Introduction to Multilevel Modeling

Concepts, Applications, and Resources

Enrico Perinelli

Department of Psychology and Cognitive Science

University of Trento

UNIVERSITÀ DI TRENTO
Dipartimento di Psicologia e Scienze Cognitive

Lecture for the *Master in Human Resources Management and Development*
Department of Economic, Business and Statistical Sciences
University of Palermo (Italy)
July 21-22, 2025

Table of contents I

1. Welcoming

1. Welcoming

Welcome to the course “Introduction to Multilevel Modeling”.

First of all, let me introduce myself.

My name is **Enrico Perinelli**, I am a Tenure Track Researcher in Work and Organizational Psychology (WOP).

My research is mainly focused on the intersections between

- WOP
- Psychometrics/Data Science
- Personality Psychology
- Educational Psychology

1.1 My Education and Career

- A snapshot of my education/career:
  - 2010, B.S. in Psychology at University of L'Aquila
  - 2013, M.S. in Clinical Psychology at University of Bologna
  - 02/2018, PhD in Personality and Organizational Psychology at Sapienza University of Rome
  - Since 09/2018 I do research and teaching at University of Trento
- See my `curriculum` or my `Google Scholar` page for more information.

1.1 My Education and Career

- A snapshot of my education/career:
  - 2010, B.S. in Psychology at University of L'Aquila
  - 2013, M.S. in Clinical Psychology at University of Bologna
  - 02/2018, PhD in Personality and Organizational Psychology at Sapienza University of Rome
  - Since 09/2018 I do research and teaching at University of Trento
- See my `curriculum` or my `Google Scholar` page for more information.

Some of my papers at the intersection between Psychometrics and WOP

1.2 Information about our course

- **Course overview:** Introduction to requirements, the interplay between substantive and methodological considerations, and Simpson's paradox.
- **Key concepts:** Definitions, equations, centering techniques, reliability indices, and other foundational elements in multilevel modeling.
- **Reference materials:** A curated list of resources and textbooks for continued learning and application.
- **Hands-on activities:** Working with both prepared datasets and your own data; optional interpretation of empirical articles of your interest.

2. Introduction

2. Introduction

Climbing the ladder of (multilevel) complexity:

1. Load, manipulate, preprocess data (80% of the time in a data science project!!)
2. Descriptive statistics
3. Correlations
4. Inferential Statistics
5. GLM, GzLM, GLMM, GzLMM
6. Psychological Measurement (reliability and validity)
7. Structural Equation Modeling
8. Multilevel SEM and Longitudinal SEM
9. Dynamic SEM

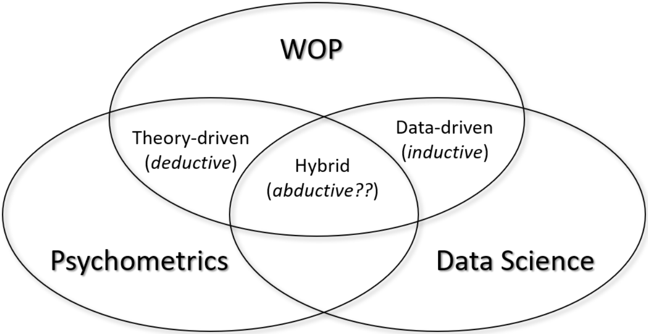
Software:

- **R**: open-source, flexible, and powerful for most statistical analyses (e.g., in data science, statistics, and psychometrics).
- **M*plus***: particularly suited for latent variable modeling.

A common suggestion is to always use open and customizable software (e.g., R and Python), and to limit the use of commercial software to operations not available in open-source tools.

In the case of multilevel modeling, R is good enough for much of operations, with exception of more advanced Multilevel SEM (e.g., Multilevel SEM with prediction on random slopes and DSEM; see `https://lavaan.ugent.be/tutorial/multilevel.html`)

WOP is historically integrated in classical **Psychometrics** models, but is also rapidly integrating with **Data Science**. See https://github.com/EnricoPerinelli/AIP-2024_YoungKeynote/blob/master/AIP_2024_YoungKeynote.pdf for a discussion.



Multilevel modeling could be considered *theory-driven models*

2.1 Substantive-methodological synergy

PSYCHOMETRIKA—VOL. 71, NO. 3, 425–440
SEPTEMBER 2006
DOI: 10.1007/s11336-006-1447-6

THE ATTACK OF THE PSYCHOMETRICIANS

DENNY BORSBOOM
UNIVERSITY OF AMSTERDAM

This paper analyzes the theoretical, pragmatic, and substantive factors that have hampered the integration between psychology and psychometrics. Theoretical factors include the operationalist mode of thinking which is common throughout psychology, the dominance of classical test theory, and the use of "construct validity" as a catch-all category for a range of challenging psychometric problems. Pragmatic factors include the lack of interest in mathematically precise thinking in psychology, inadequate representation of psychometric modeling in major statistics programs, and insufficient mathematical training in the psychological curriculum. Substantive factors relate to the absence of psychological theories that are sufficiently strong to motivate the structure of psychometric models. Following the identification of these problems, a number of promising recent developments are discussed, and suggestions are made to further the integration of psychology and psychometrics.

Key words: Psychometrics, modern test theory, classical test theory, construct validity, psychological measurement

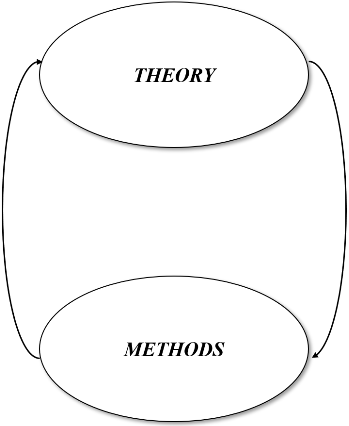
Industrial and Organizational Psychology (2021), **14**, 497–504
doi:10.1017/iop.2021.111

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

The baby and the bathwater: On the need for substantive–methodological synergy in organizational research

Joeri Hofmans[1*], Alexandre J. S. Morin[2], Heiko Breitsohl[3], Eva Ceulemans[4], Léandre Alexis Chénard-Poirier[5], Charles C. Driver[6], Claude Fernet[7], Marylène Gagné[8], Nicolas Gillet[9], Vicente González-Romá[10], Kevin J. Grimm[11], Ellen L. Hamaker[12], Kit-Tai Hau[13], Simon A. Houle[2], Joshua L. Howard[14], Rex B. Kline[2], Evy Kuijpers[1], Theresa Leyens[1], David Litalien[15], Anne Mäkikangas[16], Herbert W. Marsh[17], Matthew J. W. McLarnon[18], John P. Meyer[19], Jose Navarro[20], Elizabeth Olivier[21], Thomas A. O'Neill[22], Reinhard Pekrun[23], Katariina Salmela-Aro[24], Omar N. Solinger[25], Sabine Sonnentag[26], Louis Tay[27], István Tóth-Király[2], Robert J. Vallerand[5], Christian Vandenberghe[28], Yvonne G. T. van Rossenberg[29], Tim Vantilborgh[1], Jasmine Vergauwe[30], Jesse T. Vullinghs[1], Mo Wang[31], Zhonglin Wen[32], and Bart Wille[30]



"Multilevel" matters in WOP theories, research, and practice

> *"Organizations are multilevel systems in which organizational entities (e.g., employees, teams, departments, organizations) reside in nested arrangements (e.g., employees are nested in teams, teams in departments, and departments in organizations) [...] [The] single-level perspective is limited because it cannot explain the complexities of most organizational phenomena, where the antecedents, mediators, moderators, and outcomes involved reside at different levels"*

(González-Romá & Hernández, 2017, p. 184)

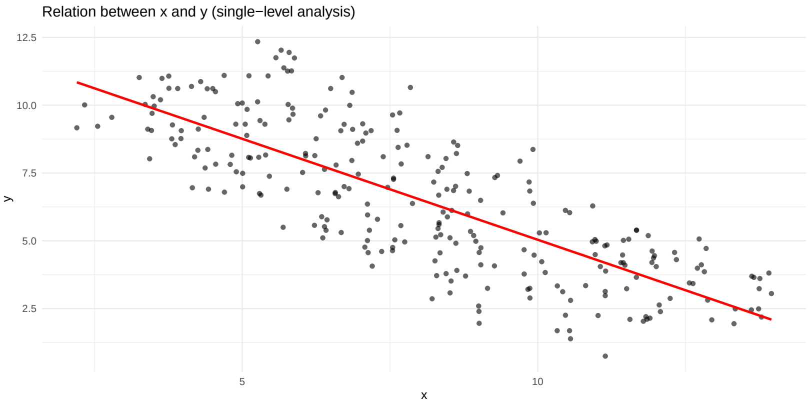
Why should we go over simple single-level (e.g., a simple linear regression) analysis or simple single-level theoretical models?

The consequences of the violation of the non-independence of observations, which is created by the presence of "clusters" in the data (e.g., countries, departments, schools, leaders, etc.), is well-represented by the Simpson's paradox (Simpson, 1951)

2.2 The Simpson's paradox

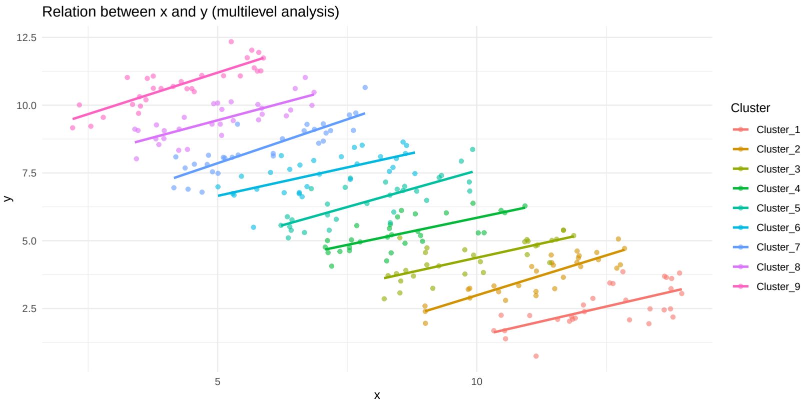
```r
library(ggplot2)
library(dplyr)
library(tibble)
library(patchwork)

set.seed(42)

# Parameters
n_clusters <- 9; n_per_cluster <- 30

# Simulation
simpson_data <- bind_rows(lapply(1:n_clusters, function(i) {
  x_shift <- 10 - i
  x <- runif(n_per_cluster, 1, 5)
  y <- 0.5 * x + i + rnorm(n_per_cluster, 0, 0.5)
  tibble(
    country = paste0("Cluster_", i),
    x = x + x_shift,
    y = y
  )}))

p1 <- ggplot(simpson_data, aes(x = x, y = y)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relation between x and y (single-level analysis)",
       x = "x", y = "y") +
  theme_minimal()
```

```
p1
```

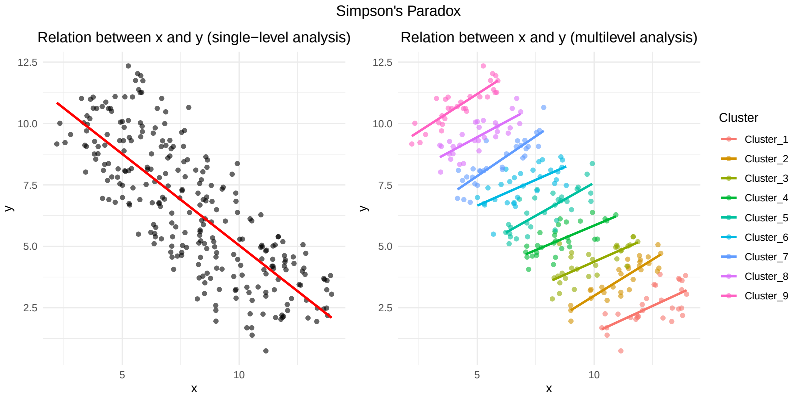

Relation between x and y (single−level analysis)
12.5
10.0
7.5
5.0
2.5
y
5
10
x

```r
p2 <- ggplot(simpson_data, aes(x = x, y = y, color = country)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relation between x and y (multilevel analysis)",
       x = "x", y = "y", color = "Cluster") +
  theme_minimal()
```

```
p2
```


Relation between x and y (multilevel analysis)
12.5
10.0
7.5
5.0
2.5
y
5
10
x
Cluster
Cluster_1
Cluster_2
Cluster_3
Cluster_4
Cluster_5
Cluster_6
Cluster_7
Cluster_8
Cluster_9

```r
(p1 + p2) +
  patchwork::plot_annotation(
    title = "Simpson's Paradox"
  ) & theme(plot.title = element_text(hjust = 0.5))
```



Any example of the above scenario in WOP?

For example $x =$ Training hours, $y =$ performance, j (cluster) $=$ company

Beyond the Simpson's paradox, other (less extreme) cases can be visualized here `http://mfviz.com/hierarchical-models/`

3. Key Concepts

3. Key Concepts

Understanding the *vocabulary of multilevel analysis* is the first step toward mastering this type of modeling.

In what follows, we will explore the most important concepts, terms, ambiguities, and characteristics of multilevel models.

3.1 One framework, a lot of names

Multilevel models (MLMs) are also referred to as **mixed-effects models**, **hierarchical linear models (HLMs)**, or **random coefficient models**.

Although the terminology varies slightly across disciplines, these terms usually refer to the same family of models. All of them are designed to account for **nested/clustered/hierarchical data structures**.

In typical applications, nesting occurs due to:[1]

- Clusters (e.g., employees within teams) or
- Repeated measures within individuals (in this case, the cluster is the ID of the participant) (Ployhart, Bliese, & Strizver, 2025)

[1] *Note:* In this course, we focus on two-level models for simplicity. Multilevel models can include more than two levels, but this is uncommon in practice due to the large sample size requirements.

Clustered (Cross-Sectional) Data: Employees Nested Within Teams

```r
library(dplyr)

set.seed(123)
n_teams <- 30
employees_per_team <- 10
total_employees <- n_teams * employees_per_team

# Between-level variable: Quality of leaderships (scale 1-10)
team_df <- data.frame(
  team_id = 1:n_teams,
  team_leadership = runif(n_teams, min = 1, max = 10)
)

# Simulate dataset
df_cluster <- data.frame(
  employee_id = 1:total_employees,
  team_id = rep(1:n_teams, each = employees_per_team)
) %>%
  left_join(team_df, by = "team_id") %>%
  mutate(
    performance = 40 + 3 * team_leadership + rnorm(n(), mean = 0, sd = 5)
  )

# Extract dataset
# writexl::write_xlsx(df_cluster, "df_cluster.xlsx")
```

Intensive Longitudinal Data: Repeated Measures Within Individuals

```
library(dplyr)

set.seed(456)
n_subjects <- 50
n_timepoints <- 9

trait_consc <- rnorm(n_subjects, mean = 5, sd = 1)  # between-person trait
subject_intercepts <- rnorm(n_subjects, mean = 50, sd = 4)  # baseline

# Simulate dataset
df_IntensiveLongitudinal <- data.frame(
  subject_id = rep(1:n_subjects, each = n_timepoints),
  time = rep(0:(n_timepoints - 1), times = n_subjects)
) %>%
  mutate(
    coscientiousn = trait_consc[subject_id],
    intercept = subject_intercepts[subject_id],
    Perform = intercept + 1.5 * time +
      2 * coscientiousn +
      rnorm(n_subjects * n_timepoints, sd = 3)
  ) %>%
  select(-intercept) %>%
  round(., 2)

# Extract dataset
# writexl::write_xlsx(df_IntensiveLongitudinal, "df_IntensiveLongitudinal.xlsx")
```

3.2 Between-level vs within-level

Within-level (variance, variables, effects, etc) refers to units, or to "Level-1".

Between-level (variance, variables, effects, etc) refers to clusters, or to "Level-2".

Note: This distinction is largely used in the Multilevel SEM framework (e.g., Heck & Reid, 2023).

3.3 Fixed Effects, Random Effects, and Random Parameters

There is often confusion between these terms (see: https://statmodeling.stat.columbia.edu/2005/01/25/why_i_dont_use/).

The terms **fixed effects** and **random effects** are commonly used in the context of *multilevel regression* (for this reason also called *mixed-effects models*), as well as in *meta-analysis*.

In contrast, the term **random parameters** (random intercept/s and random slope/s) is more frequently used in the context of *Multilevel SEM)*.

Using Heck and Thomas (2015, pp. 420 and 426) glossary:

Fixed Effects [...] In mixed effects and multilevel modeling, **fixed effects** often refer to estimates that are defined as nonvarying across higher-order units. The effects are seen as fixed to be the same within the entire sample of individuals.

In contrast, a **random effect** is assumed to vary across the higher-level units. For example, we might see the mean of salary as varying across different organizations.

In a multilevel model, the random effects are therefore the variance and covariance parameters at the group level, which may be seen to have varied effects across the sample of groups.

Using Heck and Thomas (2015, pp. 420 and 426) glossary:

Random effects When some effect in a statistical model is modeled as being random, we mean that we wish to draw conclusions about the population from which the observed units were drawn, rather than about these particular units themselves. Random effects modeling puts a focus on the variance of an effect across the population from which the units were sampled, rather than assuming the effect being fixed to one value in the population.

Using Heck and Thomas (2015, pp. 420 and 426) glossary:

- **Random intercept** This is a model where the level of the outcome mean is allowed to vary across the units in the sample. Differences in the levels of the means may be explained by unit-level predictors.
- **Random slope** This is a model where the size of the level-1 slope summarizing the regression of Y on X is allowed to vary across units in the sample. Differences in the strength of the level-1 slopes may be explained by unit-level predictors.

3.4 Equations

In simple linear regression analysis, parameters are fixed and variables vary only across units i.
In multilevel modeling, however, we introduce **clustering**, usually indexed by the subscript j.

Level-1 (within-level) predictors are typically denoted by x, while Level-2 (between-level) predictors are denoted by w.

Drawing on common notation (e.g., McNeish, 2017a), we can represent a two-level model with:

- one Level-1 predictor
- one Level-2 predictor
- random intercept and random slope

Level-1 equation: $y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + r_{ij}$

Level-2 equations:

- $\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + \mu_{0j}$
- $\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + \mu_{1j}$

The combined model will result:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + \gamma_{11}x_{ij}w_j + \mu_{0j} + \mu_{1j}x_{ij} + r_{ij}$$

This equation includes:

- fixed effects: γ's
- cross-level effect (on the intercept): $\gamma_{01}w_j$
- cross-level interaction (slope moderation): $\gamma_{11}x_{ij}w_j$
- random effects:

$$\begin{bmatrix}\mu_{0j}\\ \mu_{1j}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\tau_{00} & \tau_{01}\\ \tau_{01} & \tau_{11}\end{bmatrix}\right)$$

- residual error: $r_{ij} \sim \mathcal{N}(0, \sigma^2)$

The change in notation from simple Structural Equation Modeling (SEM) to **Multilevel SEM (MSEM)** is quite similar.

However, it places greater emphasis on the orthogonal decomposition of the total variance into the between-cluster (B) and within-cluster (W) components.

The total variance-covariance matrix is:
$\mathrm{Cov}(\mathbf{y}_{ij}) = \Sigma_T = \Sigma_W + \Sigma_B$

A multilevel factor model (multilevel CFA) is typically expressed as:
$\mathbf{y}_{ij} = \gamma + \Lambda_W \eta_{Wij} + \Lambda_B \eta_{Bj} + \varepsilon_{ij}$

Structural relations at the within level:
$\eta_{ij} = \alpha_j + \mathbf{B}_j \eta_{ij} + \Gamma_j \mathbf{x}_{ij} + \zeta_{ij}$

Structural relations at the between level:
$\eta_j = \mu + \mathbf{B} \eta_j + \Gamma \mathbf{x}_j + \zeta_j$

These equations correspond to Eq. 26.1, 26.5, 26.9, and 26.10 in Heck and Reid (2023).

3.5 Centering in multilevel models

Centering is the process of transforming a variable so that a value of zero has a meaningful interpretation: its mean.

This is typically done by subtracting a mean, either the overall (grand) mean or the cluster-specific (group) mean, from each observation.

It's a long standing debate in common multilevel observed-variable approach (Enders & Tofighi, 2007; Hamaker & Grasman, 2015; Hamaker & Muthen, 2020), as well as in multilevel latent-variable approaches (McNeish & Hamaker, 2020).

- **Group-mean centering** (also called *Centering Within Cluster*, **CWC**) is used to isolate within-level effects: $x_{ij} - \bar{x}_j$ where x_{ij} is a Level-1 variable and $\bar{x}_j$ is the group mean (e.g., average stress in a class).
- **Grand-mean centering (GMC)** is used in multilevel regression when the goal is to retain both within- and between-level variance: $x_{ij} - \bar{x}$, where x_{ij} is a Level-1 variable (e.g., student stress) and $\bar{x}$ is the grand mean. Note: GMC does not separate within- and between-level effects. If decomposition is needed, use CWC(M) or GMC(M) instead, where "(M)" refers to adding the group mean term $\beta_p \bar{x}_j$ to the model.
- **Latent-mean centering**, as implemented in M*plus*, performs centering at the model level. This avoids manual transformations and preserves the multilevel structure in latent variable models. It is particularly useful in Multilevel SEM, such as DSEM.

To make the best centering choice, **carefully read** González-Romá and Hernández (2023, p. 631, *Centering L1 predictors*) and Hamaker and Muthén (2020).

Centering in multilevel autoregressive models

When dealing with multilevel autoregressive models, carefully read Hamaker and Grasman (2015). Indeed, in this case the autoregressive $t-1$ variable (e.g., x_{t-1} predicting x_t) should **not** be group-mean centered, as this may introduce the *Nickell bias*: By removing the person mean, you risk eliminating meaningful information and biasing the autoregressive estimate downward.

Other predictors can be centered as usual (see previous slide). Instead, in the context of DSEM M*plus* automatically apply the latent centering approach (see McNeish & Hamaker, 2020).

ID i	cluster j	stress_raw x_{ij}	stress_Grand_Mean $\bar{x}$	stress_Group_Mean $\bar{x}_j$	stress_GMC $x_{ij} - \bar{x}$	stress_CWC $x_{ij} - \bar{x}_j$	performance (outcome) y_{ij}
1	A	10.99	15.02	10.92	-4.03	0.08	13.20
2	A	9.72	15.02	10.92	-5.30	-1.19	18.65
3	A	11.30	15.02	10.92	-3.73	0.38	16.65
4	A	13.05	15.02	10.92	-1.97	2.13	11.9
5	A	9.53	15.02	10.92	-5.49	-1.39	14.65
6	B	14.53	15.02	15.87	-0.49	-1.34	19.78
7	B	18.16	15.02	15.87	3.14	2.28	8.99
8	B	16.53	15.02	15.87	1.51	0.66	14.26
9	B	14.06	15.02	15.87	-0.96	-1.81	18.85
10	B	16.09	15.02	15.87	1.06	0.21	13.19
11	C	19.07	15.02	18.27	4.05	0.80	13.34
12	C	19.07	15.02	18.27	4.05	0.80	14.83
13	C	20.48	15.02	18.27	5.46	2.21	18.72
14	C	16.17	15.02	18.27	1.15	-2.10	13.76
15	C	16.55	15.02	18.27	1.53	-1.72	14.22

⚠ Simulated dataset created with ChatGPT; intended only to show Excel functions for centering (see `multilevelCenteringDataset.xlsx`)

Centering in M*plus* and R

M*plus* grand-mean centering → `DEFINE: CENTER x (GRANDMEAN);`
M*plus* group-mean centering → `DEFINE: CENTER x (GROUPMEAN);`
R (`tidyverse`) grand-mean centering (GMC; $x_{ij} - \bar{x}$):

```
df_cluster <- df_cluster %>%
  mutate(
    performance_GMC = performance - mean(performance, na.rm = TRUE)
    )
```

R (`tidyverse`) group-mean centering (or Centering Within Cluster, CWC; $x_{ij} - \bar{x}_j$)

```
df_cluster <- df_cluster %>%
  group_by(team_id) %>%
  mutate(
    performance_CWC = performance - mean(performance, na.rm = TRUE)
  ) %>%
  ungroup()
```

R (`tidyverse`): calculate the group mean (i.e., average performance at team level $\bar{x}_j$) to include as a Level-2 predictor in CWC(M) or GMC(M)

```
df_cluster <- df_cluster %>%
  group_by(team_id) %>%
  mutate(
    performance_mean_team = mean(performance, na.rm = TRUE)
  ) %>%
  ungroup()
```

3.6 Reliability in Multilevel Models

In multilevel models, reliability of multiple-indicator tools cannot be assessed in the same way as in traditional cross-sectional designs.

Moreover, it is important to distinguish between:

- Clustered designs (e.g., individuals within teams)
- Intensive Longitudinal Data (ILD) designs (e.g., repeated measures within individuals)

The topic is broad, as multiple reliability indices have been proposed (see Geldhof et al., 2014; Revelle & Wilt, 2019; Shrout & Lane, 2012).
A potential suggestion could be the follow:

Design	Within	Between
Clustered	ω_{within}	ω_{between}
ILD	R_c	R_{kF}

```r
library(multilevelTools)

data(aces_daily, package = "JWileymisc")
omegaSEM(
  items = c("COPEPrb", "COPEPrc", "COPEExp"),
  id = "UserID",
  data = aces_daily,
  savemodel = FALSE)
```

```
$Results
           label   est ci.lower ci.upper
25  omega_within 0.719    0.697    0.740
28 omega_between 0.908    0.884    0.933
```

```r
# Create a small Intensive Longitudinal Dataset
# (see `?psych::multilevel.reliability`).

shrout <- structure(
  list(
    Person = c(
      1L, 2L, 3L, 4L, 5L, 1L, 2L, 3L, 4L,
      5L, 1L, 2L, 3L, 4L, 5L, 1L, 2L, 3L, 4L, 5L),
    Time = c(1L, 1L,
             1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 3L, 3L, 3L, 3L, 3L, 4L, 4L, 4L,
             4L, 4L),
    Item1 = c(2L, 3L, 6L, 3L, 7L, 3L, 5L, 6L, 3L, 8L, 4L,
              4L, 7L, 5L, 6L, 1L, 5L, 8L, 8L, 6L),
    Item2 = c(3L, 4L, 6L, 4L,
              8L, 3L, 7L, 7L, 5L, 8L, 2L, 6L, 8L, 6L, 7L, 3L, 9L, 9L, 7L, 8L),
    Item3 = c(6L, 4L, 5L, 3L, 7L, 4L, 7L, 8L, 9L, 9L, 5L, 7L,
              9L, 7L, 8L, 4L, 7L, 9L, 9L, 6L)),
  .Names = c("Person", "Time", "Item1", "Item2", "Item3"),
  class = "data.frame", row.names = c(NA, -20L)
)
```

```
library(psych)

multilevel.reliability(
  shrout,
  grp="Person",
  Time="Time",
  items=c("Item1", "Item2", "Item3")
)
```

```
Multilevel Generalizability analysis
Call: multilevel.reliability(x = shrout, grp = "Person", Time = "Time",
    items = c("Item1", "Item2", "Item3"))

The data had  5  observations taken over  4  time intervals for  3 items.

 Alternative estimates of reliability based upon Generalizability theory

RkF  =  0.97 Reliability of average of all ratings across all items and  times (Fixed time effects)
R1R  =  0.6 Generalizability of a single time point across all items (Random time effects)
RkR  =  0.85 Generalizability of average time points across all items (Random time effects)
Rc   =  0.74 Generalizability of change (fixed time points, fixed items)
RkRn =  0.85 Generalizability of between person differences averaged over time (time nested within people)
Rcn  =  0.65 Generalizability of within person variations averaged over items  (time nested within people)

 These reliabilities are derived from the components of variance estimated by ANOVA
             variance Percent
ID               2.34    0.44
Time             0.38    0.07
Items            0.61    0.11
ID x time        0.92    0.17
ID x items       0.12    0.02
time x items     0.05    0.01
Residual         0.96    0.18
Total            5.38    1.00

 The nested components of variance estimated from lme are:
```

3.7 IRR and IRA

In organizational and multilevel research, Inter-Rater Reliability (IRR) and Inter-Rater Agreement (IRA) are two conceptually distinct but complementary families of indices used to assess data aggregated at the group level.

According to LeBreton and Senter (2008, p. 816)

- **IRR** refers to the relative consistency in ratings provided by multiple judges of multiple targets. Estimates of IRR are used to address whether judges rank order targets in a manner that is relatively consistent with other judges. The concern here is not with the equivalence of scores but rather with the equivalence of relative rankings.
- **IRA** refers to the absolute consensus in scores furnished by multiple judges for one or more targets. Estimates of IRA are used to address whether scores furnished by judges are interchangeable or equivalent in terms of their absolute value.

3.7.1 IRR + IRA indices

Although ICC indices are often considered measures of inter-rater reliability (IRR), LeBreton and Senter (2008, p. 819, Table 1) classify them under the broader category of IRR + IRA indices, as they capture both relative consistency and absolute agreement among raters.

- ICC(1): Proportion of variance in x_{ij} attributable to group membership (j). It quantifies the degree of clustering. For example, ICC$(1) = 0.15$ indicates that 15% of the variance is due to differences between clusters. Values above 0.05–0.10 are often taken as a justification for using multilevel modeling.

$$\text{ICC}(1) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \sigma^2_{\text{within}}} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

- ICC(2): Reliability of **group means**, also known as ICC$(1, k)$ or ICC(k) (LeBreton et al., 2023, pp. 243–244). "Estimates of ICC(2) should be reported when unit-level means are computed to serve as aggregate-level variables" (LeBreton et al., 2023, p. 243). When cluster sizes vary (e.g., teams of 5 to 20), K is often set to the **median group size** (LeBreton et al., 2023, pp. 243–244). When K is low, it may differ from ICC(1). Good values are $> .70$.

$$\text{ICC}(2) = \frac{K \cdot \text{ICC}(1)}{1 + (K-1) \cdot \text{ICC}(1)} = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \frac{\sigma^2_{\text{within}}}{K}}$$

3.7.1 IRR + IRA indices

Although ICC indices are often considered measures of inter-rater reliability (IRR), LeBreton and Senter (2008, p. 819, Table 1) classify them under the broader category of IRR + IRA indices, as they capture both relative consistency and absolute agreement among raters.

- ICC(1): Proportion of variance in x_{ij} attributable to group membership (j). It quantifies the degree of clustering. For example, ICC$(1) = 0.15$ indicates that 15% of the variance is due to differences between clusters. Values above 0.05–0.10 are often taken as a justification for using multilevel modeling.

$$\text{ICC}(1) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \sigma^2_{\text{within}}} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

- ICC(2): Reliability of **group means**, also known as ICC$(1, k)$ or ICC(k) (LeBreton et al., 2023, pp. 243–244). "Estimates of ICC(2) should be reported when unit-level means are computed to serve as aggregate-level variables" (LeBreton et al., 2023, p. 243). When cluster sizes vary (e.g., teams of 5 to 20), K is often set to the **median group size** (LeBreton et al., 2023, pp. 243–244). When K is low, it may differ from ICC(1). Good values are $>$ .70.

$$\text{ICC}(2) = \frac{K \cdot \text{ICC}(1)}{1 + (K-1) \cdot \text{ICC}(1)} = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \frac{\sigma^2_{\text{within}}}{K}}$$

3.7.2 IRA indices

The most used IRA index is the r_{WG} (which represents a within-group agreement index). The multi-item version is the $r_{WG(J)}$ (Biemann, Cole, & Voelpel, 2012).

$$r_{WG} = 1 - \frac{S_x^2}{\sigma_E^2} = 1 - \frac{S^2_{\text{observed}}}{S^2_{\text{expected}}}$$

"S_x^2 is the observed variance on the variable X (e.g., leader trust and support) taken over K different judges or raters and σ_E^2 is the variance expected when there is a complete lack of agreement among the judges. This is the variance obtained from a theoretical null distribution representing a complete lack of agreement among judges." (LeBreton & Senter, 2008, p. 218).

Table 1 Examples of constructs, measures, and circumstances that warrant different levels of agreement

Level of agreement	Interpretation	Illustrative examples	Explanation
.00 to .30	Lack of agreement	N/A	If calculating interrater agreement, there are few (if any) instances where you would reasonably expect and accept a lack of agreement between raters
.31 to .50	Weak agreement	Climate strength	Ideally, we would expect a group to have some degree of a climate. However, climate does not have to be strong. Weak climates exist when variability exists in the way that group members perceive the climate, so strong agreement may not be necessary
.51 to .70	Moderate agreement	Group cohesion captured using a relatively new measure	We might expect some degree of agreement for a construct like group cohesion, but if group cohesion is assessed using a newly designed measure that has not been subjected to substantial psychometric evaluation, we might not expect strong agreement
.71 to .90	Strong agreement	Group cohesion captured using a well-established, validated measure	Again, we might expect some degree of agreement for a construct like group cohesion. If group cohesion is measured using a well-established and validated measure, we might expect stronger agreement
.91 to 1.00	Very strong agreement	Panel interview ratings for critical decisions (e.g., decisions about hiring, promotion, firing, tenure)	Agreement between raters on constructs used to make important decisions should ideally be very strong

LeBreton et al. (2023, p. 248)

3.7.3 IRR and IRA in R and M*plus*

- The `multilevel` package provides functions for computing `ICC1`, `ICC2`, `rwg`, `rwg.j`, and other commonly used IRR/IRA indices.
- The `performance` package includes a general-purpose `icc()` function, compatible with mixed models fitted via `lme4`.
- The `irr` package offers several coefficients for inter-rater reliability and agreement (e.g., Cohen's κ, Kendall's W).
- M*plus* automatically computes ICCs when using `TYPE = TWOLEVEL;` in model estimation.

3.8 Model comparison and selection

As in many statistical analyses, the goal is often to compare several competing models.

In mixed-effects models, for example, we can compare models predicting the same dependent variable (i.e., y_{ij}) by testing increasingly complex models, progressively adding fixed and random effects.

A potential pipeline is the following:

Model	Question	Mixed-Effects Equation
1. Null (intercept only)	Is there substantial between-level variability to be explained in my outcome variable?	$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
2. Random intercept - Fixed slope	Does the intercept of my specified regression equation vary across clusters?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + r_{ij}$
3. Random intercept - Random slope	Does the slope of my specified regression equation vary across clusters?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + u_{0j} + r_{ij}$
4. Intercept as outcome	Are there between-level variables that can predict variability in intercept?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + u_{1j}x_{ij} + u_{0j} + r_{ij}$
5. Intercept and slope as outcomes	Are there between-level variables that can predict variability in intercept and slope?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + \gamma_{11}x_{ij}w_j + u_{1j}x_{ij} + u_{0j} + r_{ij}$

A potential pipeline is the following:

Model	Question	Mixed-Effects Equation
1. Null (intercept only)	Is there substantial between-level variability to be explained in my outcome variable?	$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
2. Random intercept - Fixed slope	Does the intercept of my specified regression equation vary across clusters?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + r_{ij}$
3. Random intercept - Random slope	Does the slope of my specified regression equation vary across clusters?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + u_{0j} + r_{ij}$
4. Intercept as outcome	Are there between-level variables that can predict variability in intercept?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + u_{1j}x_{ij} + u_{0j} + r_{ij}$
5. Intercept and slope as outcomes	Are there between-level variables that can predict variability in intercept and slope?	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}w_j + \gamma_{11}x_{ij}w_j + u_{1j}x_{ij} + u_{0j} + r_{ij}$

Model comparison can be done through:

- Likelihood ratio test (χ^2 comparison) (`lmtest::lrtest` in `R` and Satorra-Bentler $\Delta\chi^2$ for MSEM in M*plus*), where a significant *p*-value indicates that the more complex model provides a significantly better fit.
- Information criteria such as AIC and BIC, where lower values indicate better-fitting models.

3.9 Explained Variance (R^2)

In multilevel models, variance is partitioned across levels. Thus, the model may explain some variance at each level, not a single global R^2.

Nakagawa and Schielzeth (2013) proposed the

- Marginal R^2: variance explained by fixed effects
- Conditional R^2: variance explained by fixed + random effects

In `R` $\rightarrow$ `performance::r2` or `MuMIn::r.squaredGLMM`

4. Advanced topics

4. Advanced topics

Multilevel modeling includes additional methodological issues that deserve attention.

In what follows, there are selected concepts, each briefly introduced and referenced for further reading.

Comprehensive Resources

A comprehensive review of multilevel modeling topics (including those not covered here, such as statistical power and handling missing data) can be found in González-Romá and Hernández (2023) and LeBreton et al. (2023).

For detailed textbook treatments, see Raudenbush and Bryk (2002), Snijders and Bosker (2012), Hox, Moerbeek, and van de Schoot (2017).

For Multilevel SEM, see Heck and Thomas (2015), Heck and Reid (2023), Sadikaj et al. (2021).

For a parallel overview between latent growth modeling and multilevel modeling, see Grimm et al. (2017).

4.1 Cross-level Isomorphism

Cross-level isomorphism refers to the assumption that the measurement model (e.g., factor structure and loadings) is equivalent at both the within- and between-levels.

This is analogous to **measurement invariance** in longitudinal or multi-group settings.

Table 1. Analogy Between Between-Groups Measurement Equivalence and Cross-Level Isomorphism.

Between-Groups Measurement Equivalence	Cross-Level Isomorphism
Configural invariance: Pattern of zero and nonzero factor loadings holds between groups	Weak configural isomorphism: Same number of dimensions holds between levels Dimensions are generally shown to be indexed by similar indicators across levels without fixing loading patterns Strong configural isomorphism: Same number of dimensions and the pattern of zero and nonzero factor loadings holds between levels
Metric invariance: Factor loadings are equivalent between groups	Weak metric isomorphism: Relative ordering of factor loadings/item discriminations holds between levels (evidenced by high congruence of the loadings between levels) Strong metric isomorphism: Magnitude of factor loadings/item discriminations holds between levels
Scalar invariance: Indicator thresholds are equivalent between groups	No current models for estimating item thresholds across levels
Invariance in uniqueness	No statistical basis for testing across levels

Suggested technical reading: Tay, Woo, & Vermunt (2014)
Suggested theoretical reading: Aguinis, Beltran, & Marshall (2024)

4.2 Restricted Maximum Likelihood (REML)

- REML is an estimation method used to get more accurate estimates of variance components in multilevel models when the number of clusters is small.
- ML estimates fixed effects and variances at the same time, without adjusting for the fact that some variability has already been explained by the fixed effects.
- As a result, ML tends to underestimate the variance of random effects (like random intercepts and slopes).
- In contrast, REML first removes the influence of fixed effects, and then estimates the variance components using only the residual variation. This leads to less biased and more reliable estimates of random-effect variances.

Suggested reading: McNeish (2017a)

4.3 Multilevel Mediation

- Multilevel mediation extends classical mediation models to hierarchical data structures.
- Several configurations exist, such as 1-1-1 (all variables at the within-level), 2-1-1 (predictor at the between-level, mediator and outcome at the within-level), 2-2-1 (both predictor and mediator at the between-level).
- McNeish (2017b, Table 1) showed that 1-1-1 and 2-1-1 were the most used designs.
- Unlike classical mediation, multilevel mediation often involves **random slopes** and the **covariances between them**, making the indirect effect more complex than just $a \times b$ https://quantpsy.org/medmc/medmc111.htm
- For examples and M*plus* code, see: https://www.statmodel.com/download/Preacher.pdf

Suggested readings: McNeish (2017b); Preacher et al. (2010, 2011, 2016)

4.4 Standard Error Correction When Clustering is negligible or not substantive

- In some cases (e.g., Marsh et al., 2019; Perinelli, Pisanu, Checchi, Scalas, & Fraccaroli, 2022), clustering is not of substantive interest or has a negligible impact (e.g., low ICC values).
- In such situations, it is acceptable to run a single-level SEM model while adjusting standard errors to account for clustering.
- In M*plus*, use the option `TYPE = COMPLEX;` together with the MLR estimator, and specify the clustering variable via `CLUSTER = MY_CLUSTER;`.
- In `R` (with `lavaan`), use the argument `cluster = "MY_CLUSTER"` and the robust estimator `estimator = "MLR"` inside the `sem()` function.

Suggested reading: McNeish, Stapleton, & Silverman (2017)

4.5 Location-Scale Models

- Location-scale models are more commonly used in economics, but have recently been introduced in WOP.
- They allow the modeling of both the **mean (location)** and the **variance (scale)** of an outcome variable as a function of predictors.
- This means that in addition to explaining the average level of a variable (e.g., job satisfaction), we can also explain **individual differences in variability** (e.g., fluctuations in satisfaction across time or contexts).
- Adding the j subscript to the residual variance

$$r_{ij} \sim \mathcal{N}(0, \sigma_j^2)$$
$$\sigma_j^2 = \exp(\omega_0 + \mu_{2j})$$

- Such models are especially useful in ILD or EMA data, where variability in affect or behavior is a meaningful outcome.

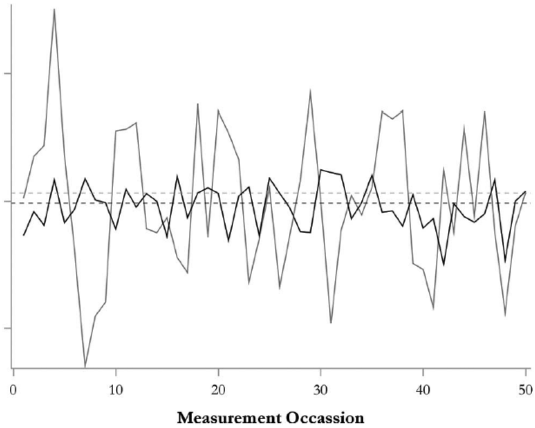


Figure 6. Comparison of trace plots for Person 5 (gray) and Person 96 (black) to highlight differences in variability across people when $N > 1$. Means for each person are shown as dashed lines.

Suggested readings: Lester et al. (2021); McNeish (2021); McNeish & Hamaker (2020)

4.6 Descriptive Statistics, Correlations, Preprocessing

When working with multilevel data, both **descriptive statistics** and **correlations** should be computed and reported separately at the **within-person** and **between-person** levels.

In addition, it is recommended to report:

- Proper operationalization of **time** (for ILD), such as fixing a meaningful zero point (usually the first measurement occasion) and using an appropriate metric (e.g., transforming dates into numeric values representing days or weeks).
- the number of **observations/measurements** within each cluster,
- adding the necessary centered variables,
- and, for Level-1 predictors, the average values within clusters.

Examples (see Supplementary Materials of these articles for the relevant code):

- Avanzi, Perinelli, & Mariani (2023): Application to cross-sectional clustered data.
- Menghini, Perinelli, & Balducci (2025): Intensive Longitudinal Design (ILD) example, including a complete preprocessing pipeline with both *required* and *recommended* steps.
- Perinelli, Vignoli et al. (2023): Preprocessing ILD for DSEM applications.

5. Data

5. Data

Well, now that you've learned a lot about multilevel modeling, it's time to apply that knowledge to data and software! 😊

We have 4 options:

- Explore and work with this R-based pipeline: https://www.rensvandeschoot.com/tutorials/lme4/.
- Explore and work with M*plus* files provided by Geiser (2013, Chapter 5): https://www.guilford.com/companion-site/Data-Analysis-with-Mplus/9781462502455.
- Explore and work with the M*plus* examples and visual guides provided in the folder `Example_Mplus`.
- Work with your own data — for example, your dataset, a paper in progress, or an article you've read and want to discuss.

6. References

6. References

- Aguinis, H., Beltran, J. R., & Marshall, J. D. (2024). Performance: Confirming, refining, and refuting theories. *Journal of Management Scientific Reports, 2*(2), 135-153. https://doi.org/10.1177/27550311241247487
- Avanzi, L., Perinelli, E., & Mariani, M. G. (2023). The effect of individual, group, and shared organizational identification on job satisfaction and collective actual turnover. *European Journal of Social Psychology, 53*(5), 956-969. https://doi.org/10.1002/ejsp.2946
- Biemann, T., Cole, M. S., & Voelpel, S. (2012). Within-group agreement: On the use (and misuse) of r_{WG} and $r_{WG(J)}$ in leadership research and some best practice guidelines. *The Leadership Quarterly, 23*(1), 66-80. https://doi.org/10.1016/j.leaqua.2011.11.006
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121
- Geiser, C. (2013). *Data analysis with Mplus*. The Guilford Press.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. https://doi.org/10.1037/a0032138
- González-Romá, V., & Hernández, A. (2017). Multilevel modeling: Research-based lessons for substantive researchers. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 183-210. https://doi.org/10.1146/annurev-orgpsych-041015-062407
- González-Romá, V., & Hernández, A. (2023). Conducting and evaluating multilevel studies: Recommendations, resources, and a checklist. *Organizational Research Methods, 26*(4), 629-654. https://doi.org/10.1177/10944281211060712

- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth Modeling: Structural equation and multilevel modeling approaches*. Guilford Press.
- Hamaker, E. L., & Grasman, R. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology, 5*, Article 1492. https://doi.org/10.3389/fpsyg.2014.01492
- Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365–379. https://doi.org/10.1037/met0000239
- Heck, R., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (3rd Ed.). Routledge.
- Heck, R. H., & Reid, T. (2023). Multilevel structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 481–499). The Guilford Press.
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.
- LeBreton, J. M., Moeller, A. N., & Wittmer, J. L. (2023). Data aggregation in multilevel research: Best practice recommendations and tools for moving forward. *Journal of Business and Psychology, 38*(2), 239-258. https://doi.org/10.1007/s10869-022-09853-9

- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815-852. https://doi.org/10.1177/1094428106296642
- Lester, H. F., Cullen-Lester, K. L., & Walters, R. W. (2021). From nuisance to novel research questions: Using multilevel models to predict heterogeneous variances. *Organizational Research Methods, 24*(2), 342-388. https://doi.org/10.1177/1094428119887434
- Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331–353. https://doi.org/10.1037/edu0000281
- McNeish, D. (2017a). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research, 52*(5), 661-670. https://doi.org/10.1080/00273171.2017.1344538
- McNeish, D. (2017b). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 609-625. https://doi.org/10.1080/10705511.2017.1280797
- McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods, 25*(5), 610–635. https://doi.org/10.1037/met0000250
- McNeish, D. (2021). Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organizational Research Methods, 24*(3), 630-653. https://doi.org/10.1177/1094428120913083

- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. https://doi.org/10.1037/met0000078
- Menghini, L., Perinelli, E., & Balducci, C. (2025). Manipulation of Intensive Longitudinal Data: A Tutorial in R With Applications on the Job Demand-Control Model. *International Journal of Psychology, 60*(2), Article e70040. https://doi.org/10.1002/ijop.70040
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in ecology and evolution, 4*(2), 133-142. https://doi.org/10.1111/j.2041-210x.2012.00261.x
- Perinelli, E., Pisanu, F., Checchi, D., Scalas, L. F., & Fraccaroli, F. (2022). Academic self-concept change in junior high school students and relationships with academic achievement. *Contemporary Educational Psychology*, *69*, Article 102071. https://doi.org/10.1016/j.cedpsych.2022.102071
- Perinelli, E., Vignoli, M., Kröner, F., Müller, A., Genrich, M., & Fraccaroli, F. (2023). Workers' emotional exhaustion and mental well-being over the COVID-19 pandemic: A Dynamic Structural Equation Modeling (DSEM) approach. *Frontiers in Psychology*, *14*, Article 1222845. https://doi.org/10.3389/fpsyg.2023.1222845
- Ployhart, R. E., Bliese, P. D., & Strizver, S. D. (2025). Intensive Longitudinal Models. *Annual Review of Organizational Psychology and Organizational Behavior, 12*, 343-367. https://doi.org/10.1146/annurev-orgpsych-110622-054803
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*(2), 161-182. https://doi.org/10.1080/10705511.2011.557329

- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods, 21*(2), 189–205. https://doi.org/10.1037/met0000052
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*(3), 209–233. https://doi.org/10.1037/a0020141
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.
- Revelle, W., & Wilt, J. (2019). Analyzing dynamic data: A tutorial. *Personality and Individual Differences, 136*, 38-51. https://doi.org/10.1016/j.paid.2017.08.020
- Sadikaj, G., Wright, A. G. C., Dunkley, D. M., Zuroff, D. C., & Moskowitz, D. S. (2021). Multilevel structural equation modeling for intensive longitudinal data: A practical guide for personality researchers. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp. 855–885). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00033-9
- Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). The Guilford Press.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological), 13*(2), 238-241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x
- Snijders, T. A., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
- Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods, 17*(1), 77-106. https://doi.org/10.1177/1094428113517008

Thanks for your attention

enrico.perinelli@unitn.it

UNIVERSITÀ DI TRENTO
Dipartimento di Psicologia e Scienze Cognitive

Presentation and material available at
`https://github.com/EnricoPerinelli/Intro-Multilevel`