8th GESIS Summer School in Survey Methodology
Cologne, August 2019

Syllabus for Course 02: "Introduction to Structural Equation Modeling: Confirmatory Factor Analysis with Mplus"

Instructors:	Prof. Dr. Jost Reinecke	Georg Kessler
E-mail:	Jost.reinecke@uni-bielefeld.de	Georg.kessler@univie.ac.at
Homepage:	www.uni-bielefeld.de	https://ufind.univie.ac.at

Date: 05.-09. August 2019
Time: 9:00–13:00, 14:00-16:00
Course starts Monday morning at 9:00

About the Instructors:

Prof. Dr. Jost Reinecke is professor of quantitative methods of empirical social research at the Faculty of Sociology at the University of Bielefeld. His research focuses on the methodology and application of structural equation models and latent class analysis, both, cross-sectionally and longitudinally. His current methodological research focuses on growth curve and growth mixture models and the development of techniques related to multiple imputation of missing data in complex survey designs. His current substantive research focuses on the longitudinal development of adolescents' delinquent behavior and applications of bi-factor models with the strengths and difficulties Questionnaire (SDQ).

Georg Kessler is a researcher at the Institute for Methods of the Empirical Social Sciences at the University of Bielefeld and involved in the project "Crime in the Modern City" – a panel study on youth criminality. He worked as a lecturer at the Universities of Vienna and Würzburg, where he taught methodology of the social sciences in undergraduate programs. His research interests include delinquent behavior in the third decade of life, the assessment of measurement quality by combining quantitative and qualitative data, panel data analysis, and the role of values and norms in society. Furthermore, he worked as a consultant at consulting and human development firms in Vienna.

Selected Publications:

- J. Reinecke (2014), Strukturgleichungsmodelle in den Sozialwissenschaften. München: Oldenbourg Wissenschaftsverlag.
- D. Seddig & J. Reinecke (2017), Exploration and Explanation of Adolescent Self-reported Delinquency Trajectories in the Crimoc Study, in: V. van der Geest & A. Blokland (Eds.), The International Handbook of Life-Course Criminology. (pp. 159-178). Routledge.
- K. Kleinke & J. Reinecke (2013), Multiple imputation of incomplete zero-inflated count data, in: Statistica Neerlandica, 67 (3), 311-336.

Short Course Description:

The course focuses on measurement models and their application within the Structural Equation Modeling (SEM) framework. We will show how a theoretical model, represented by measurement models, can be applied to empirical data and how to assess its fit to the data through the measurements' covariance matrix. Confirmatory Factor Analysis (CFA) is an important and basic aspect of the SEM-framework and its understanding and application to data is the core learning aspect of this course. Also, CFA is a necessary conceptual precondition to understand and apply the structural aspect of SEM, path modeling. Therefore, the course deals with concepts and applications of CFA such as assessing construct validity and reliability of a measurement model as well as the interpretation of calculated results. The topics addressed in the course include different modeling techniques of CFA such as single measurement models, simultaneous CFA (SCFA), the Multiple Group Comparison of the CFA

(MGCFA), and the higher-order CFA. If time permits on the last day, we can peak into topics as CFA with categorical data, path-modeling, how to handle missing data, or longitudinal analysis. Throughout the course we will work on examples provided by the lecturers using the popular SEM software package Mplus. For data preparation we accommodate needs of SPSS- or Stata-users.

Keywords:

measurement, Mplus, confirmatory, SEM

Course Prerequisites:

- We strongly encourage participants to familiarize themselves with and have a conceptual/mathematical understanding of variance, covariance, correlation, standardization, hypothesis testing (t-test, chi-square), and regression analysis [for compact refreshing we recommend http://davidmlane.com/hyperstat/];
- basic knowledge of matrix notation [a short refresher can be found on https://www.youtube.com/watch?v=G16c2ZODcg8];
- knowledge of what a linear equation system looks like and how it can be solved;
- Handling of system files (.sps; .dta; …) and transformation to portable or ASCII-data files (.dat; .csv; .txt; …) [for SPSS users: a good preparation is to import .txt-files into SPSS and use SPSS-syntax to get data; for Stata users: a good preparation is to use the stata2mplus ado in Stata to get Mplus input and data file simultaneously].
- As introductory reading we also recommend studying the chapters 1 to 3 of the Brown book (cited in the course literature).
- Basic familiarity with Mplus (can be acquired in the short course "Using Mplus for Latent-Variable Modelling: An Introduction" in week 0) and familiarity with writing syntax (Mplus input - as taught in the class-is syntax only) [we recommend looking into chapter 5 of http://www.statmodel.com/ugexcerpts.shtml].

Target Group:

Participants will find the course useful if they:
(On the level of their research questions)

- work with models that involve a complex structure of variables involving latent concepts and their relationships to each other;
- have a strong deductive framework and want to verify theoretical assumptions derived from substantive theories;
- need information on measurement quality (validity and reliability testing)
- want to apply SEM to their future analysis.

(On a more basic level)

- want to get an introduction into Structural Equation Model (SEM)-framework;
- have had prior experience with SEM, but no formal training;
- they have had prior training, but still find the whole matter rather complicated;
- they want to further their understanding of Mplus.

While this course is introductory in nature, its theoretical input should be dense enough to help more advanced users to effectively brush up their knowledge.

Course and Learning Objectives:

By the end of the course participants will:

- know how to define a latent construct through a measurement model;
- comprehend the mathematical and statistical foundation of SEM;
- be able to read, understand, and interpret an Mplus output;
- transfer the theoretical knowledge to applied research projects;
- in general be enabled to acquire the set of skills they need for their individual projects.

Organizational Structure of the Course:

The course is structured around four hours of classroom instruction (09:00-13:00) and two hours of assisted learning practice in a PC lab (14:00-16:00). During the two hours of assisted learning practice in the afternoon, we expect the participants to apply the covered material in their assignments and/or their own projects. Instructors will be present during these sessions. The class will be split so that individual coaching and class assignments will be tutored simultaneously. All the participants will have an opportunity to consult the instructors individually within the mentioned 2 hours or by appointment. The room for individual consultations will be announced in class.

Teaching will take place as a combination of lectures on the theory of CFA, getting acquainted with the program Mplus, and application of the theory in practice. To facilitate learning, participants are encouraged to bring and discuss their own projects (e.g., research questions) to get feedback from their peers and the instructors. The debate will enhance the transfer of theoretical learning to applied knowledge. We therefore expect all of the participants to take part in these 2 hours outside of the lecture to contribute to and benefit from the exchange and the extra practice.

Software and Hardware Requirements:

None. For the duration of this course, GESIS will provide participants with access to the required statistical software packages.

Long Course Description:

Attitudes, norms, values, and other key concepts in the social sciences are often not only measured by one indicator alone, but are usually conceptualized as latent concepts measured by a scale comprised of several items. If we are interested in the analysis of the latent concept, information of all these indicators need to be weighted and aggregated. Commonly, this is achieved by constructing weighted composite scores. By using CFA we may reach beyond these traditional and limited methods: first of all, by using CFA we also correct for measurement error by assigning measurement weights to each indicator and separate explained from unexplained variance. In addition, we are very flexible when it comes to specifying a particular measurement theory, like to define the relation of measurements to their latent constructs, while also allowing for selected cross-loadings. Also, relationships between measurements or between several latent constructs can be estimated freely or restricted to be a predefined value – this option extends to all parameters of a model. Regression weights (and all other parameters) don't need to be exported and used in a second step (e.g. in a regression), but in SEM they are part of one model and thus are calculated simultaneously. And what about the latent constructs? In SEM they become estimated "objects" with means, intercepts, variance, and measurement error that can be used for further analysis like regular variables. To sum it up: whichever arbitrary, linear relation a researcher can think of for his measured or latent variables, he or she can model it within the SEM-framework. Furthermore, SEM programs provide a variety of indices that offer statistical evidence for the model's fit (or lack thereof) regardless whether we are simply analyzing the measurement model's quality or use the latent constructs for further (causal) analysis. In contrast to Exploratory Factor Analysis (EFA), confirmatory factor models allow to test theoretically driven factor structures of measurements under study. Restricting the number of factors as well as the relations between factors and measurements allows testing the quality of the measurement and different assumptions for these. This pertains to aspects like reliability, construct validity, equality of parameters but also whether for example the measurement is similar in different cultural settings. Neglecting these measurement attributes could lead to biased conclusions of classical statistical analysis (e.g. mean difference testing in cross-cultural surveys, regression with composite scores). A valid measurement model is the basis for any further analysis with latent constructs.

We begin by providing a general introduction to SEM and its background and foundations. To this end, we will make use of mathematical notations and discuss mathematical procedures that are necessary to get an essential understanding of the basic algorithms. You will not have to solve any mathematical problems, but your knowledge in statistics will help you to follow the lecture better. Make sure that you meet the requirements stated above. The lecture is cognitively demanding and you need to draw on previous knowledge and combine it with the contents presented.

We propose the following itinerary through our course (subject to change): From the general introduction into the SEM-framework we will make our way to CFA – meaning that we look at the relational structure behind the measurement of a latent construct. Different specifications of measurement models are tested with CFA to get a

grip on the program, the matrix notation and how these two interact. This knowledge is necessary to specify any model – regardless whether it concerns the measurement or structural aspects. That will lead us to various techniques of CFA: the simultaneous CFA (SCFA), the Multiple Group Comparison (MGCFA) and the higher-order CFA. Our experience has shown that Confirmatory Factor Analysis (CFA) as an important aspect of the SEM-framework seems to be a good starting point for an introduction. Mastering this will give you a good start into further exploring SEM on your own. An outlook on a structural application will round off the course. In the tutorials there will be (some) room to discuss your scientific projects and get feedback.

Techniques of model estimation will be shown using the program Mplus. Please note that this course focuses mainly on the application of these techniques with continuous variables. However, the basic concepts are also applicable to categorical data. Participants will apply CFA with different survey datasets. The usefulness of measurement models for measuring attitudinal and behavioral aspects will be addressed and different aspects of cross-sectional, cross-cultural, and longitudinal designs are emphasized.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	**General introduction to Structural Equation Modeling: general latent variable concept; decomposition of variances; path diagrams and mathematical notation; model estimation and evaluation; Classical Test Theory**
	Suggested reading: ▪ Brown chapter 1, 2 and 3
2	**Basics of CFA: the model and its parameters; assumptions; model fit.**
	Compulsory reading: ▪ Brown chapter 2 and 3 Suggested reading: ▪ Muthén & Muthén 2010, chapters 1 & 2, Muthén & Muthén 2010, example 5.1
3	**Model restrictions and equality constraints, model identification, Simultaneous Confirmatory Factor Analysis (SCFA), model modifications.**
	Compulsory reading: ▪ Brown chapters 4 and 5 Suggested reading: ▪ Muthén & Muthén 2010, example 13.9
4	**Types of Errors, Reliability and Validity Estimates in Confirmatory Factor Analysis (CFA), Multiple Group Confirmatory Factor Analysis (MGCFA).**
	Compulsory reading: ▪ Brown Chapter 7 Suggested reading: ▪ Muthén & Muthén 2010, example 13.10; Muthén & Muthén 2010, pp. 421-428.
5	**SCFA with intercepts and latent means, higher order CFA, MTMM.**
	Compulsory reading: ▪ Brown, Chapter 6, 7, 8 Suggested reading: ▪ Muthén & Muthén 2010, pp. 428-429.