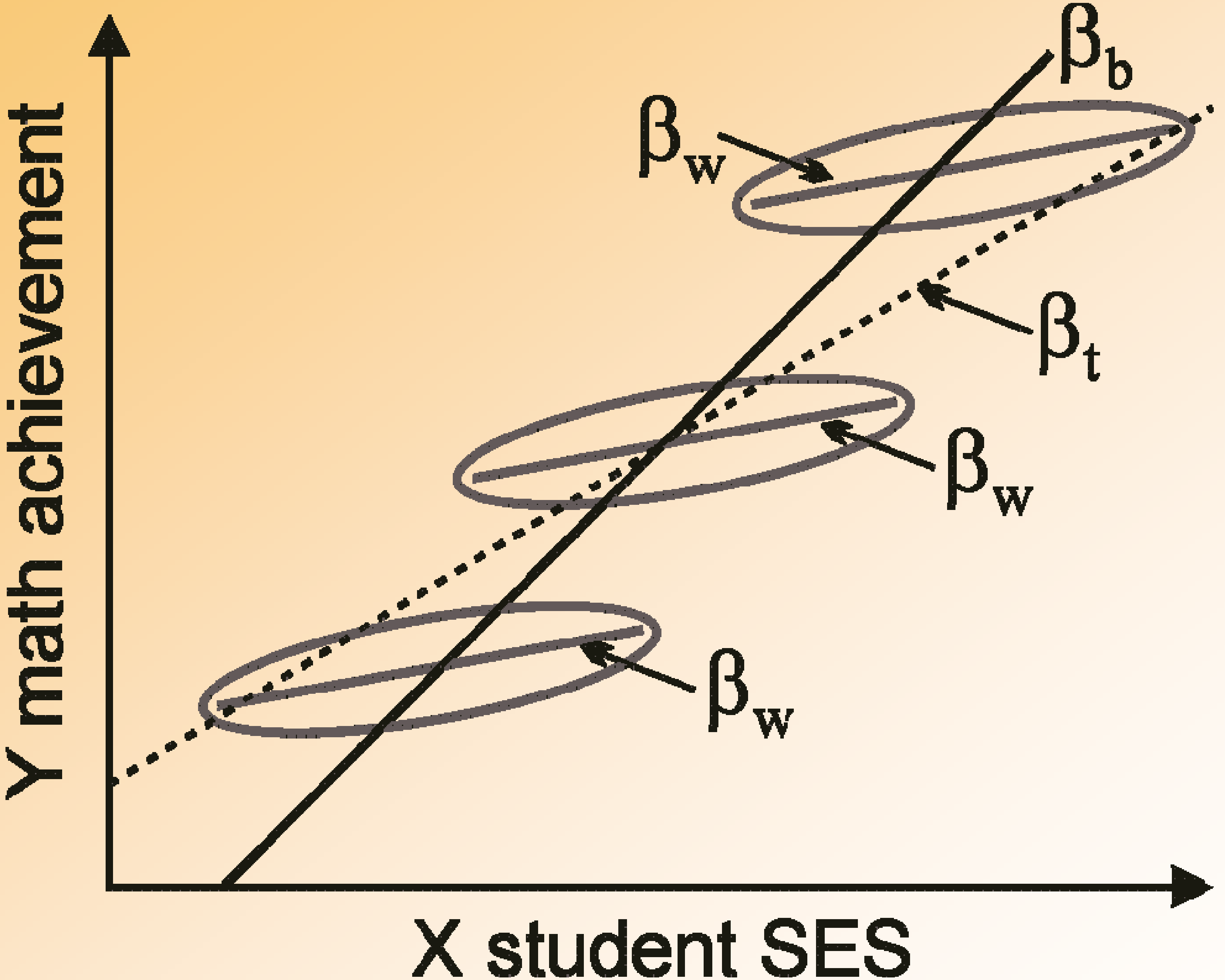
HLM 6 output complete model: graphical display of cross level interaction



Compositional effects

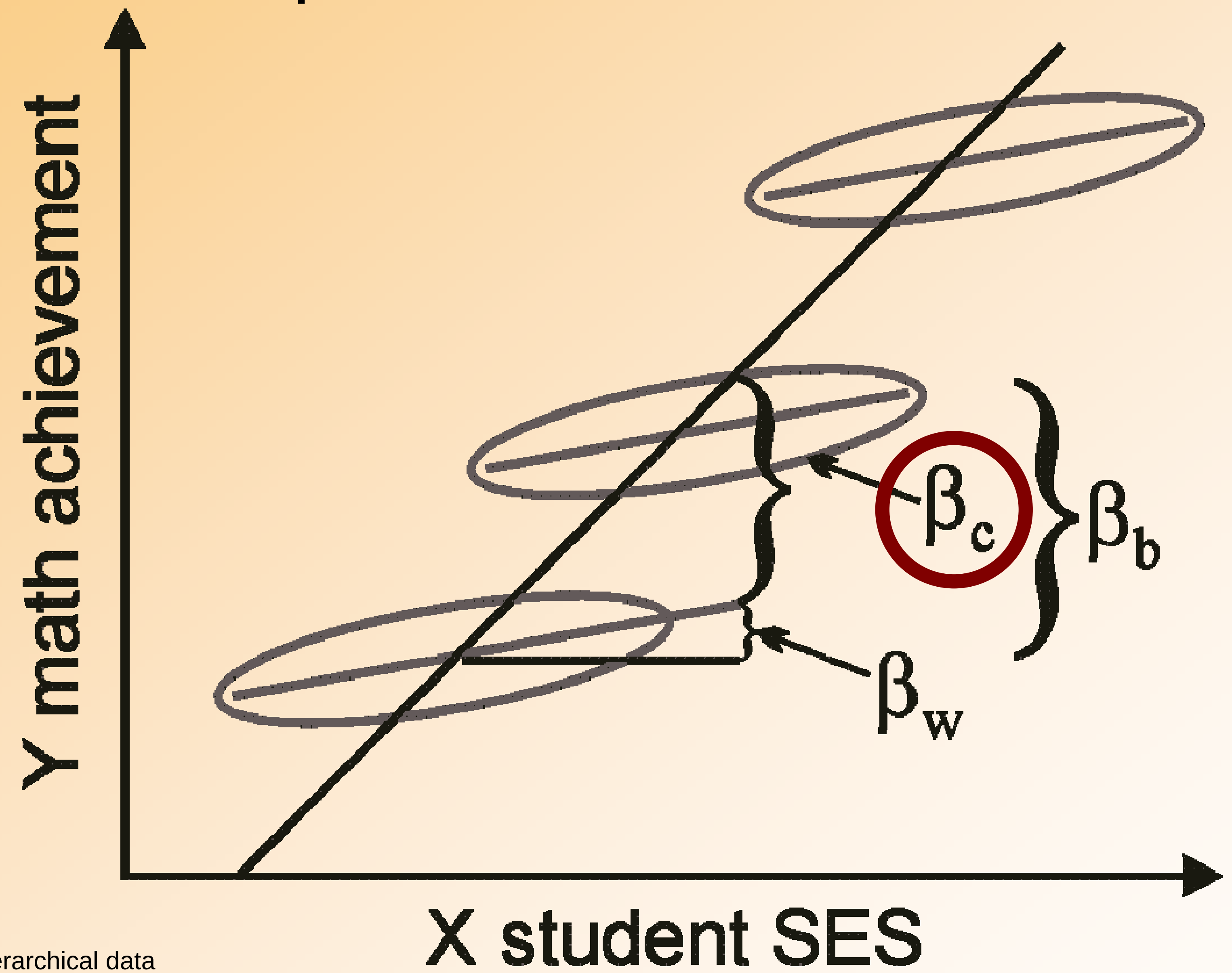
- „The statistical estimate of the additional effect obtained by the aggregated variable at the school level over-and-above the variable's effect at the individual level" (Harker & Tymms, in press)
- Multilevel regression allows the decomposition of the effect of an independent on a dependent variable into
  - effects within level 2 units
  - effects between level 2 units

Effects of a level 1 predictor within and between level 2 units



Graphical illustration of compositional effects

- Compositional effects occur if level 2 units are heterogeneous with respect to the dependent as well to the independent variable.



β_w, β_b,and β_c can be estimated separately within multilevel regressuion.

Multilevel structural equation modeling

Basic concepts of structural equation modeling

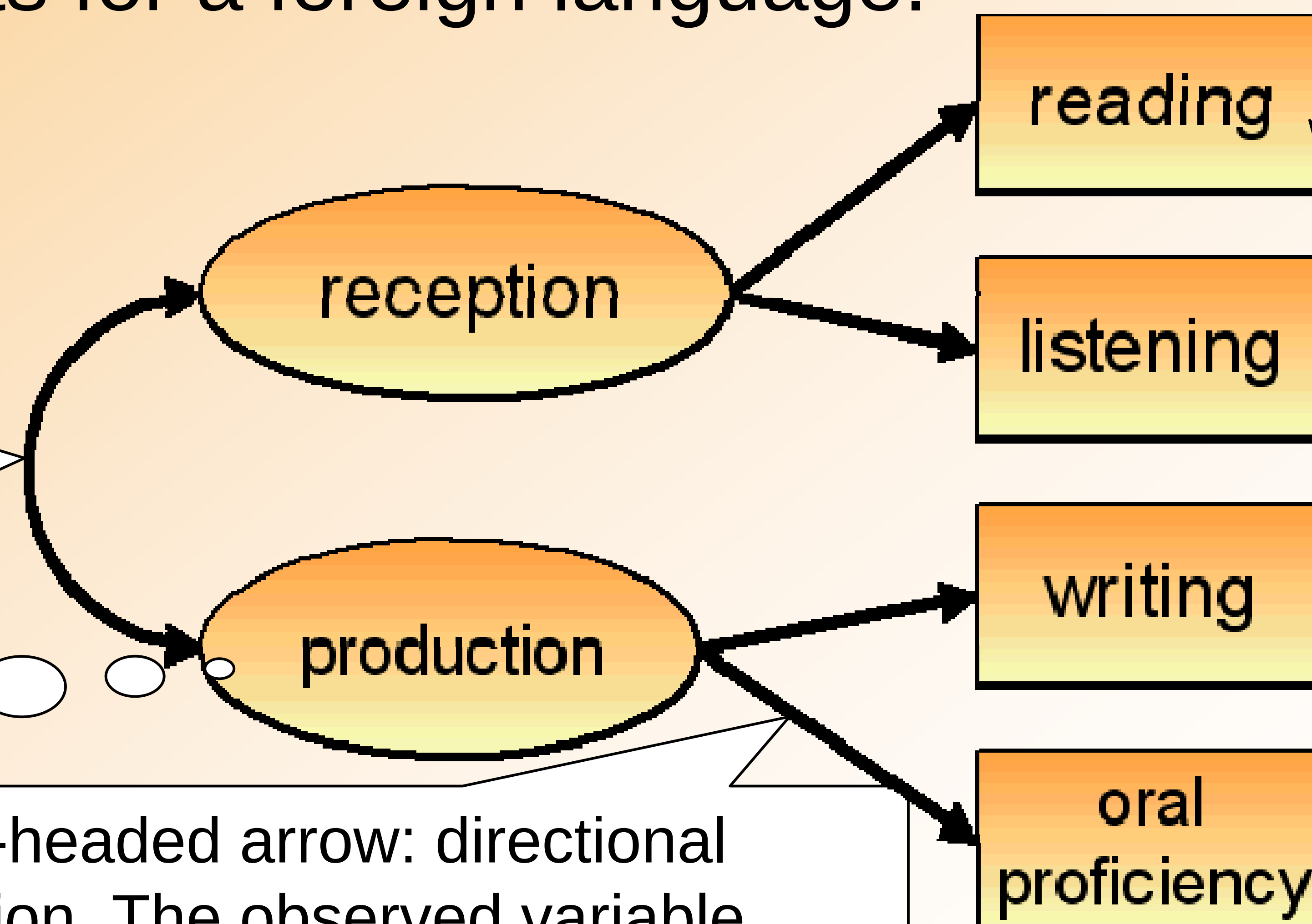
- In structural equation modeling, the observed correlations between variables (e.g. test scores) are explained by underlying latent variables.
- These latent variables are theoretical constructs, variables assumed to be inherently unobservable, but which are supposed to be useful concepts to describe and explain behavior in a specific range of observable phenomena.

Basic concepts of structural equation modeling

- Example: Two basic language skills for language reception and language production underliy the observed performance in tests for a foreign language.

Double-headed arrow: undirected effect. The latent skills “reception” and “production” are related in some way, but we do not care why or how.

Bubbles: Latent variables



Boxes: Observed or “manifest” variables

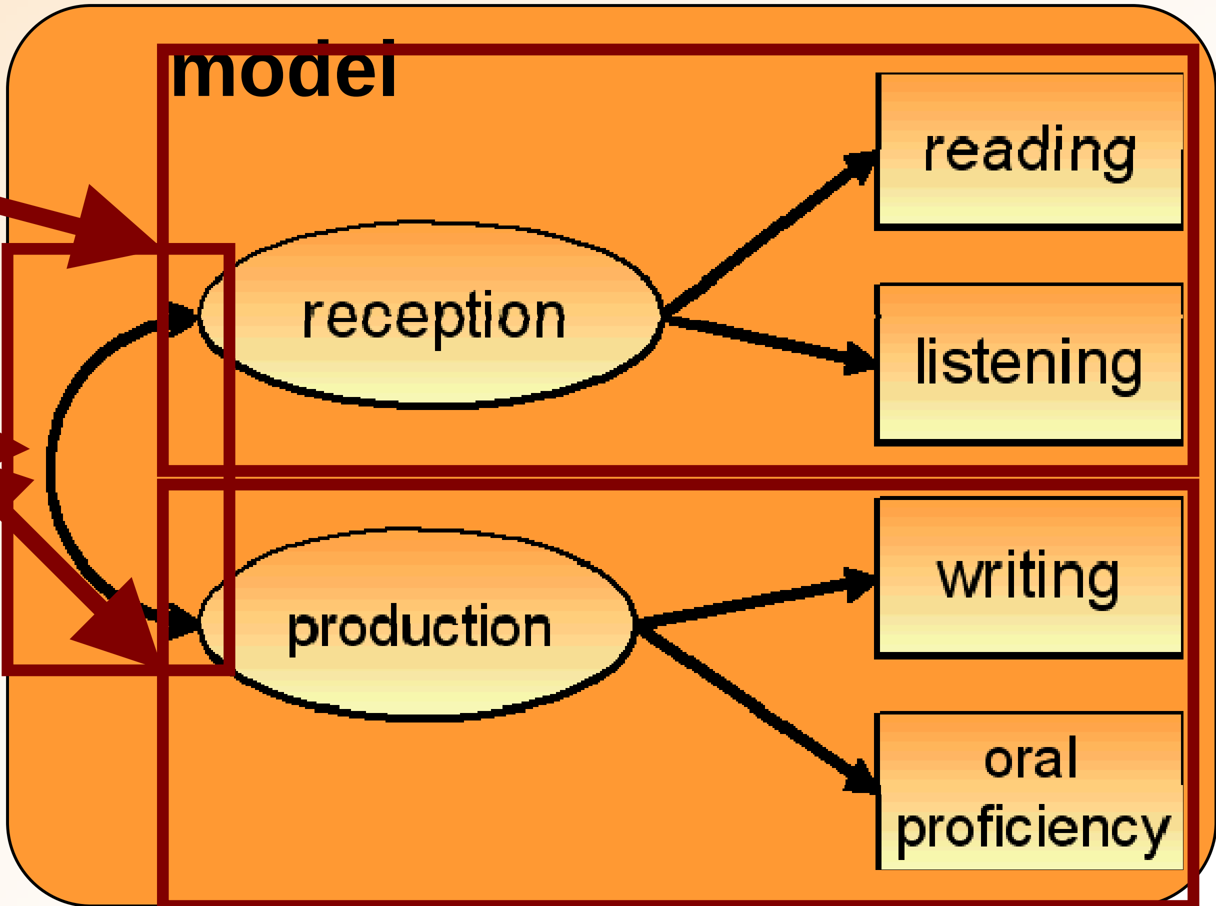
One-headed arrow: directional relation. The observed variable “reading” is influenced by the latent skill “reception”

Basic concepts of structural equation modeling

- Information about latent variables is derived from the empirical correlations of the observed variables.

empirical correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.75	1.00		
writing	0.30	0.30	1.00	
oral prof.	0.30	0.30	0.75	1.00



Basic concepts of multilevel structural equation modeling

- Structural equation models are based on empirical correlations.
- If the empirical data is collected in a multilevel structure, the correlations are a mixture of within and between group effects.
- In this case, it is advisable to separate these effects.
- To do so, the correlations between the observed variables are decomposed in correlations between and within groups.
- In multilevel structural equation modeling, separate models are fitted to the within- and between group correlations.

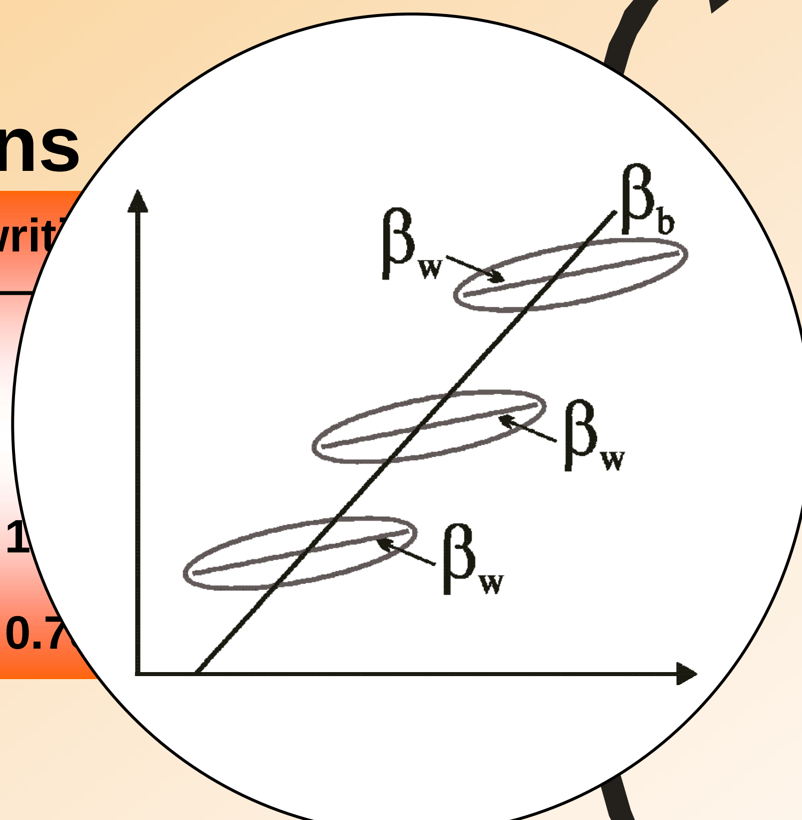
Separating Correlations within and between groups

between- group correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.88	1.00		
writing	0.88	0.88	1.00	
oral prof.	0.88	0.88	0.88	1.00

observed correlations

	read.	listen.	writ
reading	1.00		
listening	0.75	1.00	
writing	0.30	0.30	1
oral prof.	0.30	0.30	0.7



within- group correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.65	1.00		
writing	0.25	0.25	1.00	
oral prof.	0.25	0.25	0.65	1.00

Separating Correlations within and between groups

within- group correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.65	1.00		
writing	0.25	0.25	1.00	
oral prof.	0.25	0.25	0.65	1.00

→ Correlations of student performance within classes, i.e. controlling for average class performance.

between- group correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.88	1.00		
writing	0.88	0.88	1.00	
oral prof.	0.88	0.88	0.88	1.00

→ Correlations between average class performances across all classes.

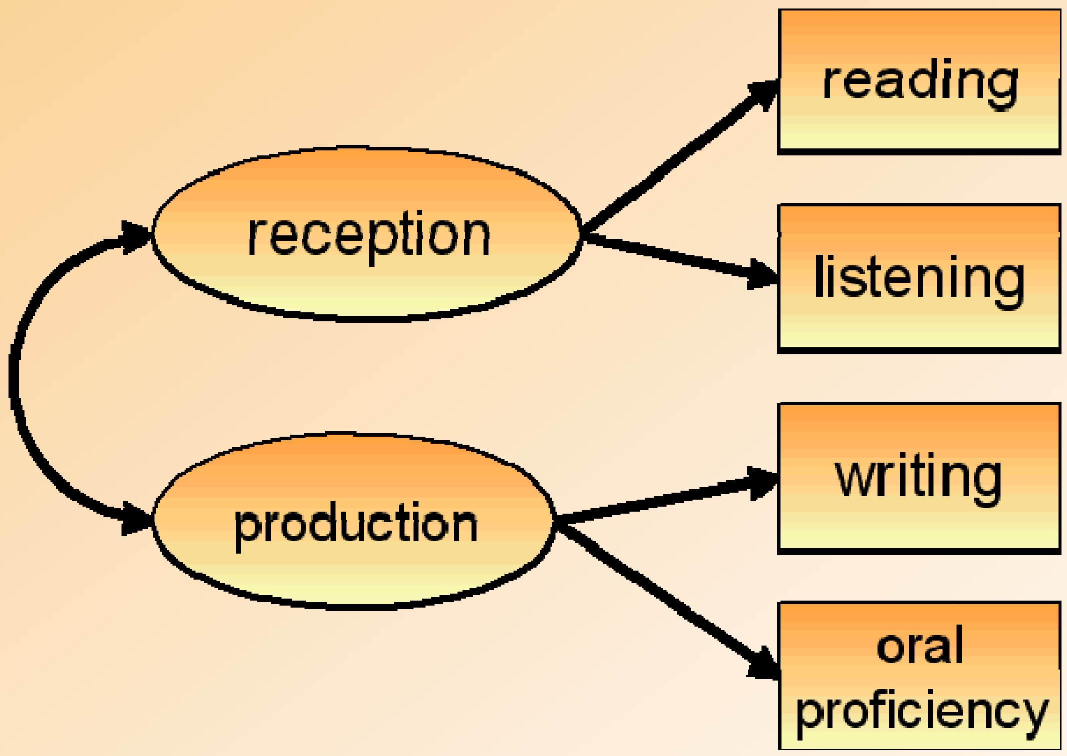
Building separate models for each levels

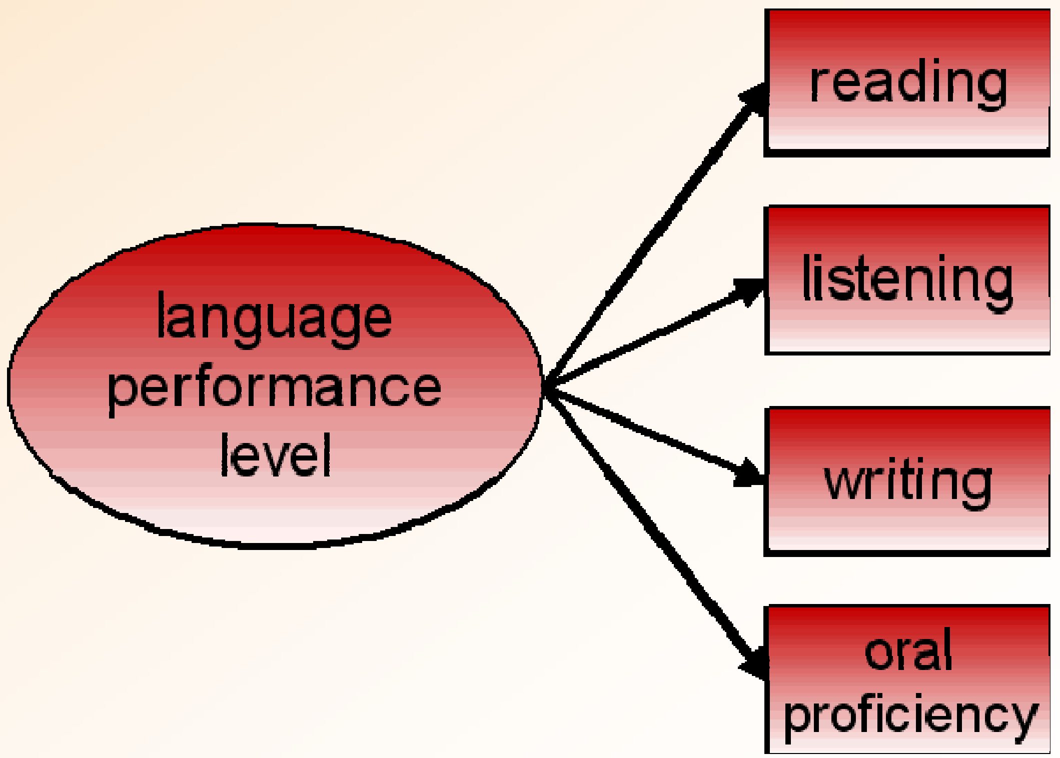
within- group correlations

	read.	listen.	writing	oral
reading	1.00			
listening	0.65	1.00		
writing	0.25	0.25	1.00	
oral prof.	0.25	0.25	0.65	1.00

between- group correlations

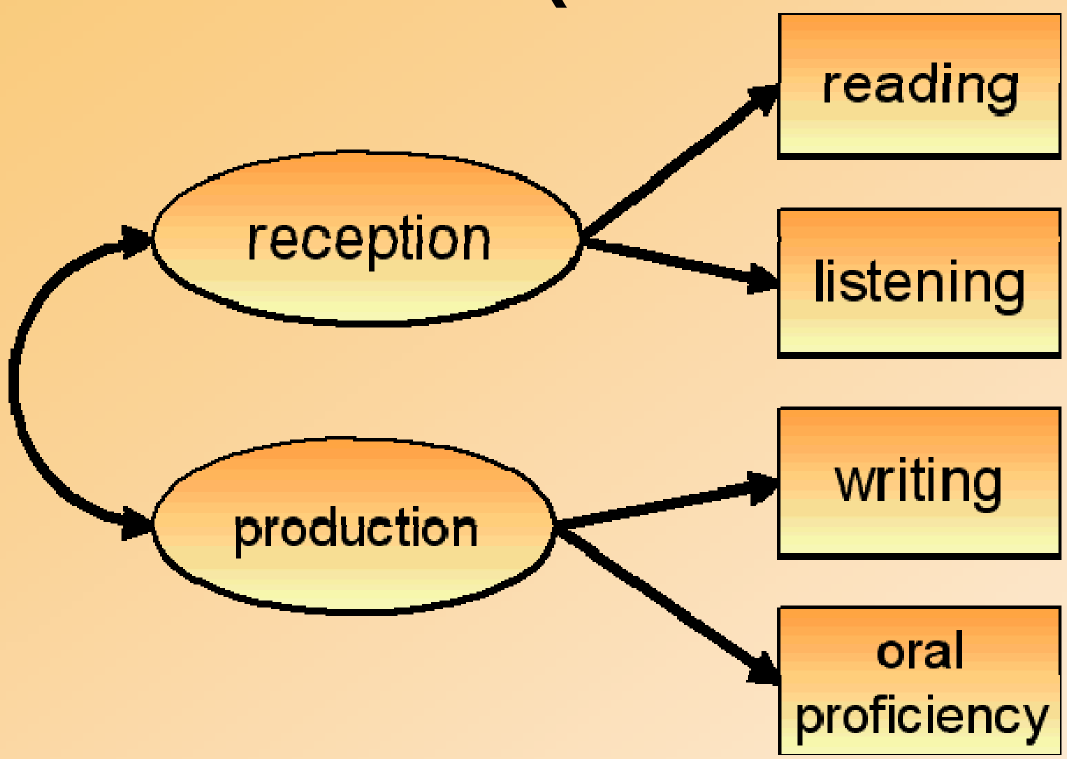
	read.	listen.	writing	oral
reading	1.00			
listening	0.88	1.00		
writing	0.88	0.88	1.00	
oral prof.	0.88	0.88	0.88	1.00





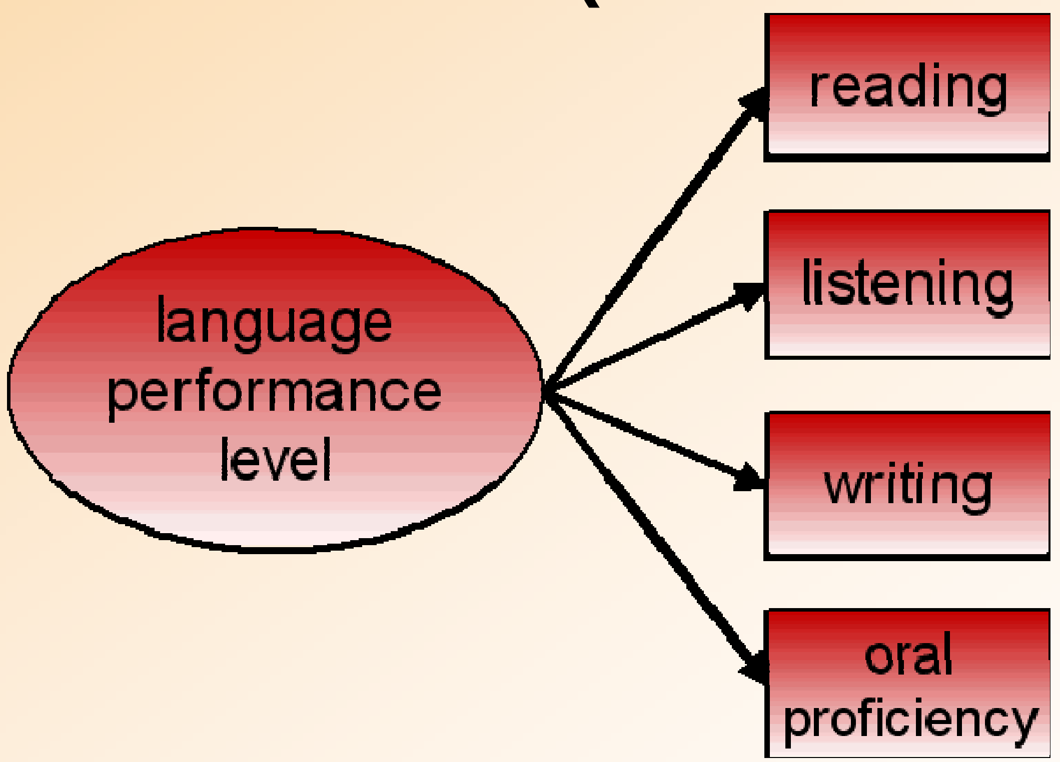
Building separate models for each levels

Level 1 model (within classes)



→ Model for relations of students' individual skills.

Level 2 model (between classes)



→ Model for relations between performance levels of whole classes.

Summary

- Hierarchical data is a common phenomenon in educational research
- Conventional statistical analysis (e.g. linear regression, ANOVA) of multilevel may lead to biased results.
- Multilevel regression analysis allows to examine effects of predictors on lower as well as higher data levels on one single outcome variables.
- In multilevel equation modeling, correlations between observed variables are decomposed in correlations within and between groups. For each level, a separate latent variable model is tested.

Graphical illustration of a two level regression model