4th ECPR Winter School in Methods and Techniques, 13-20 February 2015
University of Bamberg, Germany
Course Description Form¹ [1-week main course, 15 hours]

Course title
WD218 Structural Equation Modeling (SEM) with R

Instructor details
First name, last name: Ulrich Schroeders
Department/Unit: Department of Educational Science
Institution: University of Bamberg
Full postal address for ECPR correspondence: Feldkirchenstr. 21, 96045 Bamberg, Germany
Phone: +49 (0)951 863-2467
Fax: -
E-mail: ulrich.schroeders@uni-bamberg.de

Short Bio
http://www.uni-bamberg.de/ebf-jp

Dr. Ulrich Schroeders is Assistant Professor of Educational Research at the Bamberg Graduate School of Social Sciences (BAGSS) at the University of Bamberg. His focus is on psychological assessment, especially technology-based assessment, intelligence research, and competence testing in educational large-scale assessment.

Prerequisite knowledge
Note from the Academic Convenors to prospective participants: by registering to this course, you certify that you possess the prerequisite knowledge that is requested to be able to follow this course. The instructor will not teach again these prerequisite items. If you doubt whether you possess that knowledge to a sufficient extent, we suggest you contact the instructor before you proceed to your registration.

[For an introductory course: do indicate if no specific prerequisite knowledge is requested. Otherwise, in short, list the core prerequisite items (conceptual, technical, software, ...). Not much detail is needed.]

[For an advanced course: indicate in detail the prerequisite knowledge items. Where applicable, go to the specifics (for instance: “be able to use [software X] to perform [procedure Y] on [data type Z]”).]

Prospective participants should possess a solid knowledge of basic statistics (e.g., variance, covariance, correlation) and of manifest regression analyses (e.g., prediction, standard error). They should also be familiar with the terms often used in test theory (e.g., items, scale, reliability, validity).

With regard to software, potential participants should have worked with R on a productive level. More specifically, participants should be able to read in data from different sources, to select and to reference specific parts of a data set, and to calculate basic descriptive statistics (e.g., mean, correlation, etc.). They should be familiar with R Studio or any another programming environment. In the course a combination of Notepad++ and NppToR will be used to write and execute R Syntax; but previous experience with this software solution is not necessary.

1 Disclaimer: the information contained in this course description form may be subject to subsequent adaptations (e.g. taking into account new developments in the field, specific participant demands, group size etc.). Registered participants will be informed in due time in case of adaptations.
Short course outline
This course deals with intermediate and advanced aspects of Structural Equation Modeling (SEM), a method that is popular in psychology, educational research, and the social sciences. The course begins with a comprehensive introduction to the theoretical aspects of SEM and the terminology. Among others, issues of model identification, handling of missing data and the appropriate use of different estimators are discussed and trained in practical exercises that are an essential part of each session. Further topics include aspects of reliability, model fit evaluation, testing of measurement invariance, and modeling of longitudinal data. The overall aim of the course is to acquire a deeper understanding of latent variable modeling and to develop skills in order to estimate, interpret and tweak SEMs. All analyses will be conducted with the R package lavaan.

Long course outline
This course deals with intermediate and advanced aspects of Structural Equation Modeling (SEM), a method that is becoming more and more popular in the behavioral and social sciences. The course comprises five three-hour sessions; the time for each session is divided equally into a theoretical and a practical part. Different examples from educational research, psychology, and the social sciences are utilized. The course starts with a conceptual overview of the different classes of structural equation models and introduces the basic terminology that is used throughout the course. Furthermore, key concepts such as model identification or the distinction between covariance and mean structure are covered. After this overview, the participants are introduced to the fundamentals and the logic of the R package lavaan that is subsequently used for SEM.

Participants learn to specify Confirmatory Factor Analyses (CFA) and interpret the lavaan output. In comparison to other latent variable approaches such as IRT, SEM has the advantage of providing good omnibus tests of model fit. Accordingly, participants calculate different fit statistics (e.g., CFI, RMSEA, SRMR), learn their adequate usage and their limitations. As empirical research is often accompanied by some limitations of the data, most prominently, missingness or skewed distributions of variables, course training also includes handling of missing data and usage of different estimators (e.g., for categorical/dichotomous and continuous variables). Strengths and weaknesses of parceling of item parceling is also addressed.

In many cases an initial, theory-driven CFA model will not fit the data perfectly. We will pay close attention to sources of model mis specification and ways how to improve model fit by data- drive adjustments. Participants will become acquainted with different modeling approaches (higher order models, nested factor models, CTCM, CTCU, etc.) and understand their conceptual differences. As SEM is a powerful method to test competing theories, we will also discuss how to conduct model comparisons of nested and non-nested models. Furthermore, structural regression models that postulate latent regressions between variables are discussed as well as models that include manifest covariates (MIMIC approach).

Testing for measurement invariance is a necessary prerequisite to making valid statements about group differences. For instance, in order to ascertain that boys outperform girls in mathematics it is necessary to demonstrate that the assumptions of strong measurement invariance holds (i.e., that we are measuring the same construct in both groups identically). A common, straightforward procedure of testing for measurement invariance with Multi-Group Confirmatory Factor Analysis (MGCFA) is applied. Moreover, we extend the procedure to ordinal data and also discuss partial measurement invariance.

Because of its flexibility SEM is often employed to analyze longitudinal data as well as cross-sectional data. In the last part of the course we expand our exercises to longitudinal data. One class of models deals with modeling variability of scores around time-stable traits; the other class of models we will get to know is concerned with change in latent variables over time (i.e., latent growth curve modeling). In the last session there will also be time to answer further questions concerning issues discussed in previous sessions.
### Day-to-day schedule (Monday 16 February to Friday 20 February)

3 contact hours, split in two 90’ sessions.
For each session approximately 90 minutes of lecture and 90 minutes of training.

<table>
<thead>
<tr>
<th>Topic(s)</th>
<th>Details [NB: incl. timing of lecture v/s lab or fieldwork etc. hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day 1</strong></td>
<td>Introduction to latent variable modeling and Confirmatory Factor Analysis (CFA)</td>
</tr>
</tbody>
</table>
|          | – Different types of SEMs  
|          | – Fundamentals of the *lavaan* syntax  
|          | – Specification of measurement models based on covariance matrix and raw data  
|          | – Model identification, calculation of the observed and estimated parameters  
|          | – Interpretation of the *lavaan* output  
|          | – Draw SEM diagrams |
| **Day 2** | Evaluating model fit, dealing with missingness, using different estimators and estimating reliability |
|          | – Evaluating model fit values ($\chi^2$, $df$, CFI, RMSEA, SRMR), their calculation and limitations  
|          | – Types of missingness (MAR, MCAR, etc.)  
|          | – Estimators for categorical/dichotomous and continuous variables  
|          | – Strengths and weaknesses of parceling  
|          | – Reliability coefficients (McDonald’s $\omega$, Cronbach’s $\alpha$) |
| **Day 3** | Improving models and advanced modeling techniques |
|          | – Equivalent and nested models  
|          | – Specification of different models, among others:  
|          |   • higher order models  
|          |   • correlated trait correlated uniqueness model  
|          |   • correlated trait correlated method models  
|          |   • nested factor models  
|          | – structural regression models  
|          | – models with covariates (MIMIC) |
| **Day 4** | Testing measurement invariance with Multi-Group Confirmatory Factor Analysis (MGCFA) |
|          | – Logic of measurement invariance (MI) testing  
|          | – Common MI testing procedure across groups (configural, metric, scalar measurement invariance)  
|          | – Partial measurement invariance  
|          | – MI testing with dichotomous/ordinal data |
| **Day 5** | Latent-State-Trait-Analysis and Latent-Change-Models |
|          | – Autoregressive models  
|          | – Latent state analysis  
|          | – Latent state trait analysis  
|          | – Latent change models  
|          | – Latent change curve models |

### Day-to-day reading list
2-3 additional daily hours of homework (readings, small projects, etc.)

<table>
<thead>
<tr>
<th>Readings (please list at least the compulsory reading for the scheduled day)</th>
</tr>
</thead>
</table>
| **Day 1**  

**Software and hardware requirements**

**Software programme**
1) most recent version of R (e.g., 3.0.3)
2) most recent version of the R package *lavaan* (can be retrieved from lavaan.ugent.be)
3) Additionally, the R packages *psych, foreign, MBESS, semTools, semPlot* (via CRAN)
4) *Notepad++* (e.g., v6.6.3, see http://notepad-plus-plus.org/)
5) *NppToR* (see http://sourceforge.net/projects/npptor/)

**Hardware requirements**
No special requirements

**Literature – Further Readings**

**Lecture room requirement**
Room with PCs/Laptops

**Preferred time slots**
afternoon
**Other recommended courses (before or after this course)**
The following other ECPR Methods School courses could be useful in combination with this one in a ‘training track’. NB this is an indicative list.

**Before this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Introduction to R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**After this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>