Advances in the analysis of event-related potential data with factor analytic methods

Von der Fakultät für Lebenswissenschaften
der Universität Leipzig
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

Florian Scharf, M.Sc. Psychologie

geboren am 14.11.1990 in Grimma

Dekan:

Prof. Dr. Tilo Pompe

Gutachter:

Prof. Dr. Steffen Nestler

Prof. Dr. André Beauducel

Tag der Verteidigung: 29.03.2019

Acknowledgements

To begin, I want to thank a number of people who supported me throughout the work on this theses.

First and foremost, I want to express my gratitude to my supervisor Steffen Nestler. He let me pursue my own project ideas while teaching me the patience, the serenity, and the ability to reflect on my own work that are necessary to cope with the ups and downs of the scientific process – and academia in general. I am grateful for all the things that I could learn from him, and his guidance has shaped me in many ways.

A special thanks goes to Jana Pförtner for assisting me with supplementary simulations that helped me to better understand my simulation results – and for sharing my pleasure in programming.

I am indebted to Richard Rau for agreeing to proofread this thesis even when it turned out to be on short notice.

I would also like to thank Dagmar Müller for being an eternal source of encouragement and for paving more ways for me than I would ever have dared to ask for.

I am grateful to Andreas Widmann for introducing PCA of ERP data to me in the first place.

I also want to thank Maria Blöchl for numerous challenging and interesting discussions.

I thank Urte Roeber and Robert O'Shea for an interesting and joyful research stay in Perth.

I also want to thank all members of the Psychological Methods group for the greatest team spirit I could imagine and for bearing endless repetitions of basically the same talk.

Last but not least, I want to close by thanking Julia Heinz for her never-ending support and for constantly reminding me to take breaks from time to time.

This thesis is dedicated to Burghilde Marx.

Contents

Chapter 1

General introduction

1.1 Relevance of neurophysiological measures

The investigation of cognitive processes with psychological experiments has a long tradition in psychology (Fuchs & Milar, 2003). Through clever experimental design (e.g., Deary, Liewald, & Nissan, 2011) and modern modeling approaches (e.g., van der Linden, 2009), researchers have learned a lot about cognitive processes by studying overt behavior (i.e., reaction times or erroneous responses). Nevertheless, insights from behavioral studies are limited regarding mechanistic explanations because they do not allow direct access to the underlying brain processes. This has motivated the development of countless methods that enable the investigation of neural correlates (see Thompson & Zola, 2003, for a review). One of the oldest and most popular of these methods is the recording of brain activity from multiple electrodes that are placed on the participant's scalp via *electroencephalography* (EEG).

Technically, EEG mainly captures synchronous post-synaptic potentials (PSPs) from cortical pyramidal neurons (Jackson & Bolger, 2014). In order to be detectable from outside the skull, PSPs from thousands of neurons that have parallel orientation and are temporally synchronized need to sum up. On its way from the source to the electrodes outside the skull, the electric potential is affected by many influences (e.g., other electric potentials, transitions between tissues, external noise, etc.). Therefore, the EEG signal is best described as a mixture of spatiotemporally smoothed brain activity and electric activity from other sources (e.g., electric noise, changes in skin conductance, muscular activity). Consequently, only very strong brain activity is readily visible in raw EEG signals (e.g., epileptic activity Maganti & Rutecki, 2013). EEG has excellent temporal resolution because the electric potentials are transmitted through the head nearly instantaneously (Nunez & Srinivasan, 2006), but its spatial resolution is poor because the source activity in 3D space is projected on the 2D head surface.

Over the last decades, many research efforts were made in order to extract relevant and valid information from EEG signals for research and application purposes. The development of brain-computer interfaces (BCIs) is an outstanding example how useful EEG signals can be although they only provide partial access to the activity of the brain (see Fazel-Rezai et al., 2012, for a review). With such devices, it is possible to communicate with completely locked-in patients with whom communication through any other means is impossible (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). This example of research utilizing EEG signals underlines the potential and the relevance of the EEG method.

1.2 Event-related potentials

Cognitive psychologists are often interested in comparing brain activity between two (or more) specific points in time (e.g., the presentation of different stimuli or the same stimulus in different experimental conditions). For this purpose, the investigation of event-related potentials (ERPs) has received a lot of attention since the middle of the 20th century (Sur & Sinha, 2009). Typically, the broad term *event* is used to refer to the time point of interest indicating that a variety of features may define the relevant time points (e.g., stimulus on- and offsets, or the participant's response). Brain activity that occurs time-locked to an event is called *event-related* (e.g., Kappenman & Luck, 2012). Event-related activity is usually investigated in a specific time *epoch* around the event (e.g., 100 ms before and 500 ms after the event). The ERP can be defined as a time series of voltage deflections in such an epoch that is elicited by the event and recorded from the participant's scalp via EEG.

Figure 1 depicts a simplistic, simulated example of an ERP. The ERP is typically described by amplitude, polarity, latency, and topography of peaks in the voltage deflection (e.g., Luck, 2014). In the left-most panel of Figure 1, for instance, a *negative* peak with a latency of 120 ms (in short: N120) is followed by a *positive* peak with a latency of 300 ms (in short: P300). The term topography is used in reference to the distribution of the voltage across the electrode sites at a specific sampling point. In the middle panel in Figure 1, one can see the ERP at another electrode site. Here, the polarity of the first peak is reversed and the amplitude of the

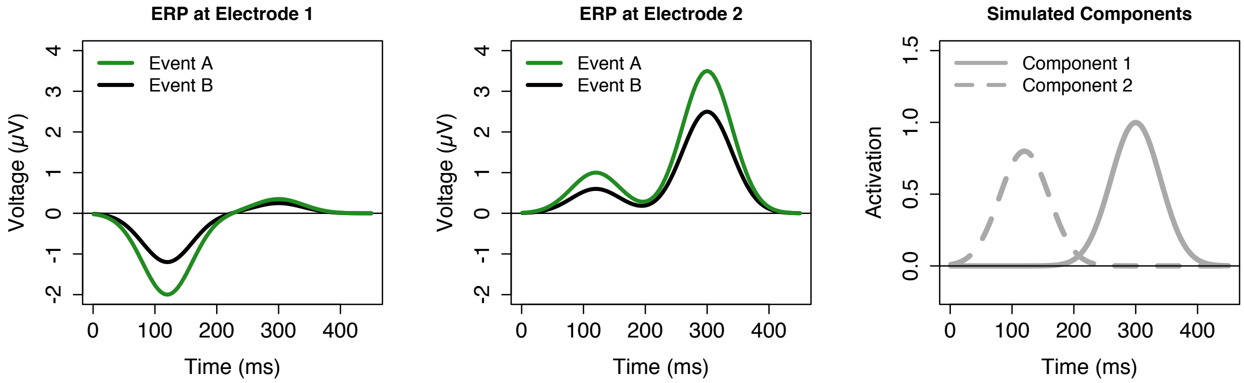

Figure 1. Left & middle panel: Simulated example ERP at two electrode sites. The colors represent the ERPs elicited by two event types (A: green, B: black). Both peaks vary as a function of the event type showing larger amplitudes for Event A than for Event B. Right panel: Factor loadings of the simulated underlying components. The components have nearly no temporal overlap. The data were simulated as described in Scharf and Nestler (2018a) and Scharf and Nestler (2018b), respectively.

second peak is much larger. That is, the topographies of the peak sampling points differ from each other. ERP researchers may be interested in differences between different events in any (or all) of these four features with the ultimate goal to attribute differential brain activity to differences in psychological processes (e.g., differential states of attention). In the remainder of this thesis, I will focus on the case where amplitude differences are investigated (Figure 1).

In the following, a brief description of a traditional recording and analysis procedure for ERPs is provided (e.g., Luck, 2014). The high noise level due to electric noise, electric activity from non-brain sources (e.g., eye blinks or changes in skin conductance), and non-event-related brain activity is an important limitation of EEG recordings. For many purposes, the resulting poor signal-to-noise ratio makes it impossible to reliably use ERPs on the level of a single presentation of an event and the most common remedy to this problem is to record a continuous EEG for a large number of repetitions of the events (e.g., experimental trials). The continuous EEG signal is then cut into epochs around the events, and averaged across all repetitions of the same event.[1,2] The averaging procedure improves the signal-to-noise ratio because it averages

[1]Typically, other pre-processing steps are performed (e.g., filtering or rejection of artifacts) to improve the signal-to-noise ratio by removing artifact-related signal contributions. For the sake of brevity, these steps are not further discussed here and the interested reader is referred to Luck (2014) for a more detailed introduction.

[2]Importantly, the averaging procedure assumes that the ERPs elicited by the events are constant (e.g., no latency-jitter across repetitions) which need not be the case (see, e.g., Mouraux & Iannetti, 2008, for a discussion of the problem of latency-jitter). Throughout this thesis, the approximate absence of latency-jitter will be assumed.

out all electric activity that is not time-locked to the event including a substantial amount of noise. Essentially, this procedure results in a data set with an average ERP for each event type, electrode, and participant. Throughout this thesis, such a data set will be referred to as an *ERP data set*. Most generally, an ERP data set can be written as a 4-dimensional sampling points × participants × electrodes × event type hypermatrix.

After the computation of an ERP data set, the amplitudes of the ERP peaks (for each participant, condition, and electrode site) need to be quantified for subsequent statistical analyses. This has been done using simple peak amplitudes (i.e., the local minima and maxima of the voltage), mean voltages in time windows around the across-participant grand average peaks, or using an area under the curve measure (Luck, 2014, chapter 9) of which mean voltages in time windows are arguably most common. Whatever measure is used to quantify the peak amplitudes, the last step of the data analytic procedure is typically to subject the chosen measure to a general linear model (GLM), most commonly an analysis of variance (ANOVA) with event type and electrode site specified as within-subject-factors - potentially followed up by post hoc pairwise comparisons. Functional interpretations of the ERPs are then made based on the results of the statistical analyses.

The described data analytic procedure has dominated ERP research for decades but it suffers from a number of severe shortcomings. The most fundamental issue is that the electric potential recorded from the scalp is a 2D projection of brain activity in 3D space in the brain. Consequently, ERPs are a linear superposition of actual event-related source signals, typically referred to as *components* (see Nunez & Srinivasan, 2006, for physiological details). The described traditional measures implicitly assume that the peaks or the time points that enter the measure, respectively, reflect activity from a single underlying component. The right-most panel in Figure 1 depicts the underlying components of the simulated example ERP. In this case, the peaks of the observed ERPs validly reflect the peaks of the underlying components and all amplitude measures would correctly quantify the intended component. The reason for this is that the components do not overlap in time at all. Figure 2 shows example where the temporal overlap of the components was increased by shifting the peak latencies of the components closer together. Especially at Electrode 2 (middle panel), the two components are clearly

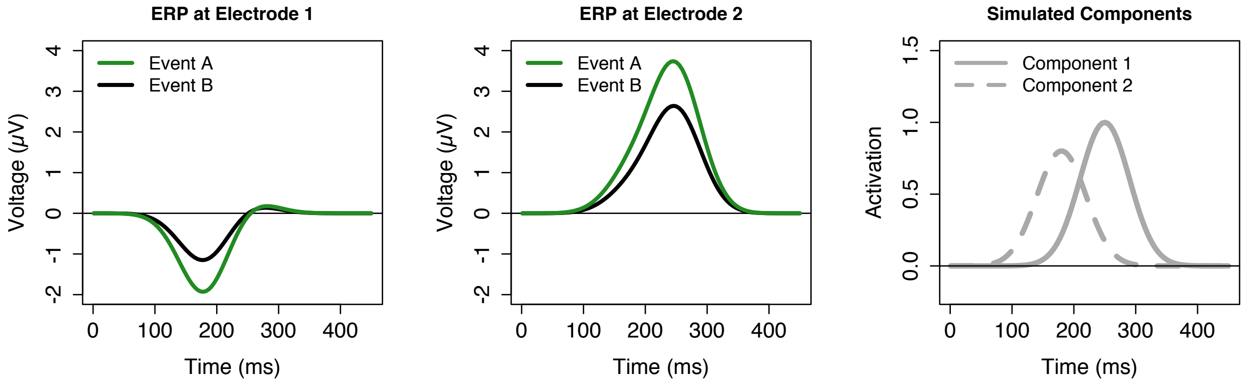

Figure 2. Left & middle panel: Simulated example ERP at two electrode sites. The colors represent the ERPs elicited by two event types (A: green, B: black). Both peaks vary as a function of the event type showing larger amplitudes for Event A than for Event B. Right panel: Factor loadings of the simulated underlying components. Please note the increased temporal overlap of the components compared to the example in Figure 1. The data were simulated as described in Scharf and Nestler (2018a) and Scharf and Nestler (2018b), respectively.

conflated in the observed ERP affecting the validity of amplitude measures. There is consensus among ERP researchers that the second example is more realistic (e.g., Luck, 2014, chapter 2).

Apart from that, modern ERP data sets challenge researchers with their sheer size. It is not uncommon for modern EEG recording systems to record data from 64 or more electrode sites at a sampling rate of 1000 Hz (i.e., 1 sampling point every millisecond). The resulting ERP data sets therefore tend to be very large, and analyses of such large data sets using the traditional data analytic approach suffer both from a multivariate comparison and an interpretation problem. For instance, in a setup with 64 electrodes, 2 event types, 64 pairwise comparisons could be conducted per peak to compare the amplitudes between the event types at all electrode sites – resulting in a massive multivariate comparison problem, and, consequently, in a loss of power due to controlling the family-wise error. In an attempt to cope with the challenging size of modern ERP data sets, a range of questionable data analytic practices emerged that suffer from extremely high type I error rates (Luck & Gaspelin, 2016).

Whereas sophisticated solutions to the multivariate comparison problem are available (Groppe, Urbach, & Kutas, 2011a, 2011b), they do not solve the other major problem which concerns the interpretation of the results. Specifically, the large number of comparisons raises questions such as how many electrode sites with a significant difference are necessary to call an experimental effect substantial or whether the significant differences at two distant electrode sites should be attributed to the same component or not. These problems are even aggravated when each sampling point is treated separately instead of computing one of the described amplitude measures. Very often, researchers choose to reduce this issue by a-priori specifying electrodes of interest (and ignoring other electrode sites in the statistical analyses) or averaging the values from several adjacent electrodes – effectively discarding information from the data. Taken together, the outlined limitations of traditional data analytic procedure motivated the development of several multivariate analysis methods that estimate the unobserved underlying components and reduce the size of ERP data sets in an information preserving way.

1.3 Exploratory factor analysis of event-related potentials

A variety of decomposition methods has been used for ERP data among them Principal Component Analysis (PCA), Independent Component Analysis (ICA), Wavelet-based decompositions, multimode PCAs, and source reconstruction methods (Groppe, Makeig, Kutas, & Diego, 2008; Möcks, 1988; Mørup, Hansen, Herrmann, Parnas, & Arnfred, 2006). From a physiological point of view, source reconstruction is clearly the most reasonable modeling approach because it considers the physical processes that underly the conduction of the electric potential from the source to the electrodes on the head. However, source reconstruction depends on accurate knowledge of both the exact positions of the electrode sites and the participants' neuroanatomy, and even small inaccuracies may result in considerable localization errors (Slotnick, 2005). Many ERP experiments are conducted in the absence of additional information from functional and structural magnetic resonance imaging so that accurate source reconstruction is not an option. The ability to provide data-driven summaries of ERP data sets under these circumstances is the main advantage of the former decomposition methods that rely only on statistical properties of ERP data sets. Although these methods lack direct physiological interpretability, they are still able to extract reliable and substantively meaningful features from ERP data sets (e.g., Beauducel, Debener, Brocke, & Kayser, 2000; Fogarty, Barry, De Blasio, & Steiner, 2018).

The present thesis focuses on the analysis of ERP data sets using PCA or, more generally, exploratory factor analysis (EFA). Technically, PCA and EFA are different in so far that the EFA model contains an explicit error term for each observed variable whereas the PCA model does not. The results of EFA and PCA differ to the extent to which the variance of these error terms differs from zero (De Winter & Dodou, 2014; Widaman, 2007, 2018) and, in this sense, PCA may be conceptualized as a restricted EFA model with all error variances restricted to zero (McDonald, 1996). For ERP data, the differences between EFA and PCA are typically considered negligible (Dien & Frishkoff, 2005). Although PCA is more common than EFA in research applications, throughout this thesis, the more general term EFA will be used to refer to this data analytic approach, and the term *factors* will be used in reference to the estimated latent entities – also to avoid confusion with the term component in reference to the source activity that generated the potential on the scalp.

In the following, the interpretation of the EFA model in the context of ERP data and the prototypical steps of an EFA are briefly outlined. For further details, including formal mathematical definitions, the reader is referred to chapter 2. The most important decision when applying EFA to ERP data is the choice of the analysis mode. In order to apply EFA to an ERP data set, the initial 4-dimensional hypermatrix must be rearranged as 2-dimensional matrix. Analysis mode refers to the dimension of the initial hypermatrix to which dimension reduction should be applied. Typically, the analysis mode is arranged as the columns of the 2D matrix, and the remaining dimensions are concatenated in the rows. Either the sampling points or the electrodes (or both consecutively; e.g., Dien, 2010; Dien & Frishkoff, 2005; Spencer, Dien, & Donchin, 2001) may be treated as analysis mode. The present thesis focusses on the case where the sampling points are the analysis mode – often referred to as *temporal EFA*.

Essentially, temporal EFA decomposes the ERP waveform into a set of latent factors where the factor loadings reflect the time courses of the latent factors, and the amplitudes are represented by the factor scores. The factor loadings are assumed to be fixed across all participants, electrodes, and conditions whereas the factor scores are allowed to vary. The observed ERP data can be recomputed by multiplying the factor loadings by the factor scores (cf. chapter 2, Footnote 4). Figure 3 shows the results of an EFA analysis for a simplistic example with two

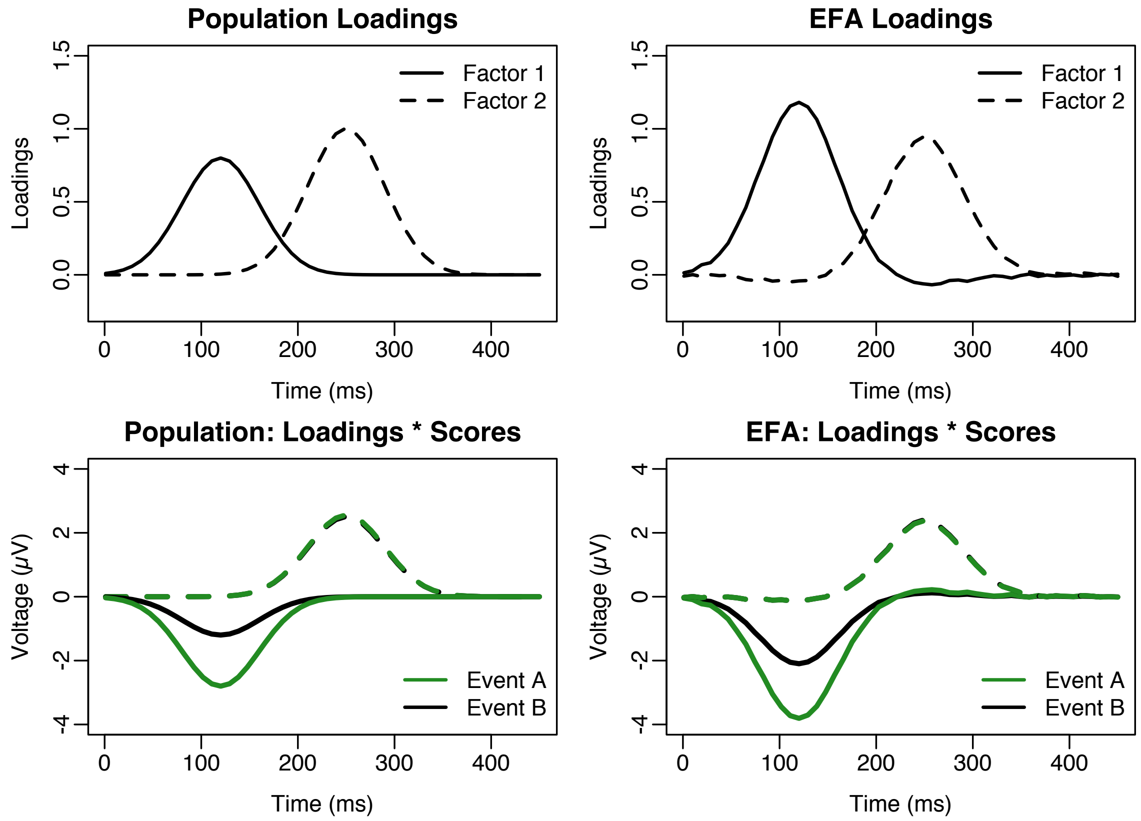

Figure 3. Upper panel: Factor loadings of a simulated example population (left) and the corresponding factor loadings estimates from a Promax-rotated EFA (right). Lower panel: Simulated ERPs for Event A (green) and B (black) computed in a factor-wise manner by multiplying the factor loadings by the factor scores for the population values (left) and the Promax-rotated EFA (right). The data were simulated as described in Scharf and Nestler (2018a) and Scharf and Nestler (2018b), respectively.

factors and two event types. The left-hand panels show the population factor loadings and the population ERP computed in a factor-wise manner. Only the factor scores of the first factor varied as a function of the event type. The right-hand panel shows the results of a Promax-rotated EFA for a sample data set. One can see that the estimated factor loadings reflect the population factor loading quite well, and, most importantly, that the estimated differences between Event A and Event B resemble the population pattern.[3]

An application of temporal EFA to ERP data typically consists of the following steps (Dien, 2012): First, the number of factors to be extracted needs to be determined which is typically done using the Parallel Test method (Horn, 1965). Then, an initial unrotated factor solution is

[3]One can also see that the estimated factor loadings are not on the same scale as the simulated factor loadings but this is irrelevant in the context of this illustration. The reader is referred chapter 2 for a detailed explanation of the rescaling.

estimated in which the factors are uncorrelated and estimated with the restriction that the first factor accounts for the largest proportion of variance, the second factor accounts for the second largest proportion of variance, and so forth. In order to achieve an interpretable solution, the initial factors are rotated, yielding the estimated factor loadings and factor correlations. The gold standard for ERP data is an oblique Promax rotation (Dien, 1998, 2010; Dien, Beal, & Berg, 2005). Finally, the factor scores are estimated, typically using a regression method (Thomson, 1935; Thurstone, 1935), and subjected to a general linear model (GLM) for subsequent statistical analysis of amplitude differences, most commonly, an analysis of variance (Dien, 2016).

As functional interpretations of the factors are based on the results of the subsequent GLM, an important methodological concern is to ensure the GLM parameters are unbiased. The term *variance misallocation* has been introduced in reference to the case where the GLM parameters are biased (Wood & McCarthy, 1984). In the worst case of variance misallocation, false positive results can occur, that is, differences in the factor scores between event types may be found although the population effect is zero. Such an extreme case of variance misallocation is illustrated in Figure 4. In the EFA solution (lower-right panel), there is a notable difference between event types in both factors although no such difference exists in the population (lower-left panel). It is also apparent (top-panel) that the estimated factor loadings do not resemble the population factors as well as in Figure 4. This is in line with previous research emphasizing the role of the factor loading estimates, and, hence, the importance of the rotation step for the occurrence of variance misallocation (Achim & Marcantoni, 1997; Dien, 1998, 2010; Möcks & Verleger, 1986).

The present thesis used a simulation approach to study the occurrence of variance misallocation. Notably, previous studies predominantly relied on simulated data from virtual head models (e.g., Dien et al., 2005) or real ERP data sets (e.g., Kayser & Tenke, 2003). Complementing these studies, a Monte Carlo approach was taken, that is, data were generated from the EFA model (cf. Chapter 2, for details) by defining a set of population parameters (i.e., factor loadings, factor correlations, and factor scores) and drawing random samples from the model-implied multivariate normal distribution. The main advantage of this approach is that biases in the

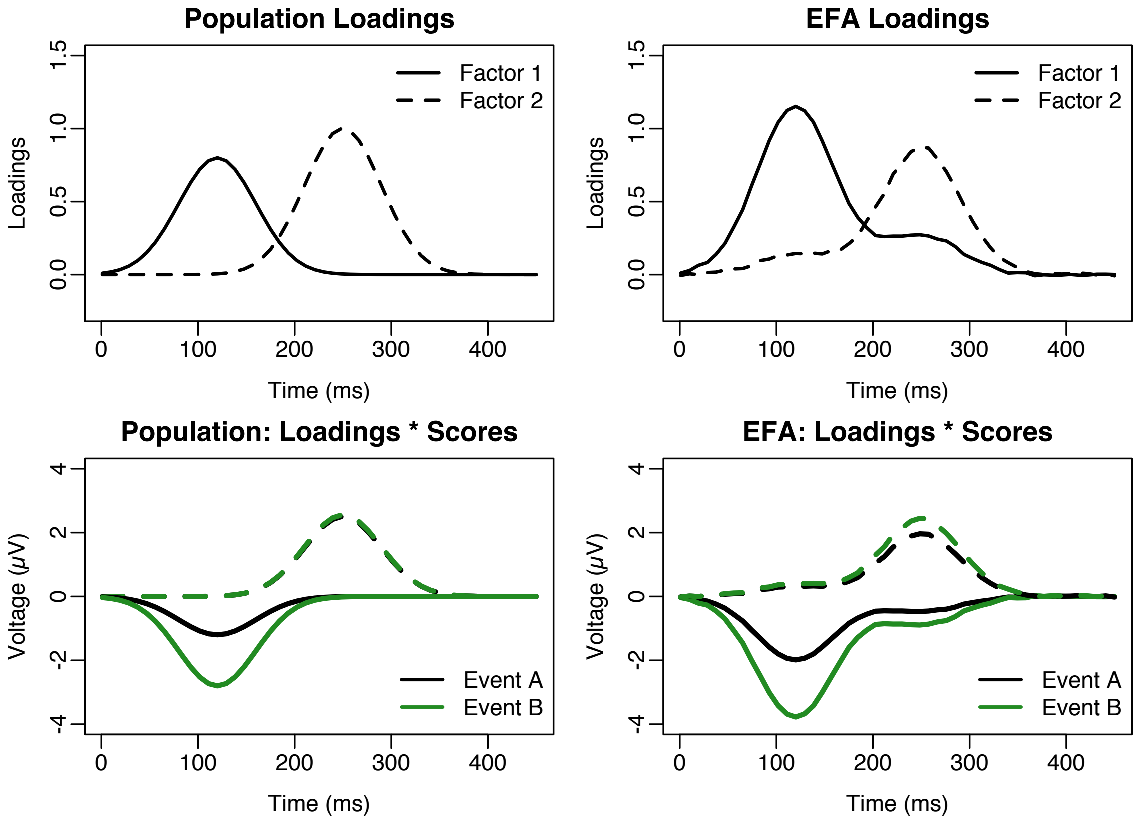

Figure 4. Upper panel: Factor loadings of a simulated example population (left) and the corresponding factor loadings estimates from a Varimax-rotated EFA (right). Notably, the estimated factor loadings are severely biased with respect to the population loadings. Lower panel: Simulated ERPs for Event A (green) and B (black) computed in a factor-wise manner by multiplying the factor loadings by the factor scores for the population values (left) and the Varimax-rotated EFA (right). The data were simulated as described in Scharf and Nestler (2018a) and Scharf and Nestler (2018b), respectively.

model parameters can be investigated more formally because a ground truth is known for each parameter. In addition, the approach is especially well suited to study how the values of the population parameters themselves affect the performance of EFA for ERP data because these parameters can be manipulated directly.

1.4 Research objectives

The main goal of the present thesis was to investigate how the risk of variance misallocation can be minimized in applications of factor analytic methods to ERP data. In the subsequent chapters, the determinants of the occurrence of variance misallocation are identified (chapter 2), and recently proposed improvements to EFA approaches are investigated that can considerably reduce the risk of variance misallocation (chapters 3 & 4). In the following, the research questions of the subsequent chapters are outlined in more detail.

In chapter 2, the principles behind variance misallocation are investigated by means of an analytic decomposition of the factor (co-)variance matrix and a Monte Carlo simulation. The study sets out from the fact that ERP data sets differ from psychometric data sets, for which EFA was originally intended, in at least two ways: First, the observations in the rows of an ERP data matrix are not independent and exchangeable. Rather, they are well structured and some observations are more strongly correlated with each other than others because they stem from the same electrode site, the same participant, and/or the same event type. This fact cannot be acknowledged in the original EFA model. Second, latent factors extracted from ERP data are likely to have a considerable temporal overlap (i.e., a considerable amount of cross-loadings), and this can hardly be influenced by researchers themselves. The study presented in chapter 2, is concerned with the consequences of these two properties for the estimation performance and interpretability of the EFA parameters.

Addressing the consequences of the neglected structure of ERP data sets, in chapter 3, exploratory structural equation modeling (ESEM) is proposed as an alternative to EFA. ESEM can properly acknowledge the structure of ERP data sets, for instance, providing substantively interpretable factor correlation estimates. Essentially, ESEM expands EFA by a structural model in which predictors of the latent variables can be specified (Asparouhov & Muthén, 2009). The study presented in chapter 3 discusses which structural model should be specified in order for ESEM to be useful for typical ERP research questions. Finally, a Monte Carlo study is reported that investigated whether ESEM is less prone to biases in the parameter estimates than EFA.

Finally, chapter 4 addresses the influence of the rotation criterion on the quality of the factor solution. In line with previous literature (Dien, 1998; Möcks & Verleger, 1986; Verleger & Möcks, 1987), the results presented in chapter 2 emphasize the importance of the rotation step in EFA for the occurrence of variance misallocation. As a rotation step is an essential part of ESEM as well, ESEM suffers from biases due to factor rotation as well. It is well-known that simple structure rotation can result in biased factor loading estimates, especially in the presence of substantial cross-loadings (Browne, 2001). Recently, regularized (or sparse) estimation of factor models has been proposed as a substitute for factor rotation. Due to it different concept of simplicity, regularized factor analaysis is able to provide good factor solutions even in conditions under which rotated EFA does not (see Trendafilov, 2014, for a review). Whereas many different regularized factor analysis methods have been proposed (Hirose & Yamamoto, 2014; Huang, Chen, & Weng, 2017; Jacobucci, Grimm, & McArdle, 2016; Trendafilov & Adachi, 2015; Trendafilov, Fontanella, & Adachi, 2017), to the best of the author's knowledge, no extensive comparison of the performance of regularization and simple structure rotation has been available. Closing this gap, the simulation study reported in chapter 4 was conducted in which the performance of simple structure rotation and regularization was compared for a wide range of factor loading patterns. Although the analysis of ERP data was not in center of interest in this study, its results have important implications for the further development of the ERP analysis framework outlined in chapter 3.

Chapter 2

Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition

To obey the publisher's copyright restrictions, this chapter contains the post-peer review author version of the manuscript. The published article can be found under the following reference:

Scharf, F., & Nestler, S. (2018b). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119–136. doi:10.1016/j.ijpsycho.2018.03.019

Nachweis über Anteile der Co-Autoren, Florian Scharf
Advances in the analysis of event-related potential data with factor analytic methods

Nachweis über Anteile der Co-Autoren:

Titel: **Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition**

Journal: International Journal of Psychophysiology

Autoren: Florian Scharf, Steffen Nestler

Anteil Florian Scharf (Erstautor):

- Projektidee
- Konzeption der Simulationsstudie
- Programmierung und Durchführung der Simulationsstudie
- Auswertung der Simulationsstudie
- Herleitung der Varianz-Kovarianz-Dekomposition
- Schreiben der Publikation

Anteil Steffen Nestler (Senior-Autor):

- Konzeption der Simulationsstudie
- Schreiben der Publikation

Florian Scharf

Steffen Nestler

Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition

Florian Scharf and Steffen Nestler

University of Leipzig

Author Note

Florian Scharf, Institute of Psychology, University of Leipzig, Germany. Steffen Nestler, Institute of Psychology, University of Leipzig, Germany. We embrace the values of openness and transparency in science (http://www.researchtransparency.org/). We have therefore published all data necessary to reproduce the reported results and provide reproducible scripts for all data analyses that were reported in this paper on the Open Science Framework.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest: none

Correspondence concerning this article should be addressed to Florian Scharf, Neumarkt 9 - 19, 04109 Leipzig, Germany Email: florian.scharf@uni-leipzig.de. Phone: +49 341 - 9735924

Abstract

Temporal exploratory factor analysis (EFA) is commonly applied to ERP data sets to reduce their dimensionality and the ambiguity with respect to the underlying components. However, the risk of variance misallocation (i.e., the incorrect allocation of condition effects) has raised concerns with regard to EFA usage. Here, we show that variance misallocation occurs because of biased factor covariance estimates and the temporal overlap between the underlying components. We also highlight the consequences of our findings for the analysis of ERP data with EFA. For example, a direct consequence of our expositions is that researchers should use oblique rather than orthogonal rotations, especially when the factors have a substantial topographic overlap. A Monte Carlo simulation confirms our results by showing, for instance, that characteristic biases occur only for orthogonal Varimax rotation but not for oblique rotation methods such as Geomin or Promax. We discuss the practical implications of our results and outline some questions for future research.

Keywords: Event-related Potential, Principal Component Analysis, Exploratory Factor Analysis, Variance misallocation

Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition

1 Introduction

The recording of electric brain activity from the scalp surface via electroencephalography (EEG) is a very popular tool among cognitive neuroscientists. Event-related potentials (ERPs) are commonly applied to investigate how brain activity reflects the differential processing of events (e.g., different stimuli or responses). Typically, the EEG signal is cut into epochs around the events of interest and averaged across all replications of the same event to improve the signal-to-noise ratio (SNR), thus resulting in an average ERP per participant, electrode, and condition. The high dimensionality of these data results in various challenges to the analyst. For instance, a rather simple experiment with 2 conditions and 20 participants using 64 electrodes and an epoch length of 500 ms (500 Hz sampling rate) produces $2 \cdot 20 \cdot 64 \cdot 250 = 640,000$ data points. As a consequence of these large data sets, ERP experiments suffer from a massive multiple comparison problem (but see, for instance, Groppe, Urbach, & Kutas, 2011a, 2011b; Maris, 2004, for solutions). In our example, comparing the two conditions at all sampling points and all electrodes would result in 16,000 possible statistical tests.

Another problem arises from the fact that the electric potential observed on the scalp surface is only a (weighted) 2D summation of the underlying electrical potentials produced by neuronal populations in 3D space. Thus, when analyzing the observed mixture of signals, it is difficult to determine the true underlying ERP components and which of them is responsible for the observed differences between conditions. Since neither shape nor allocation of effects can be determined with respect to the underlying components, it is hard to draw inferences from ERPs on the basis of the raw voltage (see e.g., Luck, 2014, p. 52). Exploratory factor analysis (EFA) has been suggested as a way to overcome the multiple test problem and to characterize the observed mixture of signals (e.g., Donchin, 1966; Donchin & Heffley, 1978; Chapman & McCrary, 1995; Dien, 2012). The goal of EFA is to describe the large number of data points as a function of a few underlying sources of variation that are called factors or components.[1] Amplitude differences between conditions and participants can then be analyzed on the level of these factors, potentially reducing the ambiguity of the raw signal. In addition, the number of comparisons can be reduced since the amplitude of each factor can be described as a single value without the need for analysis time windows. Thus, the use of EFA can significantly decrease the complexity of ERP data.

Different methods exist within the EFA framework such as *Principal Component Analysis* (PCA) or *Maximum Likelihood Factor Analysis* (MLFA). Due to its computational simplicity, previous methodological research on ERP data has focused almost exclusively on PCA (James, Witten, Hastie, & Tibshirani, 2013; Mulaik, 2010; Widaman, 2007), which is implemented in the *PCA ERP Toolbox* (Dien, 2010b), a common toolkit for ERP analyses. An important issue that was investigated in prior research was how the rotation method influences the adequacy of the EFA solution (Achim & Marcantoni, 1997; Beauducel & Debener, 2003; Dien, 1998; Dien, Beal, & Berg, 2005; Kayser & Tenke, 2003; Möcks & Verleger, 1986; Wood & McCarthy, 1984). Here, we want to examine the influence of both points with regard to a commonly studied problem with EFA for ERP data: *variance misallocation* (Wood & McCarthy, 1984). Based on a decomposition of the inter-factor covariance, we explicate why variance misallocation occurs and how its influence can be decreased by using the correct rotation method. Furthermore, we confirm our theoretical results in a Monte Carlo simulation.

The present article is organized as follows: We will first briefly describe the common factor model and its application to ERP data. Then, we will introduce the problem of variance misallocation, we theoretically explain how variance misallocation can be caused by biased factor covariance estimates and temporal overlap between factors, and highlight the consequences of our results for the analysis of ERP data with

[1]Although the resulting estimated components in a Principal Component Analysis (PCA) or Independent Component Analysis (ICA) are usually called components, throughout this article, we will refer to *estimated* components as *factors* irrespective of the estimation method that was used. We do so to be as clear as possible about the fact that the extracted factors are the result of a mathematical transformation of the data and do not necessarily reflect a physical reality although this would be a desirable outcome (for an example, see Makeig et al., 1999).

EFA. We will then report the results of a simulation study in which we manipulated both causes of variance misallocation and compared the performances of two different estimation methods (PCA and MLFA) for the factor model and three different rotation methods (Varimax, Promax, and Geomin) for the occurrence of variance misallocation. Finally, we will discuss the implications of our findings for researchers who want to apply EFA in case of ERP data.

2 Exploratory factor analysis for ERP data

In the common factor model it is assumed that the observed variables are a linear function of a set of shared *latent* sources of variation called *common factors* (see e.g., Dien, 2012; Dien & Frishkoff, 2005, for more exhaustive descriptions). The model was originally developed in a psychometric context in order to summarize the answers from multiple questionnaire items in a few psychological constructs such as personality traits (see e.g., Mulaik, 2010). Questionnaire data can be organized in a simple 2-dimensional matrix with n participants (observations) in the rows and the p items (variables) in the columns. However, data from ERP experiments tend to have a much more complicated structure. For the remainder of the article, we will assume that the data were averaged across all trials (per participant, electrode, and condition). Thus, they can be arranged in a 4-dimensional *electrodes* × *sampling points* × *participants* × *conditions* matrix.

In principle, dimension reduction can be applied in either the spatial, the temporal or in both domains (Dien, 2010a; Dien & Frishkoff, 2005), yielding factors that summarize electrodes or time points, respectively, with a common activity pattern. Here, we focus on dimension reduction in the temporal domain. According to the common factor model, the voltage at each sampling point is modeled as a function of a few underlying factors. Mathematically, this is expressed by assuming that the voltage t_j at sampling point j (where $j = 1, 2, ..., p$ and p is the number of sampling points) is a weighted linear combination of m latent factors:

$$t_j = \lambda_{j1} \cdot \eta_1 + \lambda_{j2} \cdot \eta_2 + ... + \lambda_{jm} \cdot \eta_m + \epsilon_j, \tag{1}$$

where η_k denotes the vector of factor scores for factor k, with k ranging from 1 to m and m has to be determined during the analysis. The factor scores represent the characteristic value (i.e., amplitude) of each factor for each electrode of each participant in each condition. λ_{jk} are weights called *factor loadings*, which express how much each factor is reflected by each sampling point. The higher the absolute value of λ_{jk} the stronger the influence of factor k on the voltage at sampling point j. When $\lambda_{jk} = 0$, factor k does not contribute to the observed voltage at sampling point j at all. The error term ϵ_j is the part of t_j that is not explained by the factors (i.e., noise). While the factor loadings (i.e., the time course) are fixed for all observations, amplitude differences between participants, conditions, and electrodes are reflected by the factor scores. Note that there is no intercept term in Equation 1 and therefore the model only captures the *variation* around the grand average waveform of the sampling points. That is, factors are formally defined as *sources* of variation and a voltage deflection that is constant across all participants, conditions, and electrodes would not be considered a factor.

Usually, the common factor model is expressed in matrix notation as:

$$T = \Lambda \cdot \eta + \epsilon \tag{2}$$

where T is the $p \times n$ matrix of all (centered) sampling point variables from all n observations, η is the $m \times n$ matrix of factor scores, Λ is a $p \times m$ matrix of factor loadings, and ϵ is a $p \times n$ matrix of error terms. The advantage of writing the EFA model as it is written in Equation 2 is that the covariance matrix between the sampling points, Σ, can easily be derived (see Mulaik, 2010, p. 136). This matrix is given by:

$$\Sigma = \Lambda \Phi \Lambda' + \Theta, \tag{3}$$

where Φ is the covariance matrix of the latent factors and Θ is the covariance matrix of the error terms. Equation 3 shows that the covariance between sampling points is effectively decomposed into the variances and covariances of the factors and some error (co-)variances. It is often assumed that the error covariance matrix Θ is a diagonal matrix, that is, that the error terms are mutually uncorrelated.

MLFA and PCA differ in whether an explicit error term is modeled or not. Specifically, PCA does not contain an error covariance matrix and thus does not explicitly distinguish between substantial and unique (i.e., residual) factors. In spite of that, it is one of the most popular methods for estimating the parameters of the EFA model (Fabrigar, Wegener, MacCallum, & Strahan, 1999). MLFA, on the other hand, is used to estimate the full common factor model by maximizing the likelihood function of the covariance matrix Σ (Jöreskog, 1967; Mulaik, 2010). To this end, MLFA has to assume that the data is multivariate normally distributed.[2] Furthermore, it is assumed that the factors and the error terms are uncorrelated. In general, PCA and estimation methods that explicitly consider error variance, such as MLFA, are mathematically equivalent if the latent factors explain (nearly) all of the variance in the observed variables. Specifically, differences in the loading estimates between PCA and genuine EFA methods are a function of the communality of the involved variables (i.e., the variance of a variable accounted for by the factors) and the number of variables per factor (De Winter & Dodou, 2016; Widaman, 1990, 1993, 2007). For ERP data this implies that differences between PCA and MLFA estimates are smaller 1) the higher the explained variance of a sampling point's data (typically, this variance is considered to be very high, see Dien & Frishkoff, 2005), and 2) the larger the time range of the components (i.e., stronger differences between factors that are only present over a very small time range).

In general, the factor loadings cannot be determined in a mathematically unique manner because the EFA model is unidentified. That is, an infinite set of factor loading matrices with equal fit exists for a given number of factors and a given data set. Therefore - after determining the number of factors to be retained and estimating an initial solution from the covariance matrix of the sampling points (Dien et al., 2005;

[2]One might reasonably argue that the multivariate normality assumption is unlikely to hold for ERP data. However, it has been shown that the MLFA estimators can be derived without this assumption (Howe, 1955). A violation of the multivariate normality assumption affects the calculation of standard errors for the estimates of Λ and Θ, but the estimation remains unbiased (Mulaik, 2010). Therefore, MLFA can be assumed to be applicable to ERP data.

Dien, 2012) - the initial solution may be *rotated* in order to find a more interpretable solution (e.g., Mulaik, 2010). There are two major forms of rotation: (1) orthogonal rotations (e.g., *Varimax*, Kaiser, 1958, 1959) in which the factor scores are forced to be uncorrelated, and (2) oblique rotations (e.g., *Promax*, Hendrickson & White, 1964), which allow for a certain correlation between the factor scores. Factor rotation effectively reallocates the variance between factors, aiming for a so-called *simple structure* (Thurstone, 1954). In an ideal simple structure, each factor is well-defined by a distinct set of variables, that is, the variables have high loadings on only one factor and zero loadings on the remaining factors. Most of the available rotation methods are designed to rotate the factor loadings as close as possible to a simple structure (for an overview, see Browne, 2001). Notably, in the context of ERP data, the factor loadings reflect the time courses of the factors and an ideal simple structure would be satisfied if there was no temporal overlap between the components. Hence, rotation procedures will prefer solutions in which each sampling point is as uniquely as possible assigned to a single factor rather than several factors – potentially oversimplifying situations when the temporal overlap of the underlying factors is high.

After the factor loadings have been rotated, factor scores need to be estimated to examine experimental condition effects. Several different approaches exist for this such as the commonly used *regression approach* (Thomson, 1935; Thurstone, 1935). Having obtained the factor scores, there are two ways to test for condition effects in the amplitudes: Researchers can use the factor scores as dependent measures in subsequent analyses (typically t-tests, ANOVAs, or multiple regressions; e.g., Boxtel, 1998) or they can use the factor scores to *reconstruct* the raw data in a factorwise manner and use these in subsequent analyses (Dien, 1998, 2012). Whatever of the two methods is employed, the very idea of both methods is that EFA has significantly decreased the complexity of the original ERP data.

We have now seen how EFA can be used to summarize ERP data sets. In real application scenarios, it remains unknown in how far the factor solution reflects any underlying unobserved physical entity and - at best - EFA increases the interpretability and reliability of the extracted information (Beauducel, Debener, Brocke, & Kayser, 2000). In the following sections (3 & 4), we will revert our view and assume that we know the underlying factor structure (because we constructed the data with a certain factor structure in mind) in order to see how well EFA can blindly recover our constructed factors.

3 Mechanisms of variance misallocation

In light of the central goal of testing for condition differences in the factor scores (i.e., amplitudes) or the reconstructed raw data, it is important that experimental effects are attributed to the correct factors because otherwise functional interpretations of the factors might be misguided. The term *variance misallocation* has been introduced in reference to the case where variance is incorrectly attributed to factors that are actually not affected by the experimental manipulation (Wood & McCarthy, 1984). As described earlier, the 'translation' of the data into factors (and vice versa) is mainly controlled through the factor loadings. Hence, whatever causes systematic biases in the factor loading estimates will trigger variance misallocation to some extent (Achim & Marcantoni, 1997). Two important influences on the risk of biased factor loadings have been identified in the methodological literature: biases in the factor correlations or factor covariance parameters, respectively, and variables' cross-loadings (e.g., Schmitt & Sass, 2011, for a discussion).

With regard to the factor covariance parameters, we note that - when EFA is applied to ERP data - the inter-factor covariance matrix has a special structure due to the arrangement of the initial data matrix. As stated earlier, EFA is applied to a data matrix in which the rows represent observations from all participants in all conditions and from all electrodes. This arrangement of the data has the consequence that each estimated factor *variance* is a combination of variance due to differences between participants *and* due to differences between conditions and electrodes. Furthermore, the estimated factor *covariance* is influenced by the condition effects and the topography.

To illustrate, assume a simple fictional data set that contains the average ERPs to a certain stimulus, recorded at two electrode sites from several participants. To keep the equations simple, we further assume, that two factors are sufficient to describe our fictional data set. Condition effects are considered by writing the factor scores as dependent variables in a linear regression with electrode as predictor:

$$\begin{aligned} \eta_1 &= b_{01} + b_{11} \cdot x + \delta_1 \\ \eta_2 &= b_{02} + b_{12} \cdot x + \delta_2 \end{aligned} \tag{4}$$

Here, η_1 and η_2 are vectors containing the factor scores for all observations (i.e., data from each participant at each electrode) for the respective factors, x is the effect-coded electrode, that is, a dichotomous variable that is either $+1$ (electrode 1) or -1 (electrode 2). Consequently, $b_{0\cdot}$ is the average factor score across both electrodes of the respective factor and $b_{1\cdot}$ is half the mean difference of the factor scores between the two electrodes. Finally, $\delta_\cdot$ denote error terms that reflect the deviation of each participant-electrode combination from the predicted value. We note that while x, δ_1 and δ_2 are vectors containing one element for each participant-electrode combination, all other parameters are simple scalars.

For this example, the factor covariance matrix is (see Equation 3)

$$\begin{aligned} \Phi &= \begin{pmatrix} var(\eta_1) & cov(\eta_1, \eta_2) \\ & var(\eta_2) \end{pmatrix} \\ &= \begin{pmatrix} var(b_{01} + b_{11} \cdot x + \delta_1) & cov(b_{01} + b_{11} \cdot x + \delta_1,\ b_{02} + b_{12} \cdot x + \delta_2) \\ & var(b_{02} + b_{12} \cdot x + \delta_2) \end{pmatrix} \\ &= \begin{pmatrix} b_{11}^2 \cdot var(x) + var(\delta_1) & b_{11} \cdot b_{12} \cdot var(x) + cov(\delta_1, \delta_2) \\ & b_{12}^2 \cdot var(x) + var(\delta_2) \end{pmatrix} \\ &= \begin{pmatrix} \underbrace{b_{11}^2}_{\substack{\text{topographic variance} \\ \text{of factor 1}}} + \underbrace{var(\delta_1)}_{\substack{\text{between-person variance} \\ \text{of factor 1}}} & \underbrace{b_{11} \cdot b_{12}}_{\substack{\text{covariation due to} \\ \text{topographic overlap}}} + \underbrace{cov(\delta_1, \delta_2)}_{\substack{\text{covariation across} \\ \text{participants}}} \\ & \underbrace{b_{12}^2}_{\substack{\text{topographic variance} \\ \text{of factor 2}}} + \underbrace{var(\delta_2)}_{\substack{\text{between-person variance} \\ \text{of factor 2}}} \end{pmatrix} \end{aligned} \tag{5}$$

where the third transformation follows from applying standard rules concerning the (co-)variance and the assumption that all covariances between residuals and predictors in Equation 4 are zero. The last transformation follows from the assumption that for each participant both electrode sites were recorded. This entails that the data set is balanced and that the variance of the effect-coded electrode x equals 1.

Equation 5 shows that the estimated factor variance can be decomposed into topographic variance due to electrode differences and residual variance due to unexplained between-participant differences. Similarly, the estimated factor covariance (i.e., the off-diagonal element in Φ) can be split up into covariation due to the topographic overlap and covariation due to differences between participants. Remarkably, the former occurs whenever the scores (i.e., amplitudes) of *both* factors vary systematically between electrodes. Thus, the correlation of the factors across electrodes can roughly be understood as the amount of topographic overlap and it is therefore virtually impossible that the factors are uncorrelated (see also Dien, 2010a). For instance, factor 1 might show a more positive peak at electrode 1 compared to electrode 2 while factor 2 might show a reverse pattern. Therefore, across participants, a high first factor is typically followed by a low second factor at electrode 1 and vice versa at electrode 2. In other words, the peaks are negatively correlated. This covariation due to topographic overlap will differ from zero, except if the peak of either of the factors is constant across both electrodes (i.e. either, or both, $b_{11} = 0$ or $b_{12} = 0$).

We note that one can extend these results to the multiple electrode case where the (co-)variance can also be decomposed into scalp topography across multiple electrodes, condition effects and between-participant differences (see Appendix A). Although the described mixture of (co-)variance sources has been noted before (Dien, 1998; Dien et al., 2005; Dien, 2010a; Hunt, 1985; Möcks & Verleger, 1985), the explicit derivation of the covariance decomposition for ERP data has - as far as we know - not been presented.

However, the decomposition is informative as it allows us both to understand *why* variance misallocation occurs and also how its influence can be decreased. For instance, when using an orthogonal rotation method such as Varimax, one implicitly makes the constraint during estimation that the *sum* of the covariances caused by participants, electrodes and condition effects equals zero. This might occur when the covariance terms are in fact zero or when one covariance term is the exact counterpart of the other terms. However, we believe that this precondition is rather unlikely to hold in practice, because, typically, there are multiple electrodes and conditions. The sum of the covariance terms will thus deviate from zero in most circumstances, and the more it deviates from zero, the more of the shared variance between the factors is represented as spurious cross-loadings (Sass & Schmitt, 2010; Schmitt & Sass, 2011).

This suggestion is well in line with previous simulation research on applications of EFA to ERP data showing that variance misallocation can be reduced by using oblique rotations (e.g., Promax; Beauducel, 2017; Dien, 1998; Dien et al., 2005; Dien, 2010a). However, arguments in favor of orthogonal rotations are still made (Kayser & Tenke, 2006) and orthogonal rotation methods (i.e., mainly Varimax) are still applied (e.g., Kayser & Tenke, 2015; Barry, De Blasio, Fogarty, & Karamacoska, 2016). Our theoretical results clearly show that oblique rotation methods are superior to orthogonal rotation methods and that the use of orthogonal rotations carries a high risk for the occurrence of variance misallocation. To reduce this risk, oblique rotation methods should be the default choice when applying EFA to ERP data.

Apart from the inter-factor covariance, previous research has identified *temporal overlap* between factors as an influential feature of the underlying ERP components triggering variance misallocation (Dien, 1998; Dien et al., 2005; Möcks & Verleger, 1986; Verleger & Möcks, 1987; Verleger, Paulick, Möcks, Smith, & Keller, 2013). By temporal overlap, we refer to time ranges in which two or more factors are concurrently activated. This concurrent activation is reflected by non-zero factor loadings of the respective time points on all involved factors. These cross-loadings are a major challenge for all rotation methods because they interfere with their optimization criterion to strive for factor loading matrices with simple structure, that is, with minimal cross-loadings (Browne, 2001; Mulaik, 2010). Hence, when the true population pattern deviates from the optimum that a rotation method is aiming for (i.e., there are high cross-loadings, see Asparouhov & Muthen, 2009), then the method will provide inflated factor correlation estimates and deflated estimates of the cross-loadings (De Winter, Dodou, & Wieringa, 2009; De Winter & Dodou, 2012; De Winter & Dodou, 2016; Sass & Schmitt, 2010; Schmitt & Sass, 2011), which in turn will lead to variance misallocation.

We believe that the high prevalence of slow-wave components makes ERP data especially prone to this problem as they overlap with almost all other factors in an EFA (Verleger & Möcks, 1987). A consequence of this is, that the bias could be avoided by applying rotation criteria that are more suitable for situations with high temporal overlap. For instance, there is some evidence that oblique *Geomin* rotation (Yates, 1987) performs better in the presence of substantial cross-loadings than other rotation methods (Schmitt & Sass, 2011; Asparouhov & Muthen, 2009, but see also Dien, 2010a). Alternatively, one can try to modify rotation criteria by considering likely time courses of ERP components (Beauducel, 2017).

To sum up, in this section we explained that variance misallocation is due to biases in the factor loadings and we showed that biased factor covariance parameters or cross-loadings increase the risk for the occurence of these biases. With regard to the former, we used a decomposition of the factor covariance to show that substantial inter-factor covariances are likely to be expected for ERP data due to topography and condition effects. An important consequence of this is that - in general - the usage of orthogonal rotation methods should be avoided. Furthermore, we showed why temporal overlap between factors, or cross-loadings, poses a challenge for currently available rotation methods.

4 A Simulation Study

In the following, we want to investigate the suitability of our theoretical results in a Monte Carlo simulation. We used a Monte Carlo approach here because it allowed us to examine the occurrence or non-occurence of variance misallocation in a range of different conditions defined by the size of the factor correlation, the temporal overlap between factors, and the topographic overlap between the factors. Recent studies evaluating PCA for ERP data have either used real data sets (e.g., Kayser & Tenke,

2003) or data sets simulated from virtual head models to which real EEG noise was added (e.g., Dien et al., 2005). These studies were extremely valuable for assessing the usefulness of the method (e.g., in terms of the interpretability of the solution) under realistic data situations. However, we believe that Monte Carlo studies can nicely complement these efforts, as they allow us to evaluate the statistical properties of EFA for ERP data.

In our simulation, we varied the factor correlation across participants, the temporal overlap of the time courses, the topographic overlap of the factors, and the number of participants per sample. Furthermore, we compared the performance of three rotation methods (Varimax, Promax and Geomin) with regard to the correct estimation of the factor loadings. Based on our considerations, we expected that - especially in the presence of topographic overlap - orthogonal factor rotation (i.e., Varimax) will result in biased factor loading estimates even if the factors are actually uncorrelated across participants. This bias should be substantially weaker without topographic overlap and it should not occur for the oblique Promax and Geomin rotations (Dien, 1998; Dien et al., 2005; Dien, 2010a). In addition, we expected that a differentiable bias occurs that increases only as a function of the temporal overlap. We expected that Geomin rotation can handle substantial cross-loadings better than Promax (Schmitt & Sass, 2011; Asparouhov & Muthen, 2009).

On an exploratory basis, finally, we also compared two different EFA methods (MLFA and PCA) for estimating the initial (unrotated) solution. This was done as it is typically claimed that both estimation methods yield highly similar estimates for ERP-like data (Dien et al., 2005), and we wanted to provide empirical evidence for this assumption.

4.1 Method

4.1.1 Simulation Model. Our simulation model was designed to be appropriate for investigating a variety of data situations typical of ERP analyses. We systematically varied the parameters *sample size* (2), *temporal overlap* (3), *topographic* *overlap* (2), and *population factor correlation* (2). Each sample consisted of a matrix T with 200 sampling points in the columns spread over an epoch of 450 ms and $2 \cdot 10 \cdot N$ rows (emulating ERPs from N participants recorded at 10 electrode sites in two conditions). For illustration purposes, one might think of placing 10 electrodes over the central line where *electrode 1* is the most anterior and *electrode 10* is the most posterior electrode. The sample size N was 20 or 40, representing typical sample sizes in ERP experiments. The samples were drawn from a matrix-variate normal distribution (e.g., Gupta, 2000). That is, $T \sim N(M, V, \Sigma)$ where M is the $2 \cdot 10 \cdot N \times 200$ matrix of expected time courses for each participant at each electrode site and in each condition, and V and Σ are the covariance matrices between the rows and the sampling points, respectively. The row covariance matrix V was an identity matrix of size $2 \cdot 10 \cdot N$. Note that this procedure resulted in *independent* samples, that is, data from different conditions and electrodes were not correlated within participants.

The sampling point covariance matrix Σ was derived from the common factor model (Equation 3). We specified two factors in the population factor loading matrix Λ. The loading time courses are depicted in Figure 1 (A to C). The time courses were specified by Gaussian density functions with a standard deviation of 40 ms. The mean (i.e., peak) of the first factor was at 120 ms. In order to vary the amount of temporal overlap between the factors, three latency conditions were investigated with the mean of the second factor at 200 ms (L1), 250 ms (L2), or 300 ms (L3). The loading curves had a peak loading of 0.8 and 1, respectively. The variance of the factors was normalized (i.e., $\varphi_{11} = \varphi_{22} = 1$), and the factors were either mildly correlated ($\varphi_{12} = 0.3$) or uncorrelated ($\varphi_{12} = 0$). That is, in the correlated condition, a participant with a more positive amplitude for the first factor was likely to show a more positive amplitude for the second factor as well. The error covariance matrix $\Theta = \sigma^2_{error} \cdot I_p$ was a diagonal matrix, that is, all error terms were mutually uncorrelated (white noise) and the noise level was constant over the whole simulated epoch. The noise variance (σ^2_{error}) was 0.4, simulating a moderate noise level as might be expected for average ERPs. [3]

[3]For instance, at the electrode and sampling point with the *highest* signal-to-noise ratio, about 16% of the total variance was due to the simulated noise.

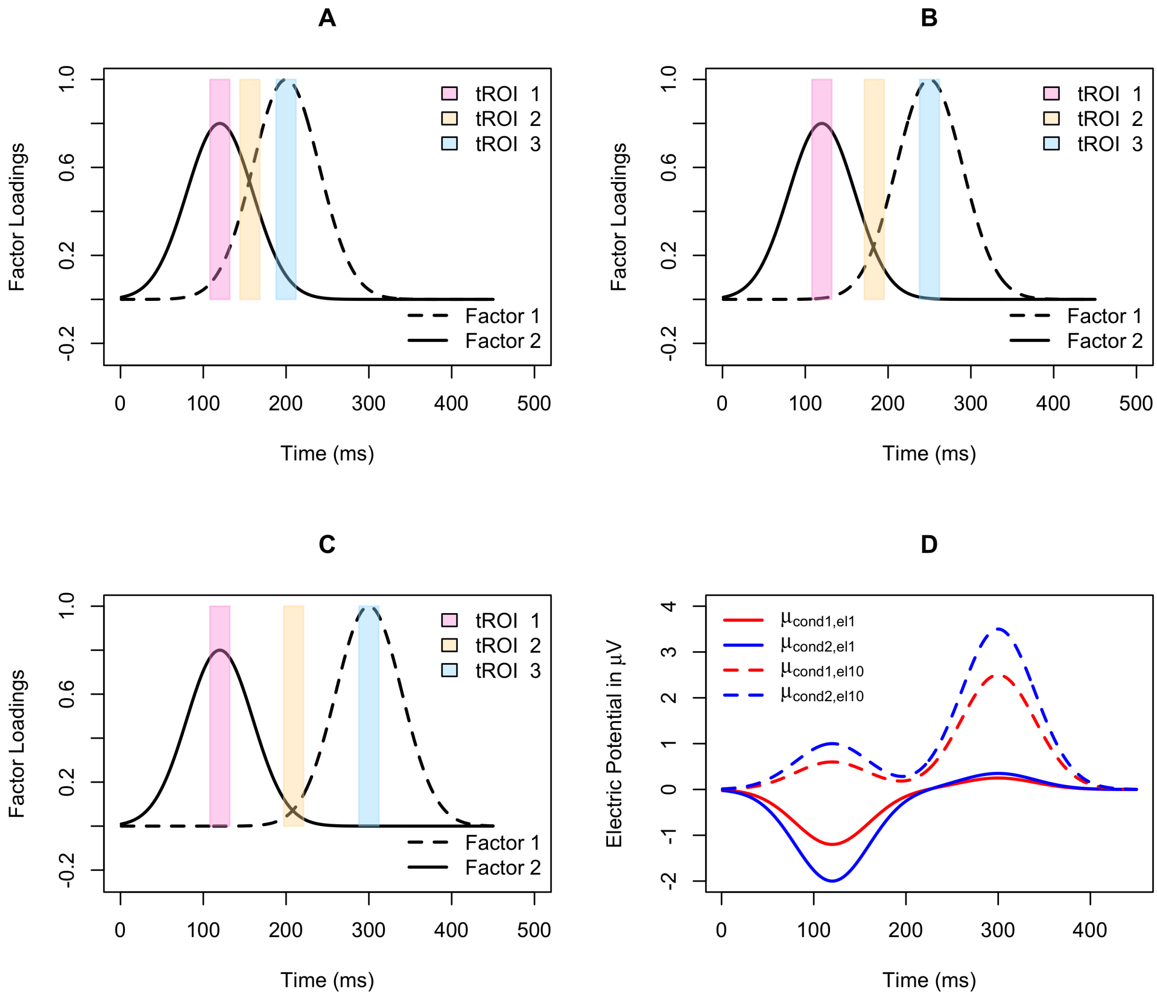

Figure 1. Illustration of the simulation model. A to C: Factors as defined by their loading time courses for the 3 different overlaps (L1 to L3 from A to C). D: Population time course for the condition with the smallest temporal overlap (L3) separately for Electrode 1 (solid lines) and Electrode 10 (dashed lines) in the condition with perfect topographic overlap. The tROIs for the loading recovery analyses are highlighted in colored boxes. Figure available at http://osf.io/xtjkn under a CC-BY 4.0 license.

The expected time courses (i.e., the population grand average ERPs) for all 20 rows per participant can be derived from Equation 2.[4] In a similar manner as Beauducel and Debener (2003), we simulated both topographic variance and condition effects. At the respective topographic maximum, the first factor had expected factor scores of -1.5 (condition 1) or -2.5 (condition 2), and the second factor had expected factor scores of 2.5 (condition 1) or 3.5 (condition 2). With respect to the topographic

[4]The factor loadings have to be multiplied by the mean vector of the factor scores in each condition: $\mu = E(t) = E(\Lambda \cdot \eta + \epsilon) = E(\Lambda \cdot \eta) + E(\epsilon) = E(\Lambda \cdot \eta) = \Lambda \cdot E(\eta)$

overlap, we contrasted the two extremes of perfectly overlapping and non-overlapping (orthogonal) factors. The first factor had its maximum peak at electrode 1 ('anterior' distribution). In the conditions with perfect topographic overlap, the second factor had its maximum peak at electrode 10 ('posterior' distribution). In the conditions with non-overlapping topographies, the second factor had its maximum peak at electrodes 5 and 6 ('central' distribution). Following the principles of the topographic component model (e.g., Achim & Bouchard, 1997; Möcks, 1988b), the remaining factor scores were calculated by multiplying the expected factor scores with topographic weights. The weights were 1 at the maximum peak and linearly decreased towards values of -0.5 (factor 1) and 0.1 (factor 2) at the other extreme of the topography. This procedure resulted in plausible topographies in the sense that the condition effects were stronger at electrode sites at which the factor was stronger and were reversed if the polarity of the factor was reversed at an electrode site. An example for the resulting ERPs at two electrodes is provided in Figure 1D. The expected condition effects and topographies were kept constant across participants.

4.1.2 Simulation Procedure. Simulations were conducted in *R* (Version 3.3.1, R Core Team, 2017). All scripts for simulations and analyses are available at http://osf.io/xtjkn. In each condition, 1000 samples were generated using the package *LaplacesDemon* (Statisticat & LLC., 2016). For each sample, the correlation matrix was calculated and then *smoothed* to ensure positive semi-definiteness in spite of the fact that there were few observations for the high number of variables. For this purpose, a principal component smoothing was applied as it is implemented in the *psych* package (Revelle, 2016; Wothke, 1993). That is, the eigenvalues of the correlation matrix were calculated such that eigenvalues smaller than 10^{-12} were fixed to 10^{-10}, and the eigenvalues were rescaled to add up to the number of variables again. The smoothed correlation matrix was calculated from the smoothed eigenvalues and was then rescaled using the original sample standard deviations. This resulted in a smoothed covariance matrix which was used in all succeeding steps.

In each sample, model parameters were estimated with PCA and MLFA. The same

PCA procedure was applied as in the ERP PCA Toolkit (Dien, 2010a). For MLFA, the algorithm by Lawley and Maxwell (1971) was used as it is implemented in the *psych* package (Revelle, 2016). The number of factors to be extracted was determined by parallel analysis (Horn, 1965). The resulting initial loading matrix was subjected to three different rotations: a *Varimax* rotation with Kaiser-normalization (Kaiser, 1958, 1959), a *Promax* rotation with Kaiser-normalization ($\kappa = 4$) (Hendrickson & White, 1964) an oblique *Geomin* rotation (Yates, 1987). All rotation procedures were conducted using the package *GPArotation* (Bernaards & I. Jennrich, 2005).

4.1.3 Dependent Measures.

4.1.3.1 Performance of Parallel Analysis based on PCA versus MLFA. We investigated the general performance of the parallel analysis and explored whether there were performance differences between parallel analysis based on the factors from PCA versus an MLFA. We classified the results of parallel analysis as *correct* when it suggested 2 factors and as *under*- or *overextracted* when parallel analysis suggested too few or too many factors, respectively. Our dependent measure was the relative proportion of *correct*, *under*-, and *overextracted* parallel analysis solutions in each condition.

4.1.3.2 Recovery of Model Parameters. To evaluate the performance of the estimation methods, we computed measures of the overall model fit, and of the recovery of the population parameters. As a measure of global model fit, we calculated the standardized root mean residual ($SRMR$) between the observed and model-implied covariance matrices in each sample (e.g., Brown, 2014, p.70). The overall model fit was calculated directly from the initial solution as it does not depend on the rotation technique due to rotational indeterminacy (e.g., Mulaik, 2010). The recovery of the model parameters (i.e., factor loadings and factor correlations) was assessed in terms of accuracy and stability of the estimates by calculating the mean and standard deviation across all samples per simulation condition. The deviation between the population values and their estimates was quantified as the absolute bias for each parameter:

$$Bias_{\Lambda} = \hat{\Lambda} - \Lambda \tag{6}$$

$$Bias_{\varphi} = \hat{\varphi}_{12} - \varphi_{12} \tag{7}$$

Here, the hat symbols denote the average estimated sample values (in contrast to the population values), and φ_{12} is the factor correlation. In order to avoid distortions due to the under- or overextraction of factors, samples were considered only if parallel analysis yielded the correct number of factors (i.e., $m = 2$). To summarize our findings more comprehensively, we specified three time ranges of interest (tROIs) and averaged the bias of the factor loadings within each tROI. Two tROIs were centered around the two peaks (± 6 sampling points). The other tROI was located around the intersection of the loading curves (i.e., 6 sampling points before and after the crossing point; Fig. 1).

We are aware that the amount of variance misallocation has been quantified by different measures in the literature. For instance, some studies focused on the amount of false positive significance tests (Wood & McCarthy, 1984; Dien et al., 2005). We quantified variance misallocation by two measures. First, we calculated the correlations between the average estimated factor loadings and the population factor loadings ($r_{\lambda\hat{\lambda}}$) separately for each factor. As in other simulation studies (De Winter & Dodou, 2016; Dien, 1998), this correlation served as a scale-independent measure of similarity between the population loading patterns (i.e., time courses) and their estimates. As variance misallocation is a consequence of biased factor loading estimates, similarity between the population loadings and their estimates is sufficient to indicate any *risk* of variance misallocation. Second, in order to evaluate the impact of the biases on statistical inferences more directly, we calculated the bias in the effect size estimates. That is, Cohen's d ($\hat{\delta}$) was calculated from the estimated factor scores using the classic formula (Cohen, 1962) and the bias was calculated as:

$$Bias_{\delta} = \hat{\delta} - \delta \tag{8}$$

Here $\hat{\delta}$ denotes the estimated effect size and δ denotes the population effect. As δ was equal to ±1 for both factors, this measure directly reflects the *relative* bias in our simulations. The bias was calculated at the electrode site of the respective topographic maximum of each factor (where the effect was maximal as well). It depends on the specific data situation (e.g., power) in how far biased factor loading estimates will affect the number of significant test results (Beauducel & Debener, 2003). Addressing this problem to some extent, our measure reflects direction and severity of the bias independent of the number of observations.

4.2 Results

Under all investigated conditions, Parallel Analysis suggested to extract 2 factors in *all* samples – no matter whether it was based on PCA or MLFA. The overall model fit as reflected by the SRMR ranged from .02 to .03 ($M = 0.03$), indicating reasonable model fit for all investigated conditions (Tab. 1 & 2). The model fit was better for larger sample sizes (N_{20}: $M = 0.02$; N_{40}: $M = 0.03$). The choice of estimation method and varying temporal or topographic factor overlap did not affect the SRMR.

The average MLFA-estimated factor loadings in comparison with the population loadings for an example condition ($N = 40$) are shown in Figure 2 and Figure 3 for factors with and without topographic overlap, respectively. These conditions were quite representative of the general trends we observed in our simulations. Differences between PCA and MLFA were negligible (see Tab. 1, 2, 3, & 4). The factor loadings were generally overestimated (Tab. 3 & 4). However, the population time courses were sufficiently recognized which is reflected by a correct representation of the peak latencies of the two factors (Fig. 2 & 3) and the high correlations ($>.90$) between the estimated and population loadings (Tab. 1 & 2). The most apparent distortions occurred for Varimax rotation in conditions with topographic overlap where it yielded a second, smaller peak in the time range around the peak of the respective other factor (where zero-loadings would have been expected). Notably, these spurious cross-loadings did not occur when the within-participant factor correlation was zero *and* the

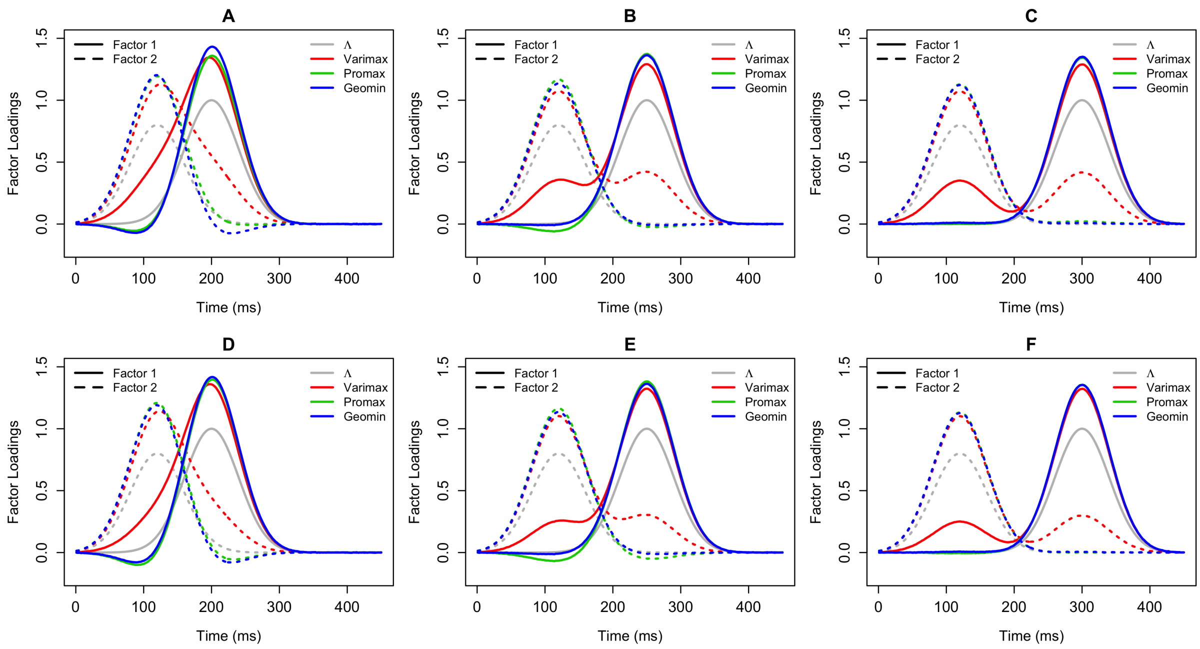

Figure 2. Average MLFA-estimated (colored) and population (Λ, gray) factor loadings of the two factors with topographic overlap (solid vs. dashed lines) for sample size $N = 40$ as a function of temporal overlap (L1 to L3: A to C; D to F) and within-person correlation in the population (A to C: correlated, D to F: uncorrelated). Different colors represent the different rotation methods. Figure available at http://osf.io/xtjkn under a CC-BY 4.0 license.

topographies were orthogonal, but they did occur, when the factors were correlated within participants – even in the absence of topographic overlap (Fig. 3, upper panels).

The obliquely rotated factor loadings where overall much closer to the population time courses than the Varimax rotated loadings. The only exception occurred for the conditions without topographic overlap but with substantial temporal overlap (L1 & L2), where Varimax rotated loadings recovered the population loadings slightly better. This pattern was also reflected by the correlations between the estimated and the population loadings (Tab. 1 & 2) where Varimax rotated loadings yielded lower correlations than obliquely rotated loadings, except for the conditions without topographic overlap but with temporal overlap. However, the benefits of Varimax rotation in these isolated conditions were much smaller than the losses in conditions with topographic overlap.

Performance differences between the oblique rotation methods were much smaller. Both Promax and Geomin perfectly recovered the population time courses in the

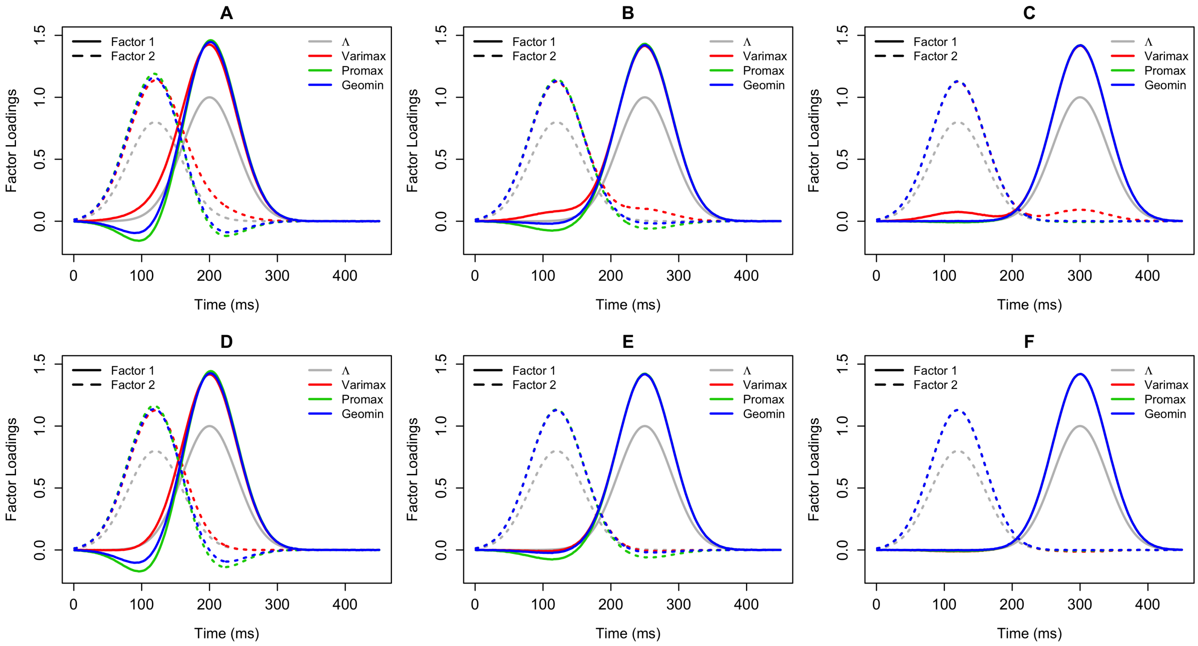

Figure 3. Average MLFA-estimated (colored) and population (Λ, gray) factor loadings of the two factors without topographic overlap (solid vs. dashed lines) for sample size $N = 40$ as a function of temporal overlap (L1 to L3: A to C; D to F) and within-person correlation in the population (A to C: correlated, D to F: uncorrelated). Different colors represent the different rotation methods. Figure available at http://osf.io/xtjkn under a CC-BY 4.0 license.

conditions without temporal overlap (L3). However, when there was temporal overlap (L1, L2), both oblique rotations tended to underestimate the population loadings in regions of temporal overlap (see Fig. 2 & 3). Neither of the oblique rotations performed unequivocally better: Geomin performed slightly better for medium overlap (L2) but for high temporal overlap (L1), Promax was slightly superior in the presence of topographic overlap. This pattern was also confirmed by the correlations between the estimated and the population loadings (Tab. 1 & 2). Finally, it should be mentioned that the factor loading estimates were very stable across samples despite the low sample size (see Appendix B & C). Geomin rotated loadings were slightly more stable than Varimax or Promax rotated loadings, especially for cross-loadings (i.e., F2 in tROI1 and F1 in tROI3, respectively). In addition, we observed that Varimax rotated loadings were more stable than Promax rotated loadings around the first peak (i.e., F1 in tROI1).

Like for the factor loadings, our simulations revealed systematic biases in the estimates of the factor correlation (φ_{12}, Tab. 3 & 4). The relative size of these biases

Table 1

Global fit and time course recovery as a function of simulation condition for conditions with overlapping topographies

					$r_{\lambda\hat{\lambda}}$					
					Varimax		Promax		Geomin	
Estimation	φ_{12}	N	Overlap	$SRMR$	F1	F2	F1	F2	F1	F2
PCA	0.3	20	L1	0.0331	0.9399	0.9663	0.9997	0.9969	0.9939	0.9956
			L2	0.0311	0.9242	0.9630	0.9998	0.9992	1.0000	1.0000
			L3	0.0312	0.9235	0.9643	0.9998	1.0000	1.0000	1.0000
		40	L1	0.0234	0.9405	0.9663	0.9995	0.9967	0.9937	0.9954
			L2	0.0220	0.9235	0.9632	0.9997	0.9991	1.0000	1.0000
			L3	0.0221	0.9237	0.9645	0.9999	1.0000	1.0000	1.0000
	0.0	20	L1	0.0332	0.9681	0.9833	0.9960	0.9924	0.9931	0.9947
			L2	0.0312	0.9635	0.9827	0.9991	0.9989	0.9999	0.9999
			L3	0.0312	0.9646	0.9835	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0235	0.9685	0.9833	0.9957	0.9921	0.9931	0.9947
			L2	0.0221	0.9635	0.9828	0.9991	0.9988	0.9999	0.9999
			L3	0.0221	0.9648	0.9837	1.0000	1.0000	1.0000	1.0000
MLFA	0.3	20	L1	0.0331	0.9376	0.9653	0.9998	0.9970	0.9939	0.9956
			L2	0.0310	0.9221	0.9622	0.9998	0.9992	1.0000	1.0000
			L3	0.0311	0.9215	0.9636	0.9998	1.0000	1.0000	1.0000
		40	L1	0.0234	0.9382	0.9652	0.9997	0.9967	0.9938	0.9955
			L2	0.0219	0.9214	0.9624	0.9997	0.9991	1.0000	1.0000
			L3	0.0220	0.9217	0.9638	0.9998	1.0000	1.0000	1.0000
	0.0	20	L1	0.0332	0.9668	0.9827	0.9963	0.9925	0.9931	0.9948
			L2	0.0311	0.9626	0.9824	0.9991	0.9989	0.9999	0.9999
			L3	0.0311	0.9638	0.9832	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0235	0.9671	0.9827	0.9959	0.9924	0.9931	0.9947
			L2	0.0220	0.9626	0.9824	0.9991	0.9988	0.9999	0.9999
			L3	0.0220	0.9640	0.9834	1.0000	1.0000	1.0000	1.0000

Note. φ_{12} = population factor correlation across participants, N = sample size, $r_{\lambda\hat{\lambda}}$ = correlation between average estimated loadings and population loadings, $F\cdot$ = factor.

was remarkable, exceeding 100% relative to the population value of 0.3 in some conditions. The pattern of these biases depended heavily on the topographic overlap: In conditions with topographic overlap, the factor correlation was generally overestimated and differences between Promax and Geomin were negligible. In conditions without topographic overlap, the biases were generally smaller but both oblique rotation methods still tended to overestimate the factor correlation when the temporal overlap was high (L1). In addition, non-zero population correlations where underestimated for the lower temporal overlaps (L2, L3) whereas the estimates were nearly unbiased for zero population correlations. This pattern was slightly more pronounced for Promax

Table 2

Global fit and time course recovery as a function of simulation condition for conditions with orthogonal topographies

					$r_{\lambda\hat{\lambda}}$					
					Varimax		Promax		Geomin	
Estimation	φ_{12}	N	Overlap	*SRMR*	F1	F2	F1	F2	F1	F2
PCA	0.3	20	L1	0.0333	0.9967	0.9988	0.9875	0.9861	0.9914	0.9932
			L2	0.0312	0.9968	0.9988	0.9987	0.9987	0.9999	0.9999
			L3	0.0310	0.9971	0.9989	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0236	0.9964	0.9988	0.9875	0.9859	0.9915	0.9932
			L2	0.0221	0.9968	0.9988	0.9987	0.9987	0.9999	0.9999
			L3	0.0219	0.9971	0.9989	1.0000	1.0000	1.0000	1.0000
	0.0	20	L1	0.0335	1.0000	0.9999	0.9840	0.9838	0.9905	0.9921
			L2	0.0313	1.0000	1.0000	0.9987	0.9986	0.9998	0.9998
			L3	0.0310	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0237	1.0000	0.9999	0.9839	0.9837	0.9906	0.9922
			L2	0.0222	1.0000	1.0000	0.9986	0.9986	0.9998	0.9998
			L3	0.0220	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
MLFA	0.3	20	L1	0.0333	0.9965	0.9987	0.9877	0.9861	0.9914	0.9932
			L2	0.0311	0.9967	0.9988	0.9987	0.9987	0.9999	0.9999
			L3	0.0309	0.9970	0.9988	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0235	0.9962	0.9987	0.9876	0.9860	0.9915	0.9933
			L2	0.0220	0.9967	0.9988	0.9987	0.9987	0.9999	0.9999
			L3	0.0218	0.9971	0.9989	1.0000	1.0000	1.0000	1.0000
	0.0	20	L1	0.0334	1.0000	0.9999	0.9841	0.9838	0.9905	0.9922
			L2	0.0313	1.0000	1.0000	0.9987	0.9986	0.9998	0.9998
			L3	0.0309	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		40	L1	0.0237	1.0000	0.9999	0.9840	0.9837	0.9906	0.9922
			L2	0.0221	1.0000	1.0000	0.9986	0.9986	0.9998	0.9998
			L3	0.0219	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000

Note. φ_{12} = population factor correlation across participants, N = sample size, $r_{\lambda\hat{\lambda}}$ = correlation between average estimated loadings and population loadings, $F\cdot$ = factor.

rotated solutions than for Geomin rotated solutions.

The relative biases of the effect size estimates (Tab. 5) corresponded very well to the biases in the factor loadings and factor correlations. That is, the biases in the effect size estimates tended to be larger in conditions where the EFA parameters were strongly biased. Overall, the relative biases for Varimax rotated solutions were stronger than for Promax and Geomin rotated solutions – exceeding 10 % in some conditions. Varimax yielded more biased effect sizes in the presence of topographic overlap and/or non-zero across-participant correlation but also when the temporal overlap was high. The bias of the effect sizes for Promax and Geomin solutions varied mainly as a

Table 3

Bias in factor loading and factor correlation estimates for conditions with topographic overlap

				$Bias_\Lambda$ (tROI1)						$Bias_\Lambda$ (tROI2)						$Bias_\Lambda$ (tROI3)						$Bias_\varphi$	
				Varimax		Promax		Geomin		Varimax		Promax		Geomin		Varimax		Promax		Geomin		Promax	Geomin
Estimation	φ_{12}	N	Overlap	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2		
PCA	0.3	20	L1	0.33	0.39	0.39	-0.05	0.40	-0.07	0.40	0.40	0.26	0.12	0.22	0.14	0.44	0.34	0.03	0.35	-0.07	0.42	0.35	0.40
			L2	0.27	0.34	0.36	-0.06	0.33	-0.01	0.19	0.18	0.11	0.07	0.10	0.09	0.40	0.29	-0.02	0.37	-0.01	0.36	0.33	0.29
			L3	0.27	0.34	0.32	0.00	0.32	0.01	0.06	0.06	0.03	0.03	0.03	0.03	0.41	0.29	0.02	0.34	0.01	0.35	0.27	0.27
		40	L1	0.33	0.39	0.39	-0.05	0.40	-0.07	0.40	0.40	0.26	0.12	0.22	0.14	0.43	0.35	0.02	0.36	-0.07	0.43	0.35	0.40
			L2	0.28	0.34	0.36	-0.06	0.33	-0.01	0.19	0.18	0.11	0.07	0.11	0.08	0.41	0.29	-0.03	0.37	-0.01	0.36	0.33	0.30
			L3	0.27	0.34	0.33	0.00	0.32	0.01	0.06	0.06	0.03	0.03	0.03	0.03	0.41	0.29	0.02	0.34	0.01	0.35	0.27	0.27
	0.0	20	L1	0.34	0.29	0.40	-0.11	0.38	-0.08	0.36	0.34	0.24	0.10	0.21	0.13	0.33	0.36	-0.04	0.39	-0.08	0.41	0.59	0.59
			L2	0.30	0.25	0.36	-0.07	0.33	-0.01	0.17	0.16	0.10	0.07	0.10	0.08	0.29	0.32	-0.05	0.38	-0.01	0.36	0.51	0.45
			L3	0.30	0.24	0.33	-0.00	0.32	0.01	0.05	0.05	0.03	0.03	0.03	0.03	0.29	0.32	0.00	0.35	0.01	0.35	0.43	0.42
		40	L1	0.34	0.29	0.40	-0.11	0.39	-0.08	0.36	0.35	0.24	0.10	0.21	0.13	0.33	0.36	-0.05	0.39	-0.08	0.41	0.59	0.59
			L2	0.30	0.25	0.36	-0.07	0.33	-0.01	0.17	0.16	0.10	0.07	0.10	0.08	0.29	0.32	-0.05	0.38	-0.01	0.36	0.51	0.45
			L3	0.30	0.24	0.33	-0.01	0.33	0.01	0.05	0.05	0.03	0.03	0.03	0.03	0.29	0.32	0.00	0.35	0.01	0.35	0.43	0.42
MLFA	0.3	20	L1	0.32	0.39	0.38	-0.05	0.39	-0.07	0.40	0.40	0.26	0.12	0.22	0.14	0.44	0.33	0.04	0.34	-0.07	0.42	0.35	0.41
			L2	0.27	0.35	0.36	-0.06	0.32	-0.01	0.18	0.18	0.11	0.07	0.10	0.08	0.41	0.29	-0.02	0.36	-0.01	0.35	0.33	0.30
			L3	0.26	0.34	0.32	0.00	0.31	0.01	0.06	0.06	0.03	0.03	0.03	0.03	0.41	0.28	0.02	0.33	0.01	0.34	0.27	0.27
		40	L1	0.32	0.39	0.38	-0.05	0.39	-0.07	0.40	0.40	0.26	0.12	0.22	0.14	0.44	0.34	0.03	0.35	-0.07	0.42	0.36	0.41
			L2	0.27	0.35	0.36	-0.06	0.33	-0.01	0.19	0.18	0.11	0.07	0.10	0.08	0.41	0.29	-0.02	0.36	-0.01	0.35	0.34	0.30
			L3	0.27	0.34	0.32	0.00	0.32	0.01	0.06	0.06	0.03	0.03	0.03	0.03	0.41	0.28	0.02	0.33	0.01	0.34	0.28	0.28
	0.0	20	L1	0.34	0.29	0.40	-0.11	0.38	-0.08	0.36	0.34	0.23	0.10	0.21	0.13	0.34	0.35	-0.04	0.38	-0.08	0.40	0.59	0.59
			L2	0.29	0.25	0.35	-0.07	0.33	-0.01	0.17	0.16	0.10	0.07	0.10	0.08	0.29	0.32	-0.05	0.37	-0.01	0.36	0.51	0.45
			L3	0.29	0.25	0.32	-0.00	0.32	0.01	0.05	0.05	0.03	0.03	0.03	0.03	0.29	0.31	0.00	0.35	0.01	0.34	0.43	0.42
		40	L1	0.34	0.30	0.40	-0.11	0.38	-0.08	0.36	0.34	0.23	0.10	0.21	0.13	0.33	0.35	-0.05	0.38	-0.08	0.41	0.60	0.59
			L2	0.30	0.25	0.36	-0.07	0.33	-0.01	0.17	0.16	0.10	0.07	0.10	0.08	0.30	0.32	-0.05	0.37	-0.01	0.35	0.51	0.45
			L3	0.30	0.24	0.32	-0.00	0.32	0.01	0.05	0.05	0.03	0.03	0.03	0.03	0.29	0.31	0.00	0.35	0.01	0.34	0.43	0.42

Note. φ_{12} = population factor correlation across participants, N = sample size, $Bias_\Lambda$ $(\cdot)$ = average difference between the estimated and population factor (averaged within the respective tROI), $Bias_\varphi$ = average difference between the estimated and the population factor correlation. Empirical standard errors are presented in Appendix B.

Table 4

Bias in factor loading and factor correlation estimates for conditions with orthogonal topographies

				$Bias_\Lambda$ (tROI1)						$Bias_\Lambda$ (tROI2)						$Bias_\Lambda$ (tROI3)						$Bias_\varphi$	
				Varimax		Promax		Geomin		Varimax		Promax		Geomin		Varimax		Promax		Geomin		Promax	Geomin
Estimation	φ_{12}	N	Overlap	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2		
PCA	0.3	20	L1	0.34	0.13	0.38	-0.17	0.35	-0.10	0.28	0.27	0.19	0.09	0.18	0.13	0.14	0.42	-0.13	0.45	-0.10	0.44	0.13	0.06
		20	L2	0.32	0.07	0.34	-0.08	0.33	-0.02	0.13	0.12	0.09	0.08	0.10	0.10	0.09	0.41	-0.06	0.43	-0.02	0.42	-0.06	-0.14
		20	L3	0.32	0.07	0.33	-0.01	0.33	0.00	0.04	0.04	0.03	0.03	0.03	0.03	0.09	0.41	-0.01	0.41	0.00	0.41	-0.16	-0.17
		40	L1	0.34	0.13	0.38	-0.17	0.35	-0.10	0.28	0.27	0.19	0.09	0.18	0.13	0.14	0.42	-0.13	0.45	-0.09	0.44	0.14	0.06
		40	L2	0.33	0.07	0.34	-0.08	0.33	-0.02	0.13	0.12	0.10	0.08	0.10	0.09	0.09	0.41	-0.06	0.43	-0.02	0.42	-0.06	-0.14
		40	L3	0.33	0.07	0.33	-0.01	0.33	0.00	0.04	0.04	0.03	0.03	0.03	0.03	0.09	0.41	-0.01	0.41	0.00	0.41	-0.16	-0.17
	0.0	20	L1	0.32	0.05	0.36	-0.19	0.33	-0.11	0.22	0.21	0.16	0.07	0.17	0.11	0.04	0.41	-0.15	0.43	-0.10	0.42	0.34	0.24
		20	L2	0.32	-0.01	0.33	-0.08	0.32	-0.02	0.10	0.10	0.09	0.08	0.10	0.09	-0.01	0.41	-0.06	0.42	-0.02	0.41	0.10	0.02
		20	L3	0.33	-0.01	0.33	-0.01	0.33	-0.00	0.03	0.03	0.03	0.03	0.03	0.03	-0.01	0.41	-0.01	0.41	-0.00	0.41	-0.01	-0.01
		40	L1	0.33	0.05	0.36	-0.19	0.33	-0.11	0.23	0.21	0.16	0.07	0.17	0.11	0.04	0.41	-0.15	0.43	-0.10	0.42	0.34	0.24
		40	L2	0.33	-0.01	0.33	-0.08	0.33	-0.02	0.10	0.10	0.09	0.08	0.10	0.09	-0.01	0.41	-0.06	0.42	-0.02	0.41	0.10	0.02
		40	L3	0.33	-0.01	0.33	-0.01	0.33	-0.00	0.03	0.03	0.03	0.03	0.03	0.03	-0.01	0.41	-0.01	0.41	-0.00	0.41	-0.01	-0.02
MLFA	0.3	20	L1	0.33	0.13	0.38	-0.17	0.35	-0.10	0.27	0.27	0.18	0.09	0.18	0.13	0.14	0.42	-0.13	0.44	-0.10	0.43	0.14	0.07
		20	L2	0.32	0.07	0.34	-0.08	0.32	-0.02	0.12	0.12	0.09	0.08	0.10	0.09	0.10	0.41	-0.06	0.42	-0.02	0.41	-0.06	-0.13
		20	L3	0.32	0.07	0.32	-0.01	0.32	0.00	0.04	0.04	0.03	0.03	0.03	0.03	0.09	0.40	-0.01	0.41	0.00	0.41	-0.16	-0.17
		40	L1	0.33	0.13	0.38	-0.17	0.35	-0.10	0.28	0.27	0.18	0.09	0.18	0.13	0.15	0.42	-0.13	0.45	-0.09	0.43	0.14	0.07
		40	L2	0.32	0.07	0.34	-0.08	0.32	-0.02	0.13	0.12	0.09	0.08	0.10	0.09	0.10	0.41	-0.06	0.42	-0.02	0.41	-0.06	-0.13
		40	L3	0.32	0.07	0.32	-0.01	0.32	0.00	0.04	0.04	0.03	0.03	0.03	0.03	0.09	0.41	-0.01	0.41	0.00	0.41	-0.16	-0.17
	0.0	20	L1	0.32	0.05	0.35	-0.19	0.33	-0.11	0.22	0.21	0.16	0.07	0.16	0.11	0.04	0.41	-0.15	0.43	-0.10	0.41	0.34	0.24
		20	L2	0.32	-0.01	0.32	-0.08	0.32	-0.02	0.10	0.10	0.09	0.08	0.10	0.09	-0.01	0.41	-0.06	0.41	-0.02	0.41	0.10	0.02
		20	L3	0.32	-0.01	0.32	-0.01	0.32	-0.00	0.03	0.03	0.03	0.03	0.03	0.03	-0.01	0.41	-0.01	0.41	-0.00	0.41	-0.01	-0.01
		40	L1	0.32	0.05	0.35	-0.19	0.33	-0.11	0.22	0.21	0.16	0.07	0.16	0.11	0.04	0.41	-0.15	0.43	-0.10	0.41	0.34	0.25
		40	L2	0.32	-0.01	0.33	-0.08	0.32	-0.02	0.10	0.10	0.09	0.08	0.10	0.09	-0.01	0.41	-0.06	0.41	-0.02	0.41	0.10	0.02
		40	L3	0.32	-0.01	0.32	-0.01	0.32	-0.00	0.03	0.03	0.03	0.03	0.03	0.03	-0.01	0.41	-0.01	0.41	-0.00	0.41	-0.01	-0.02

Note. φ_{12} = population factor correlation across participants, N = sample size, $Bias_\Lambda$ $(\cdot)$ = average difference between the estimated and population factor (averaged within the respective tROI), $Bias_\varphi$ = average difference between the estimated and the population factor correlation. Empirical standard errors are presented in Appendix C.

function of the temporal overlap. Notably, in the most favorable conditions (i.e., no temporal or topographic overlap, no across-participants correlation), none of the tested rotation methods substantially biased the effect size estimates.

Table 5

Relative biases of the effect sizes as a function of simulation condition

				Overlapping Topographies						Non-Overlapping Topographies					
				Varimax		Promax		Geomin		Varimax		Promax		Geomin	
Estimation	φ_{12}	N	Overlap	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
PCA	0.3	20	L1	-0.01	-0.16	-0.07	-0.07	-0.08	-0.06	0.03	0.09	-0.07	-0.05	-0.04	-0.03
			L2	0.07	-0.10	-0.04	0.04	-0.02	0.04	0.00	0.04	-0.04	-0.04	-0.02	-0.02
			L3	0.08	-0.13	-0.01	0.00	0.00	0.01	0.01	0.07	-0.01	0.02	-0.01	0.02
		40	L1	0.01	-0.19	-0.06	-0.07	-0.07	-0.07	0.02	0.08	-0.08	-0.06	-0.06	-0.04
			L2	0.07	-0.13	-0.03	0.02	-0.02	0.01	0.01	0.04	-0.04	-0.04	-0.02	-0.01
			L3	0.08	-0.13	0.01	0.01	0.01	0.01	0.03	0.04	0.00	-0.02	0.01	-0.01
	0.0	20	L1	-0.01	-0.16	-0.02	0.02	-0.01	0.04	0.00	0.03	-0.03	-0.03	-0.01	-0.01
			L2	0.00	-0.11	-0.02	0.06	-0.01	0.05	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01
			L3	0.00	-0.14	-0.01	0.01	-0.01	0.00	-0.01	0.02	-0.01	0.02	-0.01	0.02
		40	L1	-0.02	-0.19	-0.03	0.00	-0.02	0.01	-0.01	0.02	-0.04	-0.04	-0.03	-0.02
			L2	0.00	-0.14	-0.02	0.03	-0.01	0.01	-0.01	0.00	-0.02	-0.02	-0.01	-0.01
			L3	0.01	-0.14	0.01	0.01	0.01	0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01
MLFA	0.3	20	L1	-0.01	-0.16	-0.07	-0.07	-0.08	-0.06	0.02	0.09	-0.07	-0.05	-0.05	-0.03
			L2	0.07	-0.10	-0.04	0.04	-0.02	0.04	0.00	0.04	-0.04	-0.04	-0.03	-0.02
			L3	0.07	-0.13	-0.01	0.01	-0.01	0.01	0.01	0.06	-0.01	0.01	-0.01	0.02
		40	L1	0.01	-0.18	-0.06	-0.07	-0.07	-0.06	0.02	0.08	-0.08	-0.06	-0.06	-0.04
			L2	0.07	-0.13	-0.04	0.02	-0.02	0.01	0.01	0.04	-0.04	-0.04	-0.02	-0.01
			L3	0.08	-0.13	0.00	0.01	0.01	0.01	0.03	0.04	0.00	-0.02	0.01	-0.01
	0.0	20	L1	-0.01	-0.16	-0.02	0.03	-0.01	0.04	0.00	0.03	-0.03	-0.03	-0.02	-0.01
			L2	0.00	-0.11	-0.02	0.06	-0.01	0.05	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01
			L3	0.00	-0.14	-0.01	0.01	-0.01	0.01	-0.01	0.02	-0.01	0.02	-0.01	0.02
		40	L1	-0.01	-0.19	-0.03	0.00	-0.02	0.01	-0.01	0.02	-0.04	-0.04	-0.03	-0.02
			L2	0.00	-0.14	-0.02	0.03	-0.01	0.02	-0.01	0.00	-0.02	-0.02	-0.01	-0.01
			L3	0.01	-0.14	0.01	0.01	0.01	0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01

Note. φ_{12} = population factor correlation across participants, N = sample size, $F\cdot$ = factor.

5 Discussion

In the present article, we showed that variance misallocation occurs because of biased factor covariance estimates and because of biased cross-loading estimates caused by the temporal overlap between underlying factors. In a Monte-Carlo study, we then investigated the role of the different sources of factor covariance and the temporal overlap between factors for the occurrence of variance misallocation. We compared the recovery of EFA parameters for several rotation methods and two estimation methods (PCA and MLFA) in situations that are representative of typical ERP data. The results showed that the number of factors to be extracted was recognized with remarkable accuracy, that differences between PCA and MLFA were negligible, and that the recovery of the population factor loadings with respect to the overall time course was quite sufficient across all conditions. These findings support the notion that EFA is a useful method for ERP researchers (see also Kayser & Tenke, 2005). Therefore, before discussing the biases we found, we want to reemphasize that we view them as hints how to improve the method rather than arguments against the use of EFA.

With regard to orthogonal (here Varimax) versus oblique rotation methods, our results showed that characteristic biases occurred only for Varimax but not for Geomin or Promax rotation. Varimax produced spurious cross-loadings reflecting the shared variance of the factors that comes from systematic topographic and condition variance (Appendix A; Dien, 2010a) – even in conditions without temporal overlap (L3). In line with our reasoning, these biases clearly increased as a function of the total factor covariance: When the factors were uncorrelated across participants in the population *and* the factor topographies were orthogonal, Varimax yielded unbiased factor loadings estimates. When either factors were overlapping or factors were correlated across participants, spurious cross-loadings could be observed which were strongest when both sources of covariance differed from zero. The relative biases of the effect sizes confirm that the severity of variance misallocation due to correlated factors by far outweighs the benefits of Varimax rotation in the most favorable conditions (i.e., when the total factor correlation is indeed zero). Considering the facts that only two factors with orthogonal topography are possible for ERP data (Dien, 2010a) and that EFA for ERP data typically yields more than 2 factors, our theoretical and our simulation results suggest that researchers should abandon the use of Varimax rotation when applying temporal EFA to ERP data because it virtually guarantees variance misallocation. We note that in case the total factor covariance is indeed close to zero, the estimates of orthogonal and oblique rotations will be very similar (see Fig. 3F). Hence, there is 'nothing to lose' when regularly using oblique rotation methods (Dien, 1998; Dien et al., 2005; Dien, 2010a).

Independent from the factor correlation, oblique rotations yielded unbiased factor loadings only in absence of temporal overlap (L3), otherwise small spurious cross-loadings of opposite sign were observed that increased with temporal overlap (L1 > L2). In spite of their small size, the spurious cross-loadings clearly resulted in biased effect sizes, at least in conditions with high temporal overlap (L1). Of the two tested oblique rotation methods, Geomin performed better than Promax in medium overlap conditions (L2) and worse in some high overlap conditions (L1 with topographic overlap). Consistent with this ambiguous pattern, previous simulations found advantages for either Promax (Dien, 2010a) or Geomin rotation depending on the tested factor loading pattern (Asparouhov & Muthen, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011). These results are not contradictory because the performance of factor rotation methods depends on the degree to which the true loading pattern matches the optimum of the rotation criterion (Browne, 2001). Consequently, researchers applying EFA to ERP data should be aware that each rotation method makes implicit assumptions about the time courses of the ERP factors and the choice of the rotation method should be guided by the expected factor loading pattern (Schmitt & Sass, 2011). Indeed, researchers can easily assess the potential impact of the rotation criterion on their specific ERP factors by specifying an artificial factor loading matrix (Λ) that represents their beliefs about the expected ERP factors. When this matrix is subjected to factor rotation, the difference between the original and the rotated matrix is a valid indicator of the bias to be expected (Möcks & Verleger, 1986; Verleger & Möcks, 1987).

We note that the small size of the *rotation bias* in our simulations is a feature of the simulated loading pattern and more severe distortions can be demonstrated easily (Beauducel, 2017; Verleger & Möcks, 1987). Therefore, it is important for future research to determine which rotation method is appropriate for which situation. Kayser and Tenke (2003) suggested to use Varimax rotation including *all* factors (i.e., an unrestricted solution) instead of extracting the number of factors as suggested by Parallel Analysis (and then rotating them). However, although additionally extracted noise factors may indeed improve the general time course recovery of the meaningful factors as *temporal* autocorrelations of the noise may be better represented (see Dien, 2018, p. 101, for a similar argument), this approach cannot resolve the problem outlined here that many of the extracted factors are necessarily correlated due to *topographic* overlap[5]. Therefore, we see the questions how many factors should be extracted or how noise correlations should be considered as independent matters from the choice of the rotation method that should be further investigated in future research.

In order to improve existing rotation methods, it may be fruitful to develop rotation criteria that incorporate the tendency of ERP factors to overlap to a large extent. Recently, efforts were made to systematically improve rotation methods with additional information such as the allocation of known effects (Beauducel & Leue, 2015) or to rotate to the closest Gaussian-shaped time course (Beauducel, 2017). In addition, we suggest to explore the use of partial target rotation to utilize information about the expected time courses of the factors (Myers, Ahn, & Jin, 2013; Myers, Jin, Ahn, Celimli, & Zopluoglu, 2015). Specifically, in many situations, it might be possible to provide the rotation with a wide range of fixed zero-loadings representing time ranges in which the respective factor should *not* contribute any activity.

Regarding the factor correlation estimates, we found that they deviated heavily from the population factor correlation across participants - even for oblique rotations. At first glance, this result may be surprising as the population factor loadings were

[5] We confirmed that notion empirically by implementing another simulation condition in which 20 rather than two factors were extracted and then Varimax rotated. However, we did not find any reduction in variance misallocation compared to the other Varimax conditions in our simulation. The results for this additional simulation condition can be found on the OSF (http://osf.io/xtjkn)."

recovered relatively well. However, this finding is in line with our theoretical considerations showing that the estimated factor correlation is the result of *three* sources of covariance. We believe that this result is important for research in which the factor score correlation is substantively interpreted (e.g., Steiner, Barry, & Gonsalvez, 2016). As one does not know which source of variance dominates the formation and interpretation of the factors (Hunt, 1985; Rösler & Manzey, 1981), correlations between the factors scores can be caused by any of these sources. Hence, when the other sources of covariance are not partialled out, it is, strictly speaking, not possible to find a clear, substantive interpretation for the correlation.[6]

A number of modifications to the EFA procedure have been suggested in previous research (see, e.g., Dien & Frishkoff, 2005, for a brief overview), trying to consider the contributions of the different variance sources. For example, Möcks and Verleger (1985) suggested to partial out between-person and topographic variance by subtracting participant and electrode means from the raw signal, effectively treating them a nuisance variance and discarding all information that may be contained in them. Alternatively, separate EFAs per condition could be computed (Barry et al., 2016), resulting in one set of parameter estimates per condition. We note that this approach is especially reasonable when there are systematic *latency* shifts between the conditions. Nevertheless, the resulting factor scores will still vary due to between-person and topographic (between-electrode) differences so that the factor (co-)variance estimates still commingle these two sources of variance. Another potential modification to the temporal EFA procedure is a two-step approach where a temporal EFA is followed up by a spatial EFA on the extracted factor scores (or vice versa; Dien, 2010a; Spencer, Dien, & Donchin, 1999, 2001). An advantage of this method is that factors which were conflated in one dimension might be disentangled in the other dimension. However, this approach lacks a unified statistical model and therefore does not consider factors as unified entities across both temporal and spatial domain. Addressing this issue, *Trilinear models* such as the *topographic components model* (Achim & Bouchard, 1997;

[6]Note that even zero-correlations can be uninformative because they can be the result of summing up a positive correlation from one source of variance and a negative correlation from another one.

Möcks, 1988a, 1988b; Verleger et al., 2013) were suggested which build on the assumption that the *same* temporal factors contribute to the signal at each electrode but with varying weights (see Cong et al., 2015; Field & Graupe, 1991; Harshman & Lundy, 1984; Harshman & Lundy, 1994, for a description of these approaches). Apart from these previous suggestions, the structural equation modeling framework might offer the possibility of modeling condition (and electrode) effects as well as repeated measurement occasions within one unified model (Asparouhov & Muthen, 2009, 2012). Given this multitude of opportunities for future developments, we believe that more research is needed that directly compares these methods to determine which model is best suited both from a statistical and from a biophysiological point of view.

6 Limitations

One weakness of our simulation approach is that the simulated data contained considerable simplifications: Realistic ERP data contain more components (with varying shapes), correlated noise, and many electrode sites instead of only ten. In addition, in real ERP data sets, data from the same participant and data from neighboring electrodes within each participant should be correlated. As long as there is a sufficient number of high-loading sampling points per factor, more factors per se should not substantially degrade EFA solutions (De Winter et al., 2009). Also, the presence of more electrodes should not lead to fundamentally different conclusions. The relative impact of spatial, temporal, and participant-related autocorrelation cannot be evaluated from our simulations and needs to be evaluated in future research. A second short-coming of conducting a Monte Carlo simulations is that we were not able to investigate the effects of variance misallocation on other types of analyses that are specifically applied to EEG/ERP data such as *source localization*. There is some evidence that Promax rotated factors yield better localization results (Dien, Spencer, & Donchin, 2003; Dien, 2010a). With respect to our findings, it is likely that the localization works better the more closely the population factor loadings are recovered.

7 Conclusion

On the basis of a decomposition of the factor covariance and a Monte Carlo simulation, we investigated variance misallocation from the perspective of EFA as a statistical model and compared the performance of PCA and MLFA. For researchers applying temporal EFA for ERP data, our results should clarify three points: 1) Beyond philosophical discussions about whether the brain is orthogonal or not (Dien, 2006; Hunt, 1985; Kayser & Tenke, 2005), Varimax rotation of the temporal factors carries a very high risk of variance misallocation due to the covariance between factors caused by condition effects and topographic overlap. 2) Each factor rotation method makes implicit assumptions about the time courses of the ERP factors that one needs to be aware of. 3) Any correlation involving the factor scores must be interpreted with great caution because the factor scores commingle condition, topographic, and participant variation, except if all but one source of variance are partialled out. Finally, we want to emphasize that - although there is some room for methodological improvements - we see the EFA approach as an important tool in any ERP researcher's toolbox which can provide remarkably useful summaries of ERP data sets.

Appendix A

Derivation of the biased factor correlation estimate for the multiple-electrode case

In the following, we want to generalize our derivation of the factor covariance matrix (cf. introduction section) from the single to the multiple electrode case.

We start again from the general common factor model:

$$T = \Lambda \cdot \eta + E \tag{A.1}$$

Here, T denotes the grand-mean-centered ERP data matrix (with sampling points in the rows and participants, electrodes and conditions in the columns). Λ denotes the factor loadings for all factors and E is a matrix of error terms. Without loss of generality, we assume that the factor scores are also grand-mean-centered (i.e., $E(\eta) = 0$). If we extract m factors for the data set with p sampling points and n observations (where $n = n_{person} \cdot n_{electrode} \cdot n_{condition}$), the matrices would look as follows:

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ t_{21} & t_{22} & t_{23} & \dots & t_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ t_{p1} & t_{p2} & t_{p3} & \dots & t_{pn} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \dots & \dots & \dots & \dots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pm} \end{bmatrix} \cdot \begin{bmatrix} \eta_{11} & \dots & \eta_{1n} \\ \eta_{21} & \dots & \eta_{2n} \\ \dots & \dots & \dots \\ \eta_{m1} & \dots & \eta_{mn} \end{bmatrix} + \begin{bmatrix} e_{11} & \dots & e_{1p} \\ e_{21} & \dots & e_{2p} \\ \dots & \dots & \dots \\ e_{n1} & \dots & e_{np} \end{bmatrix} \tag{A.2}$$

Put simply: Each column in T contains the grand-mean-centered ERP recorded at a specific electrode site in a specific condition from a specific participant (in the following referred to as *observation*).[7] These ERPs are decomposed in a model-predicted part (i.e., the columns in $\Lambda \cdot \eta$) and an error term (i.e., the columns in E). Each column in Λ reflects the time course of a specific factor over the whole epoch and each column in η contains the factor scores of all factors for a specific observation.

[7] The orientation of the matrices may seem unfamiliar because observations are typically in the rows of the data matrix in typical ERP software. We chose this notation because it is quite common in text books on EFA (e.g., Mulaik, 2010).

Assuming we have l predictor variables for the condition effects (i.e., main effects *and* interactions of experimental manipulations) and k effect-coded indicator variables for the electrodes, we can consider the factor scores η as a function of three sources of variance:

$$\eta = \Gamma_{cond} \cdot X_{cond} + \Gamma_{el} \cdot X_{el} + \zeta \tag{A.3}$$

$$\begin{bmatrix} \eta_{11} & \dots & \eta_{1n} \\ \eta_{21} & \dots & \eta_{2n} \\ \dots & \dots & \dots \\ \eta_{m1} & \dots & \eta_{mn} \end{bmatrix} = \begin{bmatrix} \gamma_{11}^{(cond)} & \dots & \gamma_{1l}^{(cond)} \\ \gamma_{21}^{(cond)} & \dots & \gamma_{2l}^{(cond)} \\ \dots & \dots & \dots \\ \gamma_{m1}^{(cond)} & \dots & \gamma_{ml}^{(cond)} \end{bmatrix} \cdot \begin{bmatrix} x_{11}^{(cond)} & \dots & x_{1n}^{(cond)} \\ x_{21}^{(cond)} & \dots & x_{2n}^{(cond)} \\ \dots & \dots & \dots \\ x_{l1}^{(cond)} & \dots & x_{ln}^{(cond)} \end{bmatrix} + \tag{A.4}$$

$$\begin{bmatrix} \gamma_{11}^{(el)} & \dots & \gamma_{1k}^{(el)} \\ \gamma_{21}^{(el)} & \dots & \gamma_{2k}^{(el)} \\ \dots & \dots & \dots \\ \gamma_{m1}^{(el)} & \dots & \gamma_{mk}^{(el)} \end{bmatrix} \cdot \begin{bmatrix} x_{11}^{(el)} & \dots & x_{1n}^{(el)} \\ x_{21}^{(el)} & \dots & x_{2n}^{(el)} \\ \dots & \dots & \dots \\ x_{k1}^{(el)} & \dots & x_{ln}^{(el)} \end{bmatrix} + \begin{bmatrix} \zeta_{11} & \dots & \zeta_{1n} \\ \zeta_{21} & \dots & \zeta_{2n} \\ \dots & \dots & \dots \\ \zeta_{m1} & \dots & \zeta_{mn} \end{bmatrix} \tag{A.5}$$

Similar to the simplified example in the introduction, this equation can be seen as a regression on the level of the latent factor scores. The factor scores of each observation are predicted by condition and electrode position – leaving an observation-specific residual. The elements in the Γ matrices are regression weights, the predictor matrices X contain the predictor variables. For the condition effects, the predictor variables may be categorical or continuous (or a mixture of both). For the electrode effects, the most general formulation is to have an indicator variable for each electrode. The observation-specific residual will then be a mixture of participant-specific variation in the condition effects and the topography.

For the sake of simplicity, we derive the variance decomposition for an additive model although a model including interaction terms would arguably be more appropriate (Möcks, 1988a). However, the resulting equations are much more comprehensible for the additive case. The principle validity of our deliberations should not be affected by this simplification because, technically, the interaction terms are simply additional predictors that add even more sources of (co-)variance to the final decomposition.

The covariance matrix of the *centered* factor scores Φ can be rewritten as the expected value of the crossproduct of the factor scores (see Mulaik, 2010, for an extensive introduction into the basic mathematics of factor analysis)[8]:

$$\Phi = E[\eta \cdot \eta^T] \tag{A.6}$$

$$= E[(\Gamma_{cond} \cdot X_{cond} + \Gamma_{el} \cdot X_{el} + \zeta) \cdot (\Gamma_{cond} \cdot X_{cond} + \Gamma_{el} \cdot X_{el} + \zeta)^T] \tag{A.7}$$

$$= \underbrace{E[(\Gamma_{cond} \cdot X_{cond})(\Gamma_{cond} \cdot X_{cond})^T]}_{\Phi_{cond}} + \underbrace{E[(\Gamma_{el} \cdot X_{el})(\Gamma_{el} \cdot X_{el})^T]}_{\Phi_{el}} \tag{A.8}$$

$$+ \underbrace{E[(\Gamma_{cond} \cdot X_{cond})(\Gamma_{el} \cdot X_{el}^T] + E[(\Gamma_{el} \cdot X_{el})(\Gamma_{cond} \cdot X_{cond})^T]}_{\Phi_{cond,el}} + \underbrace{E[\zeta \cdot \zeta^T]}_{\Phi_{person}} \tag{A.9}$$

That is, the covariance of the factors is the result of covariance due to condition effects, the scalp topography, the correlation between condition and electrode predictors, and person-specific residuals. Note that even if there were no condition effects (i.e., all elements in Γ_{cond} are zero) and the factors were orthogonal within participants (i.e., Φ_{person} is a diagonal matrix), it would be extremely unlikely to get orthogonal factors as there are always topographic distributions across the scalp that overlap to some degree (Dien, 2010a).

[8]Some of the resulting product terms are zero because we assumed that predictors (here, electrodes and conditions) are uncorrelated with the residuals ζ. These zero-terms were left out of the equation.

Appendix B

Empirical standard errors of factor loading and factor correlation estimates for conditions with overlapping topographies

				$Bias_\Lambda$ (tROI1)						$Bias_\Lambda$ (tROI2)						$Bias_\Lambda$ (tROI3)						$Bias_\varphi$	
				Varimax		Promax		Geomin		Varimax		Promax		Geomin		Varimax		Promax		Geomin		Promax	Geomin
Estimation	φ_{12}	N	Overlap	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2		
PCA	0.3	20	L1	0.05	0.08	0.07	0.08	0.04	0.01	0.07	0.08	0.08	0.07	0.03	0.03	0.10	0.06	0.09	0.06	0.01	0.05	0.03	0.02
			L2	0.03	0.04	0.04	0.01	0.04	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.05	0.04	0.02	0.04	0.01	0.04	0.02	0.03
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.03
		40	L1	0.04	0.07	0.06	0.08	0.03	0.01	0.06	0.06	0.07	0.07	0.02	0.02	0.08	0.04	0.07	0.05	0.01	0.03	0.03	0.02
			L2	0.02	0.04	0.03	0.01	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.03	0.01	0.03	0.01	0.03	0.02	0.02
			L3	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.01	0.03	0.02	0.02
	0.0	20	L1	0.04	0.05	0.04	0.03	0.04	0.01	0.05	0.05	0.04	0.03	0.03	0.03	0.06	0.05	0.04	0.04	0.01	0.04	0.02	0.03
			L2	0.03	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.04
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.03	0.04
		40	L1	0.03	0.04	0.03	0.02	0.03	0.01	0.04	0.04	0.02	0.02	0.02	0.02	0.05	0.03	0.03	0.03	0.01	0.03	0.01	0.02
			L2	0.02	0.03	0.03	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.02	0.02
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.02	0.02
MLFA	0.3	20	L1	0.05	0.09	0.07	0.08	0.04	0.01	0.07	0.08	0.08	0.07	0.03	0.03	0.10	0.06	0.09	0.06	0.01	0.05	0.03	0.02
			L2	0.03	0.04	0.04	0.01	0.04	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.05	0.04	0.02	0.04	0.01	0.04	0.02	0.03
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.03
		40	L1	0.04	0.07	0.05	0.07	0.03	0.01	0.06	0.06	0.06	0.06	0.02	0.02	0.08	0.04	0.07	0.05	0.01	0.03	0.03	0.02
			L2	0.03	0.04	0.03	0.01	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.03	0.01	0.03	0.01	0.03	0.02	0.02
			L3	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.01	0.03	0.02	0.02
	0.0	20	L1	0.04	0.05	0.04	0.02	0.04	0.01	0.05	0.05	0.03	0.03	0.03	0.03	0.06	0.05	0.03	0.04	0.01	0.04	0.02	0.03
			L2	0.03	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.04
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.03	0.04
		40	L1	0.03	0.04	0.03	0.03	0.03	0.01	0.04	0.04	0.03	0.03	0.02	0.02	0.05	0.03	0.03	0.03	0.01	0.03	0.02	0.02
			L2	0.02	0.03	0.03	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.02	0.02
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.02	0.02

Note. φ_{12} = population factor correlation across participants, N = sample size, $Bias_\Lambda$ $(\cdot)$ = average difference between the estimated and population factor (averaged within the respective tROI), $Bias_\varphi$ = average difference between the estimated and the population factor correlation.

Appendix C

Empirical standard errors of factor loading and factor correlation estimates for conditions with orthogonal topographies

				$Bias_\Lambda$ (tROI1)						$Bias_\Lambda$ (tROI2)						$Bias_\Lambda$ (tROI3)						$Bias_\varphi$	
				Varimax		Promax		Geomin		Varimax		Promax		Geomin		Varimax		Promax		Geomin		Promax	Geomin
Estimation	φ_{12}	N	Overlap	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2		
PCA	0.3	20	L1	0.04	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.03	0.02	0.03	0.03	0.05	0.04	0.02	0.04	0.01	0.04	0.03	0.03
			L2	0.04	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.04
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.04	0.04
		40	L1	0.03	0.03	0.03	0.01	0.03	0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.04	0.03	0.01	0.03	0.00	0.03	0.02	0.02
			L2	0.02	0.02	0.03	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
	0.0	20	L1	0.04	0.04	0.04	0.01	0.04	0.01	0.03	0.04	0.03	0.02	0.02	0.02	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.04
			L2	0.04	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.04
			L3	0.04	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.04	0.04
		40	L1	0.02	0.03	0.02	0.01	0.02	0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.04	0.03	0.01	0.03	0.00	0.03	0.02	0.03
			L2	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
MLFA	0.3	20	L1	0.04	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.03	0.02	0.03	0.03	0.05	0.04	0.02	0.04	0.01	0.04	0.03	0.03
			L2	0.04	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.04
			L3	0.03	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.04	0.04
		40	L1	0.03	0.03	0.03	0.01	0.03	0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.04	0.03	0.01	0.03	0.00	0.03	0.02	0.02
			L2	0.02	0.02	0.03	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
	0.0	20	L1	0.04	0.04	0.04	0.01	0.04	0.01	0.03	0.04	0.03	0.02	0.03	0.02	0.04	0.04	0.01	0.04	0.01	0.04	0.03	0.04
			L2	0.04	0.03	0.04	0.01	0.04	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.04	0.01	0.04	0.01	0.04	0.04	0.04
			L3	0.04	0.03	0.04	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04	0.01	0.04	0.01	0.04	0.04	0.04
		40	L1	0.02	0.03	0.02	0.01	0.02	0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.04	0.03	0.01	0.03	0.00	0.03	0.02	0.03
			L2	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03
			L3	0.02	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.03	0.03

Note. φ_{12} = population factor correlation across participants, N = sample size, $Bias_\Lambda\ (\cdot)$ = average difference between the estimated and population factor (averaged within the respective tROI), $Bias_\varphi$ = average difference between the estimated and the population factor correlation.

References

Achim, A. & Bouchard, S. (1997). Toward a dynamic topographic components model. *Electroencephalography and Clinical Neurophysiology*, *103*, 381–385. doi:10.1016/S0013-4694(97)96055-0

Achim, A. & Marcantoni, W. (1997). Principal component analysis of event-related potentials: Misallocation of variance revisited. *Psychophysiology*, *34*, 597–606. doi:10.1111/j.1469-8986.1997.tb01746.x

Asparouhov, T. & Muthen, B. (2009). *Exploratory Structural Equation Modeling.* doi:10.1080/10705510903008204

Asparouhov, T. & Muthen, B. (2012). General random effect latent variable modeling : Random subjects, items, contexts, and parameters, 1–27. Retrieved from http://www.statmodel.com/download/NCME12.pdf

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. doi:10.1016/j.ijpsycho.2016.06.003

Beauducel, A. (2017). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific Procrustes-rotation. *Journal of Neuroscience Methods*. doi:10.1016/j.jneumeth.2017.11.011

Beauducel, A. & Debener, S. (2003). Misallocation of variance in event-related potentials: Simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *Journal of Neuroscience Methods*, *124*, 103–112. doi:10.1016/S0165-0270(02)00381-3

Beauducel, A., Debener, S., Brocke, B., & Kayser, J. (2000). On the reliability of augmenting/ reducing: Peak amplitudes and principal component analysis of auditory evoked potentials. *Journal of Psychophysiology*, *14*, 226–240.

Beauducel, A. & Leue, A. (2015). Controlling for experimental effects in event-related potentials by means of principal component rotation. *Journal of Neuroscience Methods*, *239*, 139–147. doi:10.1016/j.jneumeth.2014.10.008

Bernaards, C. A. & I. Jennrich, R. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Boxtel, G. J. M. (1998). Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments, & Computers*, *30*, 87–102. doi:10.3758/BF03209419

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* Guilford Publications.

Browne, M. W. (2001). An overview of analytic rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR360105

Chapman, R. M. & McCrary, J. W. (1995). EP component identification and measurement by principal components analysis. doi:10.1006/brcg.1995.1024

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, *65*, 145–153. doi:10.1037/h0045186

Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., & Ristaniemi, T. (2015). Tensor decomposition of EEG signals: A brief review. *Journal of Neuroscience Methods*, *248*, 59–69. doi:10.1016/j.jneumeth.2015.03.018

De Winter, J. C. F. & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*, 695–710. doi:10.1080/02664763.2011.610445

De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory Factor Analysis with small sample sizes. *Multivariate Behavioral Research*, *44*, 147–181. doi:10.1080/00273170902794206

De Winter, J. C. F. & Dodou, D. (2016). Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations. *Communications in Statistics - Simulation and Computation*, *45*, 299–321. doi:10.1080/03610918.2013.862274

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55. doi:10.1023/A:1022218503558

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clinical Neurophysiology*, *117*, 699–702. doi:10.1016/j.clinph.2005.09.029

Dien, J. (2010a). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi:10.1111/j.1469-8986.2009.00885.x

Dien, J. (2010b). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*, 138–145. doi:10.1016/j.jneumeth.2009.12.009

Dien, J. (2012). Applying Principal Components Analysis to event-related potentials: a tutorial. *Developmental Neuropsychology*, *37*, 497–517. doi:10.1080/87565641.2012.697503

Dien, J. (2018). ERP PCA Toolkit 2.66 Tutorial. College Park: University of Maryland. Retrieved from http://sourceforge.net/projects/erppcatoolkit

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*, 1808–1825. doi:10.1016/j.clinph.2004.11.025

Dien, J. & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–208). Cambridge, MA: MIT Press.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Brain research. Cognitive brain research*, *17*, 637–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14561451

Donchin, E. & Heffley, E. F. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Donchin, E. (1966). A Multivariate Approach to the Analysis of Average Evoked Potentials. *IEEE Transactions on Biomedical Engineering*, *13*, 131–139. doi:10.1109/TBME.1966.4502423

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Field, A. S. & Graupe, D. (1991). Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*. doi:10.1007/BF01129000

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*. doi:10.1111/j.1469-8986.2011.01272.x

Gupta, A. K. (2000). *Matrix variate distributions*. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics ; Boca Raton u.a.: Chapman & Hall / CRC.

Harshman, R. A. & Lundy, M. E. (1984). Data preprocessing and the extended PARAFAC model. *Research methods for multimode data analysis*, 216–284. doi:10.1109/ISCON.2002.1214589

Harshman, R. A. & Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis*, *18*, 39–72. doi:10.1016/0167-9473(94)90132-5

Hendrickson, A. E. & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70. doi:10.1111/j.2044-8317.1964.tb00244.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. doi:10.1007/BF02289447

Howe, W. G. (1955). *Some contributions to factor analysis*. Oak Ridge National Lab., Tenn.

Hunt, E. (1985). Mathematical Models of the Event-Related Potential. *Psychophysiology*, *22*, 395–402. doi:10.1111/j.1469-8986.1985.tb01621.x

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443–482. doi:10.1007/BF02289658

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. doi:10.1007/BF02289233

Kaiser, H. F. (1959). Computer Program for Varimax Rotation in Factor Analysis. *Educational and Psychological Measurement*, *19*, 413–420. doi:10.1177/001316445901900314

Kayser, J. & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*, 2307–2325. doi:10.1016/S1388-2457(03)00241-4

Kayser, J. & Tenke, C. E. (2005). Trusting in or breaking with convention: Towards a renaissance of principal components analysis in electrophysiology. *Clinical Neurophysiology*, *116*, 1747–1753. doi:10.1016/j.clinph.2005.03.020

Kayser, J. & Tenke, C. E. (2006). Consensus on PCA for ERP data, and sensibility of unrestricted solutions [7]. *Clinical Neurophysiology*, *117*, 703–707. doi:10.1016/j.clinph.2005.11.015

Kayser, J. & Tenke, C. E. (2015). Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. *International Journal of Psychophysiology*, *97*, 189–209. doi:10.1016/j.ijpsycho.2015.04.012

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed). London : Butterworths.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, Mass.: MIT Press.

Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T. J., & Courchesne, E. (1999). Functionally Independent Components of the Late Positive Event-Related Potential during Visual Spatial Attention. *The Journal of Neuroscience*, *19*, 2665–2680.

Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. doi:10.1111/j.1469-8986.2003.00139.x

Möcks, J. (1988a). Decomposing event-related potentials: A new topographic components model. *Biological Psychology*, *26*, 199–215. doi:10.1016/0301-0511(88)90020-8

Möcks, J. (1988b). Topographic components model for event-related potentials and some biophysical considerations. *IEEE Transactions on Biomedical Engineering*, *35*, 482–484. doi:10.1109/10.2119

Möcks, J. & Verleger, R. (1985). Nuisance Sources of Variance in Principal Components analysis of Event-Related Potentials. *Psychophysiology*, *22*, 674–688. doi:10.1111/j.1469-8986.1985.tb01667.x

Möcks, J. & Verleger, R. (1986). Principal component analysis of event-related potentials: A note on misallocation of variance. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *65*, 393–398. doi:10.1016/0168-5597(86)90018-3

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). CRC press.

Myers, N. D., Ahn, S., & Jin, Y. (2013). Rotation to a Partially Specified Target Matrix in Exploratory Factor Analysis: How Many Targets? *Structural Equation Modeling: A Multidisciplinary Journal*, *20*. doi:10.1080/10705511.2013.742399

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*, 494–505. doi:10.3758/s13428-014-0486-7

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.

Rösler, F. & Manzey, D. (1981). Principal components and varimax-rotated components in event-related potential research: Some remarks on their interpretation. *Biological Psychology*, *13*, 3–26. doi:10.1016/0301-0511(81)90024-7

Sass, D. A. & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within Exploratory Factor Analysis. *Multivariate Behavioral Research*, *45*, 73–103. doi:10.1080/00273170903504810

Schmitt, T. A. & Sass, D. A. (2011). Rotation criteria and hypothesis testing for Exploratory Factor Analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*, 95–113. doi:10.1177/0013164410387348

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*, S0048577299981180. doi:10.1017/S0048577299981180

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–358. doi:10.1017/S0048577201000324

Statisticat & LLC. (2016). LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com. Retrieved from https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software

Steiner, G. Z., Barry, R. J., & Gonsalvez, C. J. (2016). Sequential processing and the matching-stimulus interval effect in ERP components: An exploration of the mechanism using multiple regression. *Frontiers in human neuroscience*, *10*, 339. doi:10.3389/fnhum.2016.00339

Thomson, G. H. (1935). The definition and measurement of "g" (general intelligence). *Journal of Educational Psychology*, *26*, 241–262. doi:10.1037/h0059873

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* Chicago: University of Chicago Press. doi:10.1037/10018-000

Thurstone, L. L. (1954). An analytical method for simple structure. *Psychometrika*, *19*, 173–182. doi:10.1007/BF02289182

Verleger, R. & Möcks, J. (1987). Varimax may produce slow-wave-like shapes by merging monotonic trends with other components. *Journal of Psychophysiology*, *1*, 265–270.

Verleger, R., Paulick, C., Möcks, J., Smith, J. L., & Keller, K. (2013). Parafac and go/no-go: Disentangling CNV return from the P3 complex by trilinear component analysis. *International Journal of Psychophysiology*, *87*, 289–300. doi:10.1016/j.ijpsycho.2012.08.003

Widaman, K. F. (1990). Bias in Pattern Loadings Represented by Common Factor Analysis and Component Analysis. *Multivariate Behavioral Research*, *25*, 89–95. doi:10.1207/s15327906mbr2501{_}11

Widaman, K. F. (1993). Common Factor Analysis versus Principal Component Analysis: Differential bias in representing model parameters? doi:10.1207/s15327906mbr28031

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Lawrence Erlbaum Associates.

Wood, C. C. & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical neurophysiology*, *59*, 249–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6203715

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park: Sage Publications.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY, USA: State Univ. of New York Press.

Chapter 3

Exploratory structural equation modeling for event-related potential data – an all-in-one approach?

To obey the publisher's copyright restrictions, this chapter contains the post-peer review author version of the manuscript. The published article can be found under the following reference:

Scharf, F., & Nestler, S. (2018a). Exploratory structural equation modeling for event-related potential data—An all-in-one approach? *Psychophysiology*, e13303. doi:10.1111/psyp.13303

Nachweis über Anteile der Co-Autoren, Florian Scharf
Advances in the analysis of event-related potential data with factor analytic methods

Nachweis über Anteile der Co-Autoren:

Titel: **Exploratory structural equation modeling for event-related potential data – an all-in-one approach?**

Journal: Psychophysiology

Autoren: Florian Scharf, Steffen Nestler

Anteil Florian Scharf (Erstautor):

- Projektidee
- Implementierung des Explorativen Strukturgleichungsmodells in R
- Implementierung einer Bootstrapprozedur zur Schätzung der Standardfehler
- Konzeption der Simulationsstudie
- Programmierung und Durchführung der Simulationsstudie
- Auswertung der Simulationsstudie
- Schreiben der Publikation

Anteil Steffen Nestler (Senior-Autor):

- Konzeption der Simulationsstudie
- Schreiben der Publikation

Florian Scharf

Steffen Nestler

Exploratory structural equation modeling for event-related potential data – an all-in-one approach?

Florian Scharf and Steffen Nestler

University of Leipzig

Author Note

Florian Scharf, Institute of Psychology, University of Leipzig, Germany. Steffen Nestler, Institute of Psychology, University of Leipzig, Germany. We embrace the values of openness and transparency in science (http://www.researchtransparency.org/). We have therefore published all data necessary to reproduce the reported results and provide reproducible scripts for all data analyses reported in this paper. Correspondence concerning this article should be addressed to Florian Scharf, Neumarkt 9 - 19, 04109 Leipzig, Germany, Email: florian.scharf@uni-leipzig.de, Phone: +49 341 - 9735924.

Abstract

Event-related potential (ERP) data are characterized by high dimensionality and a mixture of constituting signals and are thus challenging for researchers to analyze. To address these challenges, exploratory factor analysis (EFA) has been used to provide estimates of the unobserved factors, and to use these estimates for further statistical analyses (e.g., analyses of group effects). However, the EFA approach is prone to biases due to assigning individual factor scores to each observation as an intermediate step and does not properly consider participants, electrodes and groups/conditions as differentiable sources of factor variance with the consequence that factor correlations are inaccurately estimated. Here, we suggest Exploratory Structural Equation Modeling (ESEM) as a potential approach to address these limitations. ESEM may handle the complexity of ERP data more appropriately because multiple sources of variance can be formally taken into consideration. We demonstrate the application of ESEM to ERP data (in comparison with EFA) with an illustrative example and report the results of a small simulation study in which ESEM clearly outperformed EFA with respect to accurate estimation of the population factor loadings, population factor correlations and group differences. We discuss how robust statistical inference can be conducted within the ESEM approach. We conclude that ESEM naturally extends the current EFA approach for ERP data and that it can provide a coherent and flexible analysis framework for all kinds of ERP research questions.

Keywords: Event-Related Potentials, Principal Component Analysis, Exploratory Factor Analysis, Exploratory Structural Equation Modeling

Exploratory structural equation modeling for event-related potential data – an all-in-one approach?

Event-related potentials (ERPs) are averaged brain responses to an event (e.g., a certain stimulus) recorded as electric potentials from several electrode sites (e.g., 64 or 128) on a participant's scalp through electroencephalography (EEG). ERPs are commonly used as a measure of differential brain activity that can be compared across experimental conditions or groups in psychological experiments. ERP data sets typically contain measurements from many time points (called *sampling points*) over a certain time range (e.g., 500 ms after the event). As the preprocessed data are averaged per participant and event type, they can be arranged in a large four-dimensional $electrodes \times participants \times conditions \times sampling\,points$ matrix. In addition to their high dimensionality, ERP data are challenging for analysts because the signal recorded from the scalp is a mixture of constituting signals called *components*. This makes it difficult to attribute differences in the observed signals to a specific component, but this must be done in order to find a functional (or physiological) interpretation of the signal. Addressing both challenges, Exploratory Factor Analysis (EFA) has been used in order to approximate the components (e.g., Dien, 2012; Dien & Frishkoff, 2005, for introductory treatments), and these estimates can be then used in further statistical analyses (e.g., analyses of group effects). Throughout this article, we will use the term EFA for this analysis approach and the term *factors* when referring to *estimated* components.

Some readers may wonder why we refer to this analysis approach as EFA instead of Principal Component Analysis (PCA). Technically, the difference between EFA and PCA is that the former contains explicit error terms for the observed variables (here: sampling points) while the latter does not (see, e.g., Widaman, 2007, 2018, for detailed discussions). In that sense, one may see PCA as a special case of EFA in which the error variances are fixed at zero (see also McDonald, 1996). Hence, when error variance approaches zero and/or the number of observed variables is high, the differences between EFA and PCA become negligible (Widaman, 2007). It has been noted that these conditions typically hold for ERP data (Dien, Beal, & Berg, 2005; Scharf & Nestler, 2018). However, it is unknown in how far this is a general condition that *always* holds for applications of the method. Therefore, we use the more general EFA model throughout this paper.

For ERP data, EFA is typically applied as follows (e.g., Dien, 2012): First, the parameters of the EFA model (i.e., factor loadings, factor correlations, and factor scores) are estimated. Second, statistical inference for group or condition effects is conducted on the factor scores (or the factor-wisely reconstructed 'raw' data), typically in the context of a general linear model (GLM; i.e., AN(C)OVAs or regressions). Such approaches are limited in at least two respects: First, it is well-known that they can result in biased estimates for GLM parameters in the second step (e.g., Devlieger, Mayer, & Rosseel, 2016; Lastovicka & Thamodaran, 1991; Lewis & Linzer, 2005; Skrondal & Laake, 2001). Second, EFA assumes only a single source of (co-)variance in the factor scores and, consequently, it commingles (co-)variance in the factors due to participants, electrodes and conditions (Scharf & Nestler, 2018).

Extended approaches are available that consider topographic variance (see Discussion) such as the two-step EFA in which a temporal EFA is followed by a spatial EFA (Dien, 2010; Spencer, Dien, & Donchin, 2001), or trilinear models (Achim & Bouchard, 1997; Field & Graupe, 1991; Möcks, 1988b; Mørup, Hansen, Herrmann, Parnas, & Arnfred, 2006; Verleger, Paulick, Möcks, Smith, & Keller, 2013; Wang, Begleiter, & Porjesz, 2000). Here, we explore how these limitations could be addressed in the framework of *Structural equation modeling* (SEM) (see Bollen, 1989; Kline, 2016, for introductions to SEMs) which is frequently used for research purposes in the social sciences (MacCallum & Austin, 2000) and for which many extensions exist (e.g., Asparouhov, Hamaker, & Muthén, 2017; Asparouhov & Muthén, 2009, 2012). Traditionally, SEMs require pre-specified measurement models which are not available for ERP data because the factor loading patterns vary heavily between experiments. This is arguably one of the reasons why neuroscientists applying SEM quantified ERP amplitudes by peak or windowed-average measures instead of utilizing EFA-estimated measurement models (e.g., Brydges, Fox, Reid, & Anderson, 2014; Charlton et al., 2008; Papaliagkas, Kimiskidis, Tsolaki, & Anogianakis, 2008; Thomas, Leeson, Gonsalvez, & Johnstone, 2013; Wilhelm, Hildebrandt, & Oberauer, 2013). Consequently, a combination of EFA and SEM would be better suited for the purposes of ERP researchers.

Recently, Exploratory Structural Equation Modeling (ESEM) was proposed that combines flexible EFA measurement models with the opportunity of SEM to model predictors of the latent variables (Asparouhov & Muthén, 2009). ESEM allows researchers to directly regress the factor scores on latent (e.g., psychometric test scores) or manifest (e.g., group or condition assignment) variables - obviating the need for a factor scoring step. Here, we suggest that ERP data can be analyzed with ESEM in order to avoid potential biases of the EFA approach and to disentangle the multiple sources of variance in ERP data.

The present article is organized as follows: First, we will explain the statistical models behind EFA and ESEM, respectively. Thereby, we briefly discuss why the EFA approach can lead to biased results and we outline the wide range of possibilities that an ESEM analysis provides. Then, we will demonstrate the use of ESEM for ERP data (in comparison with EFA) with an illustrative example. Finally, we will report the results of a small Monte Carlo simulation investigating whether ESEM effectively avoids the biases inherent to the EFA approach and whether it is able to disentangle the different sources of variance.

Exploratory Factor Analysis

The fundamental idea of temporal EFA for ERP data is that the raw voltage at a specific sampling point t_j is a weighted linear combination (i.e., a mixture) of a few (m) underlying factors:

$$t_j = \lambda_{j1} \cdot \eta_1 + \lambda_{j2} \cdot \eta_2 + ... + \lambda_{jm} \cdot \eta_m + \epsilon_j \tag{1}$$

Here, j is an index of the specific sampling point with $j = 1, 2, ..., p$, where p denotes the number of sampling points. The vector of *factor scores* η_k (with k ranging from 1 to m)

represents the characteristic values (i.e., amplitudes) of factor k for each electrode in each group/condition for all participants. The *factor loadings* λ_{jk} indicate how much each factor is reflected by each sampling point and are assumed to be constant across participants, electrodes, and groups/conditions. That is, a higher absolute value of λ_{jk} reflects a stronger influence of factor k on the voltage at sampling point j. The error term ϵ_j is the deviation of t_j from the value predicted by the model.

The EFA model is typically presented in matrix notation:

$$T = \Lambda\eta + \epsilon \tag{2}$$

where T is the $p \times n$ matrix of *all* sampling points from all n observations (i.e., data from all participants, electrodes, and conditions), η is the $m \times n$ matrix of factor scores, Λ is a $p \times m$ matrix of factor loadings, and ϵ is a $p \times n$ matrix of error terms. Essentially, EFA provides a representation of the ERP data set in which the time course and amplitude of each factor are reflected by factor loadings and factor scores, respectively.

For ERP data applications, the parameters of Equation 1 are typically estimated as follows: The factor loadings are estimated for an initial (orthogonal) model. A variety of methods can be applied to estimate the initial model such as PCA, or Maximum Likelihood Factor Analysis (MLFA) which do not differ substantially for typical ERP data sets due to the high number of variables and the high communalities (Dien & Frishkoff, 2005; Scharf & Nestler, 2018; Widaman, 1993, 2007). The initial model is typically used to determine the number of factors to be retained (but see also Kayser & Tenke, 2003), for instance, using Parallel Analysis (Horn, 1965). The factor loadings of the extracted factors are rotated utilizing an oblique rotation procedure (Dien, 1998, 2010; Dien et al., 2005; Scharf & Nestler, 2018) yielding the final (rotated) estimates of the factor loadings and the factor correlations.

Thereafter, the factor scores are obtained and submitted to subsequent statistical analyses (e.g., regressions or ANOVAs). Put simply, factor scores are an elaborate way of quantifying the amplitudes of the factors and other once common simple approaches such as peak-picking or averaging across time-windows around the peaks can be seen as naive factor scoring methods (Beauducel & Debener, 2003; Donchin, 1978). The factor scores F_η can be obtained by multiplying the (usually standardized) raw data T with the *factor scoring matrix* A_η (e.g., Mulaik, 2010):

$$F_\eta = A_\eta T \tag{3}$$

Several methods have been proposed to obtain a factor scoring matrix A_η (see DiStefano, Zhu, & Mîndrilă, 2009; Mulaik, 2010, for an overview) such as the *regression method* (Thomson, 1935; Thurstone, 1935), the *Bartlett method* (Bartlett, 1937; Thomson, 1938), or the *Anderson-Rubin* method (Anderson & Rubin, 1956). Of these methods, the regression method is still the most widely applied. In the context of PCA, it is also common to use the generalized inverse of the factor loading matrix as factor scoring matrix (Mulaik, 2010, p. 371 ff.).

The factor scores are then subjected to a GLM for further analyses (see, e.g., Myers, Montgomery, Vining, & Robinson, 2010; Rutherford, 2011, for general introductions to the GLM), that is:

$$F_\eta = X\hat{\gamma} + E \tag{4}$$

Here, $\hat{\gamma}$ denotes the vector of estimated GLM parameters (e.g., regression slopes) and X is the *design matrix* that contains the data of continuous predictor variables and, for categorical predictors, the indicator variables, respectively. The matrix E is the residual matrix of the GLM.

As ERP researchers applying EFA are typically interested in the results of the subsequent GLM, an important goal of methodological research is to investigate whether the estimated GLM parameters $\hat{\gamma}$ are unbiased with respect to their population values γ. Resembling the terminology of previous literature (e.g., Achim & Marcantoni, 1997; Dien, 1998; Kayser & Tenke, 2003; Möcks & Verleger, 1986; Wood & McCarthy, 1984), we refer to the case of biased effect size estimates in EFA as *variance misallocation*. Previous research has mainly focused on variance misallocation due to biased factor loading estimates. This may occur either as a consequence of implausibly choosing an orthogonal rotation (*orthogonality bias*; Dien, 1998; Dien et al., 2005; Scharf & Nestler, 2018) or as a consequence of the rotation itself which biases the estimates towards its simplicity criterion (*rotation bias*; Dien, 1998; Dien et al., 2005; Möcks & Verleger, 1986; Scharf & Nestler, 2018). The rotation bias is a function of the temporal overlap of the factors and gets stronger the more two factors load on the same sampling points. Another contribution to variance misallocation, which has received less attention in the context of ERP data, is that the effect size estimates may be attenuated due to the factor scoring procedure (Skrondal & Laake, 2001) - resulting in a loss of power. In the following, we will briefly explain why and when this *factor scoring bias* can occur.

When subjecting factor scores to a GLM, the GLM does not acknowledge potential uncertainty inherent to the procedure of factor scoring. In general, this uncertainty is a function of the number of indicators (i.e., sampling points) per factor and the communalities (i.e., the amount of variance explained by the factors). That is, the bias *decreases* as the number of indicators (per factor) and the communalities increase (Acito & Anderson, 1986; Devlieger et al., 2016; Skrondal & Laake, 2001). For ERP data, this implies that the bias gets smaller the more sampling points are recorded (i.e., the higher the sampling rates) and the wider the time range of the respective factor is.[1] Based on experience, it is typically assumed that high communalities and high samplings rates make the factor scoring bias negligible for ERP data (e.g., Dien & Frishkoff, 2005). However, there is no guarantee that this assumption always holds in practice. For instance, when predictors with measurement error (e.g., cognitive test scores) are used in the GLM, the factor scoring bias may become substantial.[2] In order to avoid any *risk* of biases due to factor scoring, the structural parameters can be

[1]We note that this statement is only true if the rotation bias remains unchanged. In practice, the estimated effect size is a function of the population effect size and the outlined biases. For instance, when a factor has a wide time range (i.e., loads on many sampling points), the factor scoring bias will be relatively small. However, in practice, a wider time range is typically confounded with higher temporal overlap with other factors, which in turn increases the rotation bias. Therefore, it is hard to predict the relative impact of each contribution to variance misallocation for a specific application.

[2]As outlined above, other peak measures that are based on the raw data are simplified special cases of factor scores and, hence, are prone to the factor scoring bias as well (see, e.g., Mai & Zhang, 2018, where variable averages are used as factor scores).

estimated using SEMs – or, alternatively, factor scoring methods may be used that yield unbiased regression coefficient estimates within the EFA approach (Croon, 2002; Devlieger et al., 2016; Skrondal & Laake, 2001).

Apart from the potential bias due to factor scoring, the EFA approach does not properly consider the multiple sources of variance in ERP data (Scharf & Nestler, 2018). Typically, the data matrix subjected to (temporal) EFA contains the data from all combinations of electrodes, participants and groups/conditions in the rows and treats the sampling points as variables in the columns. Consequently, the three sources of (co-)variance in the data are commingled. This has the consequence that factor (co-)variances must be interpreted as a mixture which makes it very difficult to interpret them substantively, especially because the topographic overlap between the factors virtually guarantees that the factors are substantially correlated. Perhaps the most important implication of this mixture is that orthogonal rotations must not be applied because neglecting the topographic overlap causes a substantial amount of variance misallocation (Dien, 1998; Dien et al., 2005; Dien, 2010; Scharf & Nestler, 2018). In the following, we will describe ESEM as a framework that both avoids the risk of biases due to factor scoring and that is able to disentangle the multiple sources of variance.

Exploratory Structural Equation Models

ESEM combines EFA with SEM (Asparouhov & Muthén, 2009; Marsh, Morin, Parker, & Kaur, 2014). In general, an SEM consists of a measurement model and a structural model. The measurement model describes the relationship between the observed variables (here: sampling points) and the latent factors. The structural model is a regression model in which latent variables can be predicted by other manifest or latent variables. The only difference between SEM and ESEM is that some factor loadings are fixed to zero in SEM in order to achieve a unique solution, whereas rotated EFA solutions are used as measurement models in ESEM (Marsh et al., 2014). As it is hard to pre-specify measurement models for ERP factors (because they vary between experiments), the flexibility of ESEM's exploratory measurement model is more appropriate for ERP data.

Applied to ERP data, the ESEM consists of two equations [3]:

$$\text{Measurement model: } T = \Lambda\eta + \epsilon \tag{5}$$

$$\text{Structural model: } \eta = \alpha + \Gamma X + \zeta \tag{6}$$

The measurement model (Equation 5) is the same as in EFA (Equation 2). The structural model (Equation 6) is a regression in which each observation's factor scores (η) are predicted by a set of independent manifest (X) predictors multiplied by their respective regression weights (Γ). In the case of ERP data, the manifest predictor variables could, for instance, be categorical variables encoding group assignment and electrode site (see example below). In order to adapt the model to the behavior of Dien's (2012) toolkit, one can add an intercept term α to the structural model so that the factor scores are not centered. The structural error term matrix ζ denotes the deviation of each observation's factor score from its predicted value. The interested reader is referred to the Appendix for some technical details on the estimation of the ESEM parameters.

Less technically, one could say that the structural model replaces the GLM in the ESEM approach - obviating the need for an intermediate step of factor scoring. That is, the structural model may serve for the same purposes as the GLM. In experimental contexts, mean comparisons are arguably the most prominent application. Therefore, in the following, we will focus on structural models that resemble latent versions of ANOVAs in regression formulation (see Cohen, 1968; Rutherford, 2011, for basic treatments in the context of the GLM). However, in principle, linear models with any combination of categorical (e.g., electrodes or conditions) and continuous (e.g., test scores) predictors can be specified in the structural model. The only limitation is that the same structural model must be specified for all factors, even if some predictors are

[3] In the service of comprehensibility, we simplified the ESEM equations by leaving out possible regressions among the manifest and latent variables, respectively. Fixing the intercept terms of the sampling points to zero reflects the notion that the expected voltage at each sampling point is zero in the absence of a signal (i.e., a factor). The full model can be found in Asparouhov and Muthén (2009).

relevant only for a subset of the factors (Asparouhov & Muthén, 2009).

ESEM can address both problems of the EFA approach outlined above: First, it is not prone to factor scoring biases. As can be seen from Equation 6, the factor scores (η) instead of factor score approximations (F_η, cf. Equation 4) are specified as a function of the predictor variables, avoiding the estimation of individual factor scores and eliminating the associated risk of biases (Mai, Zhang, & Wen, 2018). This is possible because ESEM operates on sufficient sample statistics (i.e., means and variances), and therefore avoids the uncertainties inherent to assigning an individual factor score to each observation. Second, the capability of ESEM to model several sources of factor variation in the structural model (cf. Equation 6) allows researchers to specify models that consider condition/group differences and topographic variance in the factors. Specifically, this can be achieved with categorical variables (e.g., 'dummies') encoding condition/group assignment and electrode site of each observation. This allows the average factor score to vary between electrodes. In addition, a plausible model must allow condition/group effects to vary between electrode sites (Möcks, 1988a). Therefore, it is crucial to also specify all interaction terms between the condition/group and electrode site in the structural model.

For instance, in the hypothetical case of 2 groups of participants and two recorded electrode sites, the following structural model can be specified that roughly resembles an electrodes × group ANOVA:

$$\eta_k = \alpha_k + \gamma_{1k} \cdot x_{group} + \gamma_{2k} \cdot x_{el} + \gamma_{3k} \cdot x_{group} \cdot x_{el} + \zeta_k \quad (7)$$

Here, the index k denotes the respective factor. The manifest predictors x_{group} and x_{el} are dummy-coded variables with a value of 1 for group 2 (x_{group}) or electrode 2 (x_{el}), respectively, and 0 otherwise. The dummy-coding of the indicators determines the interpretation of the model parameters (see Table 1 for an example): The intercept α_k represents the average factor score of the kth factor when both indicators are zero, that is, for group 1 at electrode 1. The topographic difference between the electrodes is encoded by γ_{2k}. Hence, the average factor score of the first group for the kth factor at electrode 2 is $\alpha_k + \gamma_{2k}$. The parameter γ_{1k} stands for the difference in the average factor score between the groups at electrode 1. The topographic variation of the group difference is represented by the remaining interaction parameter γ_{3k}, that is, the group difference at electrode 2 is $\gamma_{1k} + \gamma_{3k}$. The residual term ζ_k is the difference between the predicted factor score and the actual factor score of each observation (i.e., participant-electrode combination). Analogous to EFA, where the factor variance is fixed to 1, the variance of the residual ζ_k is fixed to 1 in ESEM and the remaining partial correlation between the factor scores $\varphi_{12} = cov(\zeta_1, \zeta_2)$ is estimated.

An ESEM specified in this way explicitly models the factor topographies (i.e., factor scores as a function of electrode site) assuming that the factor topography is constant across participants. It is an advantage of ESEM that the topography is directly represented by the model parameters because this enables researchers to directly test hypotheses regarding the topographic parameters within the model. One might object that the assumption of constant topographies does not necessarily hold for real data. However, we note that our parametrization is equivalent to presenting averaged topographies of (bias-corrected) EFA factor scores. Therefore, this is problematic only for cases for which the topographies are so variable across participants that the mean is not a useful representation of the sample topography any more. Apart from that, the only consequence of violating this assumption would be that the factor residuals (ζ) loose some interpretability because they would reflect both variation in the factor amplitudes and variation in the factor topographies across participants instead of only the former.

ESEM offers further advantages over the EFA approach because it provides a more general and flexible analysis framework: An ESEM model can contain any combination of manifest variables, exploratory (rotated) factors (e.g., ERP factors) and confirmatory factors (e.g., test scores) that can be regressed on manifest or latent variables – enabling researchers to test a wide range of hypotheses (see Asparouhov & Muthén, 2009). For instance, ESEMs can answer substantive questions regarding the factor correlation (e.g., Barry, De Blasio, Fogarty, & Karamacoska, 2016) because the factor correlation matrix ($\Phi = cov(\zeta)$) may be interpreted as a partial correlation matrix – controlled for the predictors in the structural model (e.g., topographic and condition/group-related variance).

Finally, we note that ESEM contains EFA as a special case: When the factors are not predicted by any other variable (intercept-only model; i.e., Γ is a zero-matrix), the ESEM and EFA solutions will be equivalent. This notion has the profound implication that all recommendations from previous methodological research on the use of EFA for ERP data also apply to ESEM – at the very least for that special case. Therefore, ESEM can be seen as a logical extension of the well-established EFA method. Having mentioned this, we would like to emphasize that ESEM is an extension of the *statistical* model of EFA. That is, it does not include any form of ERP source modeling and the resulting factors should not be interpreted as source signals. While this fact should be acknowledged as a limitation of the ESEM approach, the estimated factors may still be useful quantifications of the ERP signal and make functional interpretations easier (e.g., Beauducel, Debener, Brocke, & Kayser, 2000; Beauducel & Debener, 2003).

Illustrative Example

In the following, we will use a simplified artificial example data set based on previous simulations (Scharf & Nestler, 2018) in order to illustrate differences and similarities between EFA and ESEM and to give an example of how to apply an ESEM to ERP data. In the service of comprehensibility, we will focus on the basic principles of ESEM for an independent sample case with two groups. We note that this is done exclusively to focus on the core principles and does not represent a limitation of ESEM per se. Statistical inference for the structural parameters (Eq. 7) will be conducted with bootstrapped confidence intervals (e.g., Efron & Tibshirani, 1993) because they do not require strong assumptions about the form of the sampling distribution, and because they can be used to account for the dependencies within the data (Cameron, Gelbach, & Miller, 2008; Kreiss & Paparoditis, 2011). The bootstrapping procedure can be summarized as follows:

(1) Draw a random bootstrap sample (of original sample size) from the original sample with replacement. The sample should be drawn in a participant-wise manner (so-called block bootstrap), that is, *all* data from each participant are drawn as a fixed block if that participant is (randomly) chosen.

(2) Estimate the ESEM for the new sample.

(3) Repeat (1) and (2) many times (e.g., 1000 times) and collect the ESEM parameters each time.

(4) Obtain confidence intervals for all parameters by determining the 2.5% and the 97.5% quantiles of the parameter distributions across all bootstrap samples.

Due to the blocked resampling in step (1) this procedure preserves the correlations between the repeated measures and enables valid statistical inferences (Davison & Hinkley, 1997, chapter 8).

For illustration purposes, one may think of a typical Oddball paradigm, in which two tones that differ with respect to their probability of occurrence are presented. For instance, high tones (standards) are presented more often than low tones (deviants) and participants might be instructed to watch a silenced video while they are passively presented with the sequence of tones. The typical ERP for such an experiment is well known: Deviants elicit a negative deflection in the time range between 100 and 200 ms (N1/MMN) followed by a positive deflection around 300 ms (P3) relative to the frequent (standard) tone (e.g., Horváth, Winkler, & Bendixen, 2008). It has been attempted to use these ERPs as biomarkers for a variety of psychological disorders. For instance, reduced MMN and P3 amplitudes were found for schizophrenic patients compared to healthy controls (e.g., Kaur et al., 2011).

Inspired by this line of research, we simulated a data set with two factors recorded at two electrode sites mimicking the described deviant ERPs in a very simplified manner[4]: We created epochs with a length of 100 sampling points equally spaced over a

[4] The simulated data and the R- and Mplus-codes used to generate and analyse the simulated data are available from the Open Science Framework http://osf.io/t8mau.

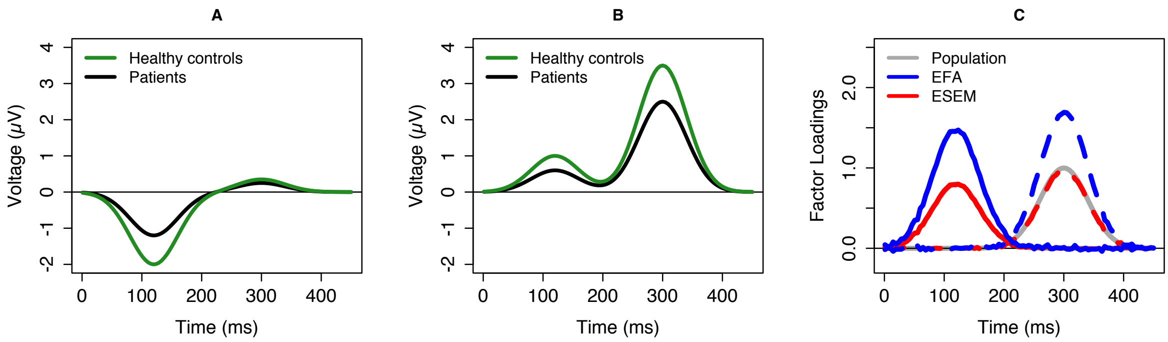

Figure 1. A & B: Population average ERPs of the example data in the two groups at the anterior (A) and posterior (B) electrode site, respectively. C: Factor loading estimates for factor 1 (solid) and factor 2 (dashed) from EFA (blue) and ESEM (red) compared with the true population value (grey). Note that the population loadings of factor 1 are hidden behind the ESEM estimates because the estimates perfectly resembled the population values. Available from http://osf.io/t8mau under a CC-BY 4.0 license.

time range of 450 ms. Roughly emulating volume conduction, data from the two electrodes *within* each participant were highly correlated (0.9). The simulated population ERPs are depicted in Figure 1 (A & B). The first factor[5] (N1/MMN) caused a negative voltage deflection around 120 ms peaking at electrode 1 ('anterior') and showed a polarity reversal at electrode 2 ('posterior'). The second factor (P3) had its maximum positive deflection around 300 ms at electrode 2 and showed a much weaker positive deflection at electrode 1. Both factors had Gaussian shaped factor loading curves with an *SD* (i.e., width) of 40 ms. The peak loadings were 0.8 (N1/MMN) and 1 (P3), that is, P3 explained more variance than N1/MMN. The factor amplitudes had a moderate correlation across participants of 0.3.

We simulated two groups (healthy controls vs. patients) with amplitude differences by defining average factor scores of -2.5 (healthy-controls-N1/MMN), -1.5 (patient-N1/MMN), 3.5 (healthy-controls-P3), and 2.5 (patients-P3), at the respective topographic maximum. That is, the N1/MMN and P3 amplitudes were higher for healthy controls than for patients. For the sake of comprehensibility, we reduced the influence of sampling error and drew an unrealistically large sample of 500 participants per group from a multivariate normal distribution. For more detailed simulations with

[5]Following the naming of ERP peaks, we will name the corresponding factors by their temporal latency, that is, η_1 refers to the factor with the *lowest* latency.

realistic sample sizes, the reader is referred to the subsequent section of the article.

For ESEM, binary indicator variables ('dummies') were used to predict the factor scores, applying the same principles as in the GLM (Cohen, 1968; Rutherford, 2011). Again, one may think of this model as a two-way latent variable ANOVA with the factors *group* (healthy controls vs. patients) and *electrode* (anterior vs. posterior). The specified structural model was the same as in Equation 7 with the group indicator (x_{group}) being 1 for patients and 0 for healthy controls. For EFA, no predictor was entered into the structural model (intercept-only model, i.e., $\eta = \alpha + \zeta$). We estimated the EFA parameters (including the factor scores) in *Mplus* (Version 8 Muthén & Muthén, 2015). The ESEM parameters were estimated with a custom implementation in *R* (R Core Team, 2018) that was based on the packages *lavaan* (Rosseel, 2012), *GPArotation* (Bernaards & Jennrich, 2005), and *boot* (Canty & Ripley, 2017; Davison & Hinkley, 1997). We employed a maximum likelihood estimator to obtain the initial solutions and subsequently rotated these initial solutions by applying an oblique *Geomin* rotation (Asparouhov & Muthén, 2009). Then, bootstrapped confidence intervals were obtained from 1000 block bootstrap samples. That is, bootstrap samples were drawn with replacement from the original sample in a participant-wise manner so that the dependency structure between the repeated measures was preserved (Cameron et al., 2008; Kreiss & Paparoditis, 2011).

For our example data set, both EFA and ESEM fit the data reasonably well as was indicated by overall model fit indices, for instance, the standardized root mean residual (SRMR) of 0.018 and 0.021, respectively. Roughly, the SRMR quantifies how well the observed covariance matrix between the sampling points can be reconstructed from the EFA/ESEM model estimates (e.g., Brown, 2014, p. 70). Taking a closer look at the model parameters, both EFA and ESEM recovered the time courses of the factors very well but only ESEM loadings were asymptotically unbiased (Fig. 1C). We note that the differences in the factor loading estimates (Fig. 1C) can be treated as a simple rescaling. In EFA, the factor variance is the sum of variance due to differences between participants, electrodes, and conditions (see Scharf & Nestler, 2018, for analytic expressions) whereas, in ESEM, it is the residual variance after controlling for condition effects and topography. As an identification constraint, the total factor variances are fixed to 1 in EFA whereas the residual variances are fixed to 1 in ESEM. The bias is therefore not relevant if the ultimate goal of the analysis is to analyse condition effects. Differences in the factor correlation estimates were more profound: The factor correlation was strongly inflated by EFA ($\hat{\varphi}_{12} = 0.74$), whereas ESEM indeed provided much less biased factor correlation estimates ($\hat{\varphi}_{12} = 0.23$). These results are in line with the notion that (co-)variance estimates are the sum of the different sources of (co-)variance in the raw data matrix (Dien, 2010; Scharf & Nestler, 2018) and demonstrate that ESEM but not EFA could separate group variance, topographic variance, and participant variance in our example data set.

Overall, the simulated topography was sufficiently recovered by ESEM. The first factor (N1/MMN) was negative at electrode 1 and showed reversed polarity (and effect-direction) at electrode 2, and the second factor (P3) had its positive maximum at electrode 2 and a weaker positive peak (and a weaker group effect) at electrode 1. This notion is further supported by the fact that the estimates of the structural ESEM parameters (Table 1) closely resembled the simulated population with respect to average factor scores (α_1, $\alpha_2 + \gamma_{22}$) in the group of patients at the topographic maximum of the respective factor. Finally, both EFA and ESEM estimated the group differences with sufficient accuracy. For ESEM, this was reflected by the respective structural parameters (γ_{11}, $\gamma_{12} + \gamma_{32}$). For EFA, separate t-tests on the regression-estimated factor scores revealed that the average factor scores of the two groups were significantly different for both factor 1 (N1/MMN), $t(998) = 15.09$, $p < .001$, $d = 0.966$, and factor 2 (P3), $t(998) = -15.28$, $p < .001$, $d = -0.954$, yielding comparable but slightly lower effect size estimates as the ESEM.

To sum up, this simple example demonstrates how ESEM extends the EFA approach. Both approaches served the goal of analysing group/condition effects well and for the given data set and yielded nearly unbiased effect size estimates. In addition to that, ESEM was able to decompose the total factor (co-)variances into the contributions due to participants, electrodes, and groups/conditions as reflected by the unbiased factor correlation estimate. Thus, apart from the ability to analyse group/condition effects, ESEM directly provides interpretable estimates of the average factor topography and the factor correlations. Robust inference can be conducted with bootstrapped confidence intervals.

A Simulation Study

To demonstrate the generalizability of the results of the illustrative example for realistic sample sizes, we conducted a simulation study that compared the performance of ESEM and EFA under varying temporal and topographic overlap of the factors, under varying sizes of the population factor correlation and for an orthogonal (Varimax; Kaiser, 1958, 1959) and an oblique rotation method (Geomin; Yates, 1987). We are aware that PCA is most commonly chosen to estimate the initial unrotated model for ERP data. For typical ERP data sets, EFA and PCA can be expected to yield equivalent results (Dien & Frishkoff, 2005; Scharf & Nestler, 2018). Nevertheless, the results of a Geomin-rotated PCA can be found as online supplementary material on the Open Science Framework (http://osf.io/t8mau).

Methods

Simulation Model. We varied *temporal overlap* (2), *topographic overlap* (2), and *population factor correlation* (2) in this simulation. In order to avoid unfeasibly long estimation times per model, we simulated epochs of only 50 sampling points equally spaced over a time range of 450 ms. The simulated data for each sample were arranged in a $(2 \cdot 10 \cdot 40) \times 50$ sample matrix T with the sampling points in the columns and, in the rows, the data from 2 groups of 40 participants at 10 electrode sites. In reference to the illustrative example, one may imagine the 10 electrodes being placed over the central line from electrode 1 ('anterior') to electrode 10 ('posterior').

Each sample was drawn from a matrix-variate normal distribution, that is, $T \sim N(M, V, \Sigma)$ where M denotes the $2 \cdot 10 \cdot 40 \times 50$ matrix of expected time courses, and V and Σ are the covariance matrices between the observations (i.e., the rows) and

Table 1

ESEM structural parameter estimates for example data

Parameter	Interpretation	Estimate	*SE*	*CI*
α_1	Average η_1 score of healthy controls at E1	-2.44	0.07	[-2.56; -2.29]
γ_{11}	Group difference in average η_1 scores at E1	1.00	0.07	[0.86; 1.12]
γ_{21}	Topographic difference in α_1 (E2 minus E1)	3.75	0.09	[3.57; 3.91]
γ_{31}	Topographic difference in γ_{11} (E2 minus E1)	-1.54	0.04	[-1.63; -1.45]
$\alpha_1 + \gamma_{21}$	Average η_1 score of healthy controls at E2	1.32	0.05	[1.20; 1.42]
$\gamma_{11} + \gamma_{31}$	Group difference in average η_1 scores at E2	-0.54	0.06	[-0.67; -0.42]
α_2	Average η_2 score of healthy controls at E1	0.32	0.05	[0.23; 0.42]
γ_{12}	Group difference in average η_2 scores at E1	-0.05	0.06	[-0.18; 0.08]
γ_{22}	Topographic difference in α_2 (E2 minus E1)	3.21	0.07	[3.06; 3.34]
γ_{32}	Topographic difference in γ_{12} (E2 minus E1)	-0.91	0.04	[-0.97; -0.84]
$\alpha_2 + \gamma_{22}$	Average η_2 score of healthy controls at E2	3.53	0.09	[3.35; 3.69]
$\gamma_{12} + \gamma_{32}$	Group difference in average η_2 scores at E2	-0.96	0.07	[-1.10; -0.82]
φ_{12}	Residual correlation of η_1 and η_2 scores	0.23	0.03	[0.18; 0.29]

Note. *SE* = Standard error, *CI* = Bootstrapped confidence interval, η_1 = Factor 1 (N1/MMN), η_2 = Factor 2 (P3), E1 = Anterior electrode, E2 = Posterior electrode. Group differences are parameterized as patients minus healthy controls. The full R and Mplus code is available from the Open Science Framework (http://osf.io/t8mau).

the sampling points (i.e., the columns), respectively (e.g., Gupta, 2000). The covariance matrix of the observations V was an identity matrix resulting in *independent* samples. This simplification was made because it avoids the necessity for bootstrapped standard errors which would result in infeasibly long simulation times.

The covariance matrix of the sampling points (Σ) was derived from the common factor model:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta, \tag{8}$$

where Λ, Φ, and Θ denote the factor loading, the factor correlation and the error covariance matrix, respectively. We specified a two factor model in the simulated population in which Gaussian density functions with a standard deviation of 40 ms emulated the time courses. The first factor peaked at 120 ms with a peak loading of 0.8, the second factor peaked either at 200 ms (L1) or at 300 ms (L2) with a peak loading of 1 – resulting in simulation conditions with high (L1) or low (L2) temporal overlap of the factors (Figure 2 A & B). The factor variance was normalized (i.e., $\varphi_{11} = \varphi_{22} = 1$), and the factors were moderately correlated ($\varphi_{12} = 0.3$) or uncorrelated ($\varphi_{12} = 0$). All error terms were mutually uncorrelated (white noise) with a moderate noise level (σ^2_{error} $= 0.4$) that was constant over the whole simulated epoch (i.e., $\Theta = \sigma^2_{error} \cdot I_p$).

The time courses were derived exactly as in Scharf and Nestler (2018): The first factor had its topographic maximum at electrode 1 ('anterior' distribution) and the second factor peaked either at electrodes 5 and 6 ('central' distribution) in simulation conditions without topographic overlap (T1) or at electrode 10 ('posterior' distribution) in simulation conditions with perfect topographic overlap (T2). At the respective topographic maximum, the expected factor scores were -1.5 (group 1) or -2.5 (group 2) for the first factor, and 2.5 (group 1) or 3.5 (group 2) for the second factor. For the remaining electrodes, these expected factor scores were multiplied with topographic weights (e.g., Achim & Bouchard, 1997; Möcks, 1988a). The weights were 1 at the topographic maximum and linearly decreased to -0.5 (factor 1) and 0.1 (factor 2) at the most distant electrode sites. Consequently, the group differences were strongest at the topographic maximum ($\gamma = 1$) and otherwise followed the topography of the factor.

The resulting ERPs at electrode 1 and electrode 10 are depicted in Figure 2 (C & D). The expected condition effects and topographies were constant across participants.

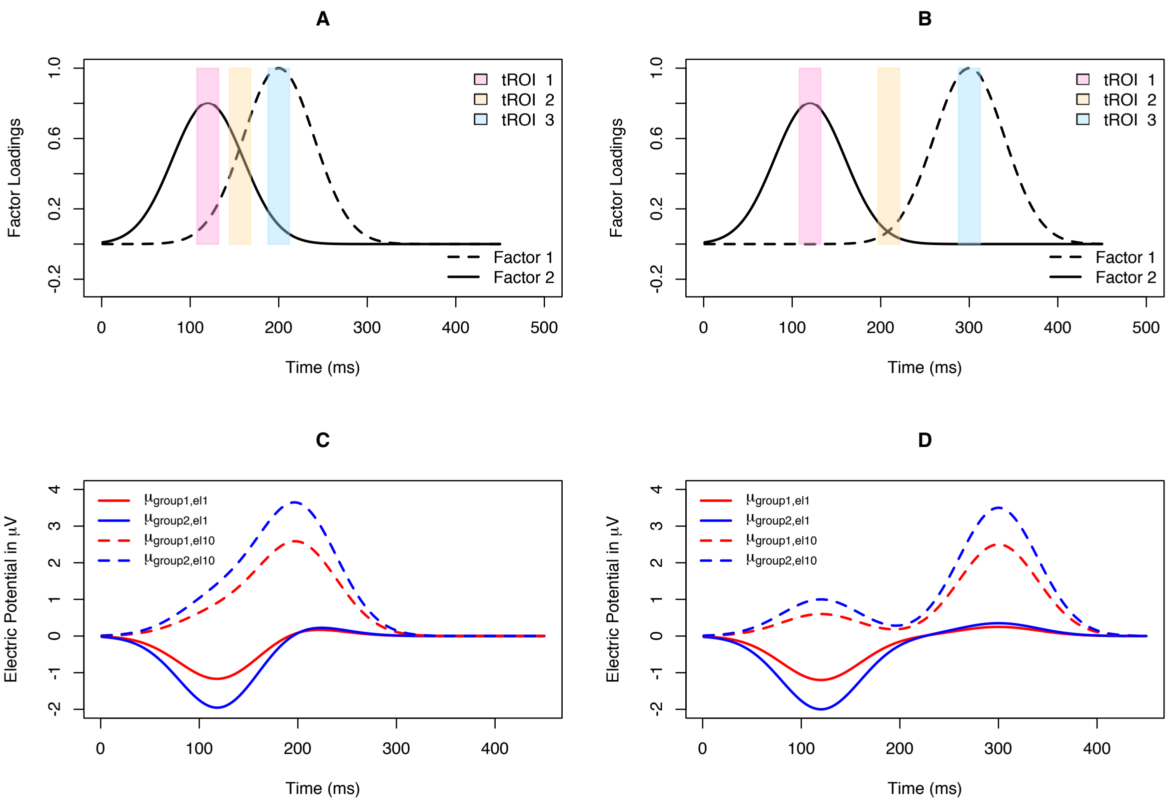

Figure 2. Upper panel: Population factor loadings of the first factor (solid) and the second factor (dashed) used in our simulations for (A) simulation conditions with high temporal overlap (L1), and (B) simulation conditions with low temporal overlap (L2). Lower panel: Population time course for the simulation condition with high (C) and low (D) temporal overlap, separately for group 1 (blue lines) and group 2 (red lines) at Electrode 1 (solid lines) and Electrode 10 (dashed lines) in the simulation condition with perfect topographic overlap. Color-shaded regions mark the temporal regions of interest (tROIs) for the analyses of the factor loading estimates. Available from http://osf.io/t8mau under a CC-BY 4.0 license.

Procedure. Simulations were conducted in *R* (Version 3.4.4, R Core Team, 2018). The following packages were used: *abind* (Plate & Heiberger, 2016), *arrayhelpers* (Beleites, 2016), *MplusAutomation* (Hallquist & Wiley, 2018), *LaplacesDemon* (Statisticat & LLC., 2016), *psych* (Revelle, 2016; Wothke, 1993), *tarRifx* (Friedman, 2014), *reshape2* (Wickham, 2007), and *xtable* (Dahl, 2016). EFA and ESEM models were estimated in *Mplus* (Version 7.4, Muthén & Muthén, 2015). All scripts for simulations and analyses are available at https://osf.io/t8mau/.

In each simulation condition, 1000 samples were generated and, for each sample, a two-factor solution was estimated for the ESEM and the EFA models employing a maximum likelihood estimator.[6] The measurement models were rotated by applying an oblique *Geomin* rotation with $\epsilon = .0001$ (Yates, 1987) or an orthogonal *CF-Varimax* rotation (Crawford & Ferguson, 1970; Yates, 1987).[7] Alignment of the factors across samples was ensured by reordering the two factors according to the rank order of their correlations with the population factor loadings. For the ESEM models, we specified dummy-coded binary indicator variables for group, electrode site and their interaction in the same manner as for the presented example – except that more dummy-variables were necessary to encode 10 instead of 2 electrode sites. For EFA, the factor scores of each factor were estimated with the regression method (see Thomson, 1935; Thurstone, 1935). The factor scores were then subjected to an independent-sample *t*-test comparing the average factor scores between the two groups.

Dependent Measures. We evaluated the performance of EFA and ESEM with respect to overall fit with the SRMR (e.g., Brown, 2014, p.70). The accuracy of the measurement model estimates was evaluated by means of the absolute biases of the parameters (θ):

$$Bias_{\theta} = \hat{\theta} - \theta \tag{9}$$

Here, the hat symbol denotes the respective estimated parameter averaged across all samples, and the symbol without hat denotes the true population value. In particular, we calculated the bias in the factor loadings ($Bias_{\Lambda}$) and the factor correlation ($Bias_{\varphi}$). We also calculated empirical standard errors (i.e., *SD*s across all 1000 samples per simulation condition) in order to quantify the precision of the estimates. For the factor loadings, we defined three time ranges of interest (tROIs) to summarize the results in a comprehensive way. Two tROIs were centered around the peaks ($\pm$ 6 sampling points) of the factors, and the remaining tROI was set around the intersection of the loading

[6]We skipped the step of determining the number of factors because previous simulations showed that this could be accurately done in similar data situations (see e.g., Scharf & Nestler, 2018).

[7]Technically, CF-Varimax is a member of the quartimin family but it yields equivalent results to the original Varimax rotation (Browne, 2001). It was chosen because the original Varimax rotation is not available for ESEM in Mplus.

curves (i.e., 6 sampling points before and after the crossing point; Figure 2 A & B). Both bias and empirical standard errors of the factor loadings were averaged within each tROI.

Finally, the risk of variance misallocation was quantified by two measures: First, the Pearson correlation between the averaged estimated factor loadings and the population loadings ($r_{\lambda\hat{\lambda}}$) was calculated for each factor as a scale-free measure of time course recovery (i.e., similarity between estimated and population factor loading vectors). Second, we calculated the bias in the estimated structural coefficient of the group effect for ESEM at the topographic maximum of each factor ($Bias_{\gamma}$). We note that due to the dummy-coding and the normalization of the factor variances, this measure is equivalent to the group difference in Cohen's d (Cohen, 1962). Therefore, for EFA, we calculated Cohen's d from the parameters of the t-test on the factor scores representing the structural parameter estimate $\hat{\gamma}$.

Results

All results of our simulations are summarized in Table 2. For applied readers, the rightmost columns containing the bias of the estimated effect sizes may be in the center of interest as they most directly reflect the consequences of biases for the statistical inferences. Both ESEM and EFA achieved reasonable overall model fits in all simulation conditions as indicated by an average SRMR of 0.025 (Hu & Bentler, 1999). None of the simulation conditions differed substantially with respect to overall model fit.

Both ESEM and EFA recovered the general time courses of the factors very well (Figure 3 & Table 2) but some characteristic distortions were observed: First, ESEM loading estimates were relatively unbiased, whereas the EFA estimates were consistently too high. Second, Varimax rotation yielded inflated cross-loadings when the factors were correlated ($\varphi_{12} = 0.3$), and, for EFA, when the topographies were overlapping (T2). Third, Geomin rotation tended to underestimate the cross-loadings when the temporal overlap was high (L1) for both EFA and ESEM. The empirical standard errors indicated that the estimates had sufficient stability.

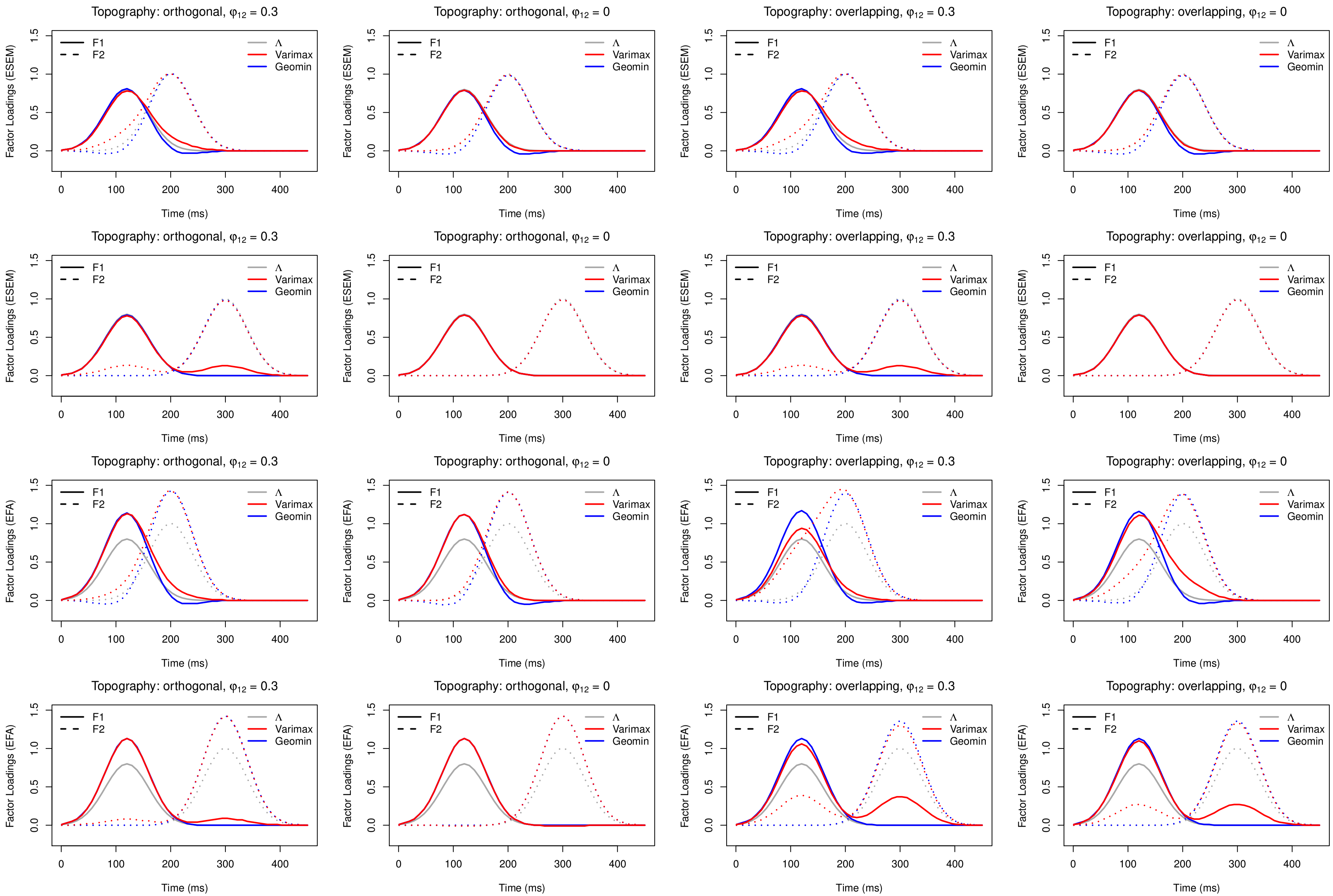

Figure 3. Average factor loadings (F1 solid lines; F2 dashed lines) estimated by ESEM (upper half) and EFA (lower half) as a function of temporal overlap (high overlap: 1st & 3rd row; low overlap: 2nd & 4th row), topographic overlap (overlapping vs. orthogonal), and between-participant correlation (correlated vs. uncorrelated). The population loadings (grey) are depicted as a reference line. Available from http://osf.io/t8mau under a CC-BY 4.0 license.

The correlation of the estimated and the population loadings ($r_{\Lambda\hat{\Lambda}}$) was consistent with this pattern yielding very high correlations (>.99) except for Varimax rotated loadings in simulation conditions with correlated factors and, for EFA only, in simulation conditions with topographic overlap (T2). Notably, the size of the correlations was similar for Geomin rotated ESEM and EFA, whereas Varimax rotated loadings yielded much higher similarity with the population loadings in ESEMs than in EFAs.

The factor correlations as estimated by the oblique Geomin rotation also revealed considerable biases. For ESEM, the factor correlation estimate was substantially inflated ($Bias_{\varphi} > 0.10$) when the temporal overlap was high (L1) but otherwise unbiased. For EFA, the bias was also more positive in the presence of temporal overlap, but, in addition, it depended on the topographic overlap: With topographic overlap (T2), the factor correlation was generally overestimated, and, without topographic overlap (T1), the factor correlation was underestimated when the factors were correlated in the population ($\varphi_{12} = 0.3$).

ESEM yielded relatively unbiased estimates of the group differences except for considerable overestimation ($Bias_{\varphi} > 0.05$) when Varimax rotation was used for correlated factors ($\varphi_{12} = 0.3$). EFA yielded generally underestimated group differences – with the strongest biases for Varimax in the presence of topographic overlap (T2) and for Geomin in the presence of temporal overlap (L1). Notably, the empirical standard errors for the group effects were quite high (~.24) and EFA and ESEM did not differ in that respect.

Discussion

In the present article, we introduced ESEM as an extension of the commonly applied EFA approach to analyzing ERP data sets. ESEM offers advantages for ERP researchers because it explicitly represents the latent factors as a function of other manifest (especially group or condition) or latent variables in the structural model. This allows researchers to examine the effects of these predictors without a separate

Table 2

Simulation results for all dependent measures as a function of estimation method, rotation method, and simulation condition

Method	Topo	φ_{12}	Latency	Rotation	$SRMR$	$Bias_{\varphi}$	$Bias_{\Lambda}$ (tROI 1) F1	$Bias_{\Lambda}$ (tROI 1) F2	$Bias_{\Lambda}$ (tROI 2) F1	$Bias_{\Lambda}$ (tROI 2) F2	$Bias_{\Lambda}$ (tROI 3) F1	$Bias_{\Lambda}$ (tROI 3) F2	$r_{\lambda\hat{\lambda}}$ F1	$r_{\lambda\hat{\lambda}}$ F2	$Bias_{\gamma}$ F1	$Bias_{\gamma}$ F2
ESEM	T1	0.3	L1	Geomin	0.025	0.14 (0.03)	0.00 (0.02)	-0.05 (0.01)	-0.04 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.9965	0.9971	-0.03 (0.23)	-0.03 (0.24)
				Varimax	0.025	-0.30 (0.00)	0.00 (0.02)	0.12 (0.03)	0.06 (0.02)	0.03 (0.03)	0.01 (0.02)	0.00 (0.01)	0.9935	0.9862	0.08 (0.23)	0.08 (0.24)
			L2	Geomin	0.025	0.00 (0.04)	-0.01 (0.02)	0.00 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
				Varimax	0.025	-0.30 (0.00)	-0.02 (0.02)	0.10 (0.02)	-0.01 (0.02)	0.09 (0.02)	0.01 (0.02)	0.03 (0.01)	0.9872	0.9914	0.10 (0.24)	0.06 (0.24)
		0.0	L1	Geomin	0.026	0.19 (0.04)	-0.02 (0.02)	-0.06 (0.01)	-0.06 (0.02)	-0.02 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.9950	0.9962	-0.01 (0.23)	-0.01 (0.24)
				Varimax	0.026	0.00 (0.00)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.00 (0.01)	0.9999	0.9999	0.02 (0.23)	0.01 (0.24)
			L2	Geomin	0.025	0.00 (0.04)	-0.01 (0.02)	0.00 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
				Varimax	0.025	0.00 (0.00)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
	T2	0.3	L1	Geomin	0.024	0.14 (0.03)	0.00 (0.02)	-0.05 (0.02)	-0.04 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.9966	0.9971	0.03 (0.24)	-0.03 (0.24)
				Varimax	0.024	-0.30 (0.00)	0.00 (0.02)	0.12 (0.03)	0.06 (0.02)	0.03 (0.03)	0.01 (0.01)	0.00 (0.01)	0.9937	0.9861	0.00 (0.24)	0.08 (0.24)
			L2	Geomin	0.024	0.00 (0.04)	-0.01 (0.02)	0.00 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
				Varimax	0.024	-0.30 (0.00)	-0.02 (0.02)	0.10 (0.02)	-0.01 (0.02)	0.09 (0.02)	0.01 (0.02)	0.03 (0.01)	0.9872	0.9914	-0.02 (0.24)	0.06 (0.24)
		0.0	L1	Geomin	0.025	0.19 (0.04)	-0.02 (0.02)	-0.06 (0.01)	-0.06 (0.02)	-0.02 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.9950	0.9962	0.05 (0.24)	-0.01 (0.24)
				Varimax	0.025	0.00 (0.00)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.00 (0.01)	0.9999	0.9999	0.02 (0.24)	0.01 (0.24)
			L2	Geomin	0.024	0.00 (0.04)	-0.01 (0.02)	0.00 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
				Varimax	0.024	0.00 (0.00)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.01)	1.0000	1.0000	0.02 (0.24)	0.00 (0.24)
EFA	T1	0.3	L1	Geomin	0.025	-0.02 (0.03)	0.23 (0.03)	0.04 (0.02)	0.01 (0.02)	0.30 (0.03)	-0.01 (0.02)	0.05 (0.02)	0.9965	0.9974	-0.06 (0.24)	-0.08 (0.24)
				Varimax	0.025	-0.30 (0.00)	0.25 (0.03)	0.17 (0.03)	0.13 (0.03)	0.32 (0.03)	0.01 (0.02)	0.05 (0.02)	0.9978	0.9980	0.01 (0.24)	-0.02 (0.24)
			L2	Geomin	0.025	-0.16 (0.03)	0.24 (0.03)	0.00 (0.01)	0.22 (0.03)	0.00 (0.01)	0.08 (0.02)	0.06 (0.02)	1.0000	1.0000	-0.01 (0.24)	-0.04 (0.24)
				Varimax	0.025	-0.30 (0.00)	0.24 (0.03)	0.06 (0.02)	0.21 (0.03)	0.06 (0.02)	0.09 (0.02)	0.08 (0.02)	0.9974	0.9987	0.02 (0.24)	-0.02 (0.24)
		0.0	L1	Geomin	0.026	0.16 (0.03)	0.22 (0.03)	0.02 (0.02)	0.01 (0.02)	0.29 (0.03)	-0.01 (0.02)	0.05 (0.02)	0.9960	0.9964	-0.04 (0.24)	-0.05 (0.24)
				Varimax	0.026	0.00 (0.00)	0.23 (0.03)	0.10 (0.03)	0.07 (0.03)	0.30 (0.03)	0.00 (0.02)	0.05 (0.02)	0.9999	1.0000	-0.02 (0.24)	-0.04 (0.24)
			L2	Geomin	0.025	-0.01 (0.03)	0.24 (0.03)	0.00 (0.01)	0.22 (0.03)	0.00 (0.01)	0.08 (0.02)	0.06 (0.02)	1.0000	1.0000	-0.01 (0.24)	-0.04 (0.24)
				Varimax	0.025	0.00 (0.00)	0.24 (0.03)	-0.01 (0.02)	0.21 (0.03)	0.00 (0.02)	0.08 (0.02)	0.06 (0.02)	0.9999	1.0000	-0.01 (0.24)	-0.04 (0.24)
	T2	0.3	L1	Geomin	0.024	0.36 (0.02)	0.26 (0.03)	0.05 (0.02)	0.03 (0.03)	0.28 (0.03)	-0.01 (0.02)	0.05 (0.02)	0.9978	0.9984	0.00 (0.25)	-0.09 (0.24)
				Varimax	0.024	-0.30 (0.00)	0.11 (0.07)	0.55 (0.07)	0.07 (0.08)	0.41 (0.04)	0.00 (0.02)	0.04 (0.02)	0.9986	0.9146	-0.02 (0.24)	0.01 (0.24)
			L2	Geomin	0.024	0.29 (0.02)	0.24 (0.03)	0.00 (0.01)	0.22 (0.03)	0.00 (0.01)	0.08 (0.02)	0.05 (0.01)	1.0000	1.0000	-0.01 (0.24)	-0.06 (0.24)
				Varimax	0.024	-0.30 (0.00)	0.19 (0.03)	0.28 (0.02)	0.17 (0.03)	0.26 (0.02)	0.12 (0.02)	0.14 (0.02)	0.9360	0.9558	-0.11 (0.24)	0.03 (0.24)
		0.0	L1	Geomin	0.025	0.53 (0.02)	0.25 (0.03)	0.04 (0.02)	0.03 (0.03)	0.27 (0.03)	-0.01 (0.02)	0.05 (0.02)	0.9973	0.9982	0.03 (0.24)	-0.05 (0.24)
				Varimax	0.025	0.00 (0.00)	0.26 (0.03)	0.32 (0.03)	0.22 (0.03)	0.32 (0.03)	0.02 (0.02)	0.04 (0.02)	0.9816	0.9741	-0.09 (0.24)	-0.05 (0.24)
			L2	Geomin	0.024	0.43 (0.03)	0.24 (0.03)	0.00 (0.01)	0.22 (0.03)	0.00 (0.01)	0.08 (0.02)	0.05 (0.01)	1.0000	1.0000	-0.01 (0.24)	-0.04 (0.24)
				Varimax	0.024	0.00 (0.00)	0.22 (0.03)	0.20 (0.02)	0.20 (0.03)	0.18 (0.02)	0.11 (0.02)	0.11 (0.02)	0.9692	0.9803	-0.12 (0.24)	-0.04 (0.24)

Note. Topo = topography (T1 = orthogonal, T2 = overlapping), L1 = high temporal overlap, L2 = low temporal overlap, φ_{12} = population factor correlation, $SRMR$ = standardized root mean squared residual, $Bias_{\varphi}$ = average difference between the estimated and the population factor correlation, $Bias_{\Lambda}$ $(\cdot)$ = average difference between the estimated and population factor (averaged within the respective tROI), $r_{\lambda\hat{\lambda}}$ = correlation between average estimated loadings and population loadings, $Bias_{\gamma}$ = bias in the group effect size, $F1$ = factor 1, $F2$ = factor 2. Standard deviations (i.e., empirical standard errors) are provided in parentheses where appropriate. Note that Varimax rotation always fixes $\hat{\varphi}_{12}$ to zero, resulting in SDs of zero across samples.

factor scoring step before the statistical inference. In addition, ESEM allows for interpretable estimates of the latent factor covariance parameters which separate (co-)variance from different sources. The results of an illustrative example and the results of a Monte Carlo simulation study supported that ESEM can disentangle topographic, group-effect related and between-participant (co-)variance, and that Maximum Likelihood estimation of ESEMs is feasible even for very small sample sizes as long as the factor loadings are high (see also De Winter, Dodou, & Wieringa, 2009).

The biases in the factor loadings for both ESEM and EFA followed the same pattern as in previous simulations (Scharf & Nestler, 2018), that is, we observed differentiable biases due to the violation of orthogonality constraints (i.e., $\varphi_{12} \neq 0$ or, for EFA, topographic overlap) and due to the rotation criterion. The former bias caused spurious cross-loadings when Varimax was used in spite of substantial factor correlations in the population. Importantly, only for EFA, these cross-loadings depended on the topographic overlap, providing further evidence that ESEM succeeded in separating the topographic variance. The latter bias due to the rotation criterion resulted in underestimated cross-loadings and inflated factor correlations for high temporal overlap (L1). The reason for this bias is that the rotation criterion aims for simple structure (i.e., small cross-loadings; Browne, 2001), and can – to some extent – achieve that at the cost of inflated factor correlations (Asparouhov & Muthén, 2009; Schmitt & Sass, 2011; Scharf & Nestler, 2018). Notably, ESEM was as prone to the rotation bias as EFA.

The results for the group effect sizes in ESEM demonstrate that these biases propagated to the structural model. Even moderate factor correlations in the population of 0.3 were enough to cause considerable (relative) biases in the group effect sizes (between 6% and 10%) when Varimax rotation was applied. In our simulations, the rotation bias for oblique rotation methods had clearly less serious consequences ($\leq$ 5% relative bias). That is, although Varimax rotation was overall less prone to the rotation bias, its benefits were smaller than its costs when the orthogonality constraint was violated. Therefore, we tend to recommend avoiding orthogonal rotation also for ESEM – except if there are very strong reasons to believe that orthogonality of the factors is a valid assumption for a given data set.

Apart from the propagated factor loading biases, EFA group effect size estimates suffered from biases due to factor scoring as indicated by the general underestimation of group effect sizes. This is consistent with previous results comparing the EFA approach with SEM or ESEM (Skrondal & Laake, 2001; Devlieger et al., 2016; Mai et al., 2018). However, although the same simulation conditions were used, these results differ from the previous simulations with a comparable setup (Scharf & Nestler, 2018) which did not observe the consistent underestimation throughout all simulation conditions. These differences can be explained by the lower sampling rate (50 vs. 200 sampling points) used here to improve the speed of the ESEM simulations. As mentioned earlier, the differences between EFA approaches and (E)SEM approaches are a function of the number of variables per factor and the communality (Acito & Anderson, 1986; Devlieger et al., 2016). Consistent with this interpretation, the biases for EFA were clearly weaker for the illustrative example with 100 sampling points.

We tested this interpretation empirically by running a supplementary simulation over a range of different numbers of sampling points. In addition, we investigated the influence of the communalities by comparing our simulation setup ($\sigma^2_{error} = 0.4$) with a setup in which the error variances were almost zero ($\sigma^2_{error} = 0.01$). In order to avoid confounds due to other contributions to variance misallocation, we simulated a single-electrode case. The number of observations was set to 4000 to minimize sampling error and we used the setup from the illustration example in which a rotation bias is avoided because the temporal overlap is very small. Figure 4 depicts the difference between factor-score-based and ESEM-based estimates of the condition effect as a function of the number of sampling points and for both investigated communality conditions. These results clearly demonstrate that differences between ESEM and factor scoring approaches depend both on the communalities and on the sampling rate. In our view, these results reemphasize that biases due to factor scoring can be substantial even for ERP data when the factors do not span a sufficient time range (relative to the sampling rate). Taken together, the results of our simulation and the supplementary

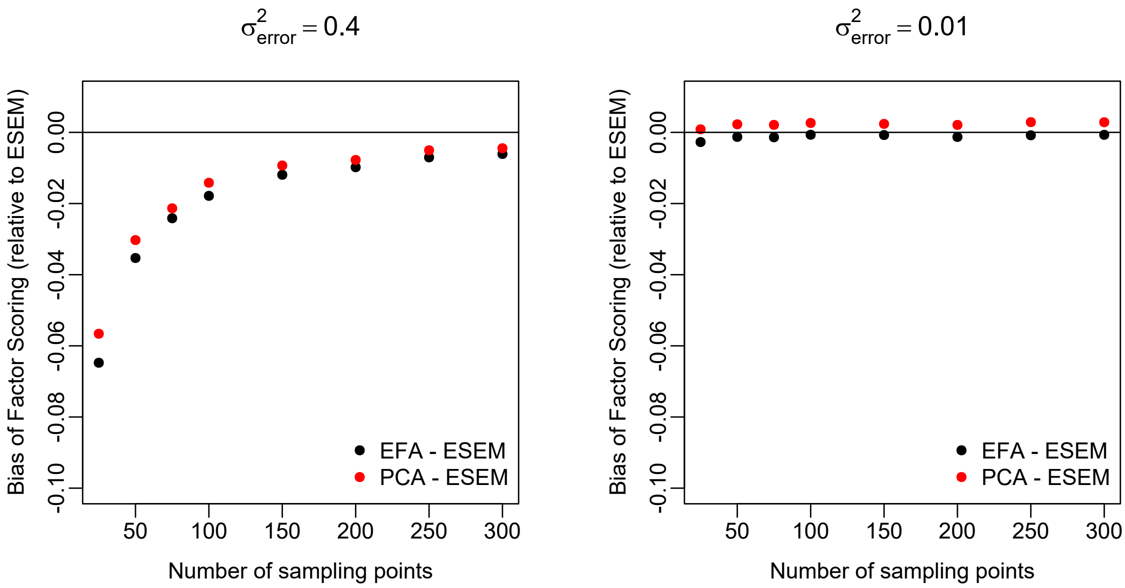

Figure 4. Differences between factor-score-based effect size estimates and ESEM-based effect size estimates as a function of the number of sampling points, separately for the noise level in our simulations (left-hand panel) and almost noise-free data (right-hand panel). EFA scores were obtained by the regression method, PCA scores were obtained by multiplying the Moore-Penrose inverse of the rotated factor loading matrix with the raw data. As the simulation study showed that ESEM yields unbiased estimates of the group effects (if there is no rotation bias) and as the raw ESEM estimates (not depicted) were within ±1 standard error around the population value, we present the EFA- and PCA-based estimates relative to ESEM as the reference method. The figure and related scripts are available from http://osf.io/t8mau under a CC-BY 4.0 license.

simulation support the notion that ESEM instead of EFA should be applied to ERP data in order to avoid the risk of biases.

Having said this, we want to stress that the present results should not be read as an argument in favor of peak-picking or averaged time-window approaches over the EFA approach. Both approaches can be described as EFA-approaches with very strict assumptions regarding the factor loading pattern (Donchin, 1978). More specifically, peak-picking assumes that the peaks of the ERPs exclusively reflect activity of the respective factor. Averaging across a time-window of interest further assumes that the factor loadings within the time-window are constant. It can be shown with very simple examples that both assumptions are not reasonable (see, e.g., Luck, 2014). As a consequence, both approaches carry an even higher risk of variance misallocation than

EFA (Beauducel & Debener, 2003).

Finally, we want to outline some tasks for future research. First of all, although our results regarding the application of ESEM to ERP-like data are promising, we acknowledge that experience with real data applications is necessary to establish ESEM as an extension of EFA. In order to support such pioneer work, we describe how an ESEM is estimated in four simple steps in the Appendix. In addition, a working basic implementation of this procedure (requiring only open source software) can be found on the OSF (http://osf.io/t8mau). Our implementation demonstrates that, in principle, any existing SEM software can be combined with rotation algorithms in order to estimate the ESEM parameters – making ESEM a very accessible method from a technical point of view.

In order to apply ESEM routinely, another crucial task is to establish reasonably robust methods for statistical inference that work both for independent groups and repeated measures. In both ESEM and EFA, the rows of the data matrix are assumed to be independent. This assumption is typically violated for repeated measures, for instance, when multiple electrodes and experimental conditions are analyzed per participant. For EFA, this is typically resolved by analyzing condition effects at selected electrode sites with *dependent* sample tests on the factor scores (or on reconstructed 'raw' data; Dien, 2012, 2016). For ESEM, we proposed a structural model that is equivalent to a two-way electrode $\times$ conditions ANOVA combined with a block bootstrap procedure to obtain confidence intervals that are valid in the presence of repeated measurements. We validated our approach both with respect to the point estimates (by comparison with the corresponding Mplus estimates), and with respect to the accuracy of the standard error estimates (and, hence, the statistical tests) by running a supplementary Monte Carlo simulation. The results confirmed that the statistical inferences hold the nominal alpha level and that the bootstrap standard errors consistently estimated the empirical sampling variation of the parameters. More detailed results and the corresponding scripts are available from the OSF.

The proposed ESEM-bootstrap approach has the advantage that it does not require strict assumptions regarding the form and the covariance structure of the sampling distributions. Furthermore, it can easily be generalized to within-participant manipulations of experimental conditions - only requiring that all data from each participant including data from all repeated measurement conditions are treated as fixed blocks. This procedure closely resembles the typical participant-wise sampling scheme in experimental contexts and therefore provides a general basis for valid bootstrap procedures (Davison & Hinkley, 1997). Nevertheless, bootstrapping is computationally extensive, and the proposed percentile bootstrap confidence intervals require nearly symmetric sampling distributions (Efron & Tibshirani, 1993). Therefore, it is important for future research to compare different statistical inference methods such as bootstrapping techniques (Efron & Tibshirani, 1993; Davison & Hinkley, 1997), latent versions of the repeated measurement ANOVA (see Rutherford, 2011, chapters 6 & 7; for a detailed technical treatment), SEM-specific random effects approaches (Asparouhov & Muthén, 2012; Rabe-Hesketh, Skrondal, & Zheng, 2007) or robust standard errors (e.g., Yuan & Bentler, 1997; Zhang, 2014) in order to determine which method is favorable under which conditions.

Another open question is how the rotation bias can be reduced. Recently, it has been suggested to develop alternative rotation criteria that do not aim for simple structure but for plausible time courses, for instance, assuming that time courses are monophasic (Beauducel, 2018). If available, prior knowledge regarding the allocation of effects (Beauducel & Leue, 2015) or the time courses of the factors (Scharf & Nestler, 2018) may be utilized in target rotations (Myers, Ahn, & Jin, 2013; Myers, Jin, Ahn, Celimli, & Zopluoglu, 2015). Further research is necessary to directly compare these approaches and to determine for which factor loading patterns they are appropriate. To this end, it is important to consider more complex factor patterns than in the present simulation because, for instance, slow-wave potentials present a considerable challenge for available rotation methods (Verleger & Möcks, 1987).

Apart from questions regarding the details of ESEM implementation, future research should compare ESEM with other proposed enhancements of the EFA procedure. A variety of methods have been proposed that take the topography into account (see also Dien & Frishkoff, 2005), among them the trilinear *topographic components model* (Achim & Bouchard, 1997; Möcks, 1988a, 1988a), and two-step PCAs (Dien, 2010; Spencer et al., 2001). Systematic comparisons of their relative performance are rare (but see, e.g., Verleger et al., 2013, for a comparison of trilinear models and PCA). For instance, formal comparisons could explore the mathematical relations between the models so that their relative strengths and weaknesses can be judged. In this context, it is especially interesting under which conditions the proposed methods yield similar (or even equivalent) results and under which conditions they differ substantially.

Beyond these basic considerations, enhancements of the ESEM could be developed that relax some of the assumptions that are currently made. An important assumption of ESEM (and EFA) is that the time courses of the factors are constant across groups (or conditions) and participants. In principle, this assumption can be relaxed and tested within the ESEM framework (Asparouhov & Muthén, 2009; Marsh et al., 2014), which may improve the appropriateness of ESEM for some experimental paradigms (Barry et al., 2016). Another interesting direction would be to relax the assumption that all participants have the same fixed condition effect so that individual differences in condition effects could be investigated. Similar approaches have already been proposed to analyse raw ERPs (Frömer, Maier, & Abdel Rahman, 2018; Pernet, Chauveau, Gaspar, & Rousselet, 2011; Tremblay & Newman, 2015).

Conclusion

In the present article, we suggested ESEM as an enhancement of EFA to analyze ERP data. The ESEM approach allows researchers to analyze ERP factor amplitudes (i.e., scores) as a function of a set of predictors while maintaining the flexibility of exploratory measurement models. Using an illustrative example, we explained, how the factor topography and group effects can be specified in an ESEM and how the model parameters are interpreted. A Monte Carlo simulation confirmed that ESEM could avoid the potential biases related to the EFA approach. We believe that this makes ESEM a suitable framework to model the full complexity of the ERP data structure. Therefore, despite the outlined open questions, we believe that ESEM is a powerful statistical technique that can become a significant addition to the ERP researcher's toolbox.

Acknowledgements

We thank three anonymous reviewers for the productive discussion through which the paper improved a lot. We also thank Jana Pförtner for helpful remarks on an earlier version of the manuscript.

Appendix

Technical details of an ESEM implementation

As a service to the interested reader, we provide some more details on the technical implementation of ESEM (see Asparouhov & Muthén, 2009, for a more detailed treatment). A basic implementation in the form of an R-script utilizing the packages *lavaan* (Rosseel, 2012), *GPArotation* (Bernaards & Jennrich, 2005) and *boot* (for bootstrapped standard errors; Canty & Ripley, 2017; Davison & Hinkley, 1997) is available from the Open Science Framework (http://osf.io/t8mau).

The estimation of an ESEM can be described in four steps:

(1) An initial SEM is estimated, in which the factors are orthogonal (i.e., $\mathrm{cov}(\zeta) = I$, where I denotes the identity matrix) and in which the upper diagonal of the factor loading matrix Λ is fixed to zero while all other elements of Λ are freely estimated for the exploratory factors (i.e., Λ is an upper-echelon matrix). This results in a solution that is rotationally indeterminate (as for any other EFA model).

(2) The factor loading matrix Λ for the exploratory factors is rotated utilizing any method of factor rotation (e.g., *Geomin* or *Infomax*).

(3) The rotation matrix (which we denote as H) is used to rotate all other model parameters. The necessary transformations can be derived from Equations 5 & 6 by applying the well-known rotation transformation (e.g., Mulaik, 2010, p. 276):

$$T = \underbrace{\Lambda H^{-1}}_{\Lambda_{rot}} \underbrace{H\eta}_{\eta_{rot}} + \underbrace{\epsilon}_{\epsilon_{rot}} \tag{A.1}$$

$$\eta_{rot} = H\eta = H \cdot (\alpha + \Gamma X + B\eta + \zeta) \tag{A.2}$$

$$= \underbrace{H\alpha}_{\alpha_{rot}} + \underbrace{H\Gamma}_{\Gamma_{rot}} X + \underbrace{HBH^{-1}}_{B_{rot}} \eta_{rot} + \underbrace{H\zeta}_{\zeta_{rot}} \tag{A.3}$$

Hence, the residual covariance matrices after rotation are:

$$cov(\epsilon_{rot}) = cov(\epsilon) \tag{A.4}$$

$$cov(\zeta_{rot}) = E(\zeta_{rot}\zeta_{rot}^T) = H \cdot cov(\zeta) \cdot H^T \tag{A.5}$$

(4) Standard errors of the model parameters are estimated. Several parametric and non-parametric approaches are available for this purpose. Parametric standard errors can, for instance, be calculated from the Maximum likelihood standard errors of the initial model via the (multivariate) delta theorem (Asparouhov & Muthén, 2009; Dorfman, 1938; Jennrich, 2007). Non parametric standard errors can be obtained, for instance, using resampling methods such as bootstraps (see Zhang, 2014, for detailed considerations in the context of EFA). Here, we used a block bootstrap, resampling the data in a participant-wise manner, because this procedure inherently controls for dependencies due to repeated measurement without requiring assumptions about the structure of the dependencies (e.g., sphericity). This approach performed very well for our example (see OSF for some validation results). However, it remains up to future research to investigate the application of other robust statistical inference techniques frequently used in ERP research (e.g., permutation tests) in the context of ESEM.

References

Achim, A. & Bouchard, S. (1997). Toward a dynamic topographic components model. *Electroencephalography and Clinical Neurophysiology*, *103*, 381–385. doi:10.1016/S0013-4694(97)96055-0

Achim, A. & Marcantoni, W. (1997). Principal component analysis of event-related potentials: Misallocation of variance revisited. *Psychophysiology*, *34*, 597–606. doi:10.1111/j.1469-8986.1997.tb01746.x

Acito, F. & Anderson, R. D. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research*, *23*, 111–118. doi:10.2307/3151658

Anderson, T. W. & Rubin, H. (1956). Statistical Inference in Factor Analysis, 111–150.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 257–269. doi:10.1080/10705511.2016.1253479

Asparouhov, T. & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. doi:10.1080/10705510903008204

Asparouhov, T. & Muthén, B. (2012). General random effect latent variable modeling : Random subjects, items, contexts, and parameters. Retrieved from http://www.statmodel.com/download/NCME12.pdf

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. doi:10.1016/j.ijpsycho.2016.06.003

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section*, *28*, 97–104. doi:10.1111/j.2044-8295.1937.tb00863.x

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. *Journal of Neuroscience Methods*, *295*, 20–36. doi:10.1016/j.jneumeth.2017.11.011

Beauducel, A. & Debener, S. (2003). Misallocation of variance in event-related potentials: Simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *Journal of Neuroscience Methods*, *124*, 103–112. doi:10.1016/S0165-0270(02)00381-3

Beauducel, A., Debener, S., Brocke, B., & Kayser, J. (2000). On the reliability of augmenting/ reducing: Peak amplitudes and principal component analysis of auditory evoked potentials. *Journal of Psychophysiology*, *14*, 226–240.

Beauducel, A. & Leue, A. (2015). Controlling for experimental effects in event-related potentials by means of principal component rotation. *Journal of Neuroscience Methods*, *239*, 139–147. doi:10.1016/j.jneumeth.2014.10.008

Beleites, C. (2016). *arrayhelpers: Convenience Functions for Arrays.* Retrieved from https://cran.r-project.org/package=arrayhelpers

Bernaards, C. A. & Jennrich, R. I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118619179

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* Guilford Publications.

Browne, M. W. (2001). An overview of analytic rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR360105

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). Predictive validity of the N2 and P3 ERP components to executive functioning in children: a latent-variable analysis. *Frontiers in Human Neuroscience*, *8*, 1–10. doi:10.3389/fnhum.2014.00080

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, *90*, 414–427. doi:10.1162/rest.90.3.414

Canty, A. & Ripley, B. D. (2017). *boot: Bootstrap R (S-Plus) Functions.*

Charlton, R. A., Landau, S., Schiavone, F., Barrick, T. R., Clark, C. A., Markus, H. S., & Morris, R. G. (2008). A structural equation modeling investigation of age-related variance in executive function and DTI measured white matter damage. *Neurobiology of Aging*, *29*, 1547–1555. doi:10.1016/j.neurobiolaging.2007.03.017

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, *65*, 145–153. doi:10.1037/h0045186

Cohen, J. (1968). Multiple Regression As a General Data-Analytic System. *Psychological Bulletin*, *70*, 426–443. doi:10.1037/h0026714

Crawford, C. B. & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, *35*, 321–332. doi:10.1007/BF02310792

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195–223). Mahwah, NJ: Lawrence Erlbaum.

Dahl, D. B. (2016). *xtable: Export Tables to LaTeX or HTML*. Retrieved from https://cran.r-project.org/package=xtable

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap methods and their application* (11. print.). Cambridge series in statistical and probabilistic mathematics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511802843

De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory Factor Analysis with small sample sizes. *Multivariate Behavioral Research*, *44*, 147–181. doi:10.1080/00273170902794206

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis Testing Using Factor Score Regression: A Comparison of Four Methods. *Educational and Psychological Measurement*, *76*, 741–770. doi:10.1177/0013164415607618

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55. doi:10.1023/A:1022218503558

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi:10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying Principal Components Analysis to event-related potentials: a tutorial. *Developmental Neuropsychology*, *37*, 497–517. doi:10.1080/87565641.2012.697503

Dien, J. (2016). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology*. doi:10.1016/j.ijpsycho.2016.09.006

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*, 1808–1825. doi:10.1016/j.clinph.2004.11.025

Dien, J. & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–208). Cambridge, MA: MIT Press.

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*, 1–11. doi:10.1.1.460.8553

Donchin, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Dorfman, R. (1938). A note on the delta-method for finding variance formulae. *The Biometric Bulletin*, *1*, 92.

Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap* ([Repr.]). Monographs on statistics and applied probability. Boca Raton u.a.: Chapman & Hall/CRC. Retrieved from https://katalog.ub.uni-leipzig.de/Record/0003155546

Field, A. S. & Graupe, D. (1991). Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*. doi:10.1007/BF01129000

Friedman, A. B. (2014). *taRifx: Collection of utility and convenience functions.* Retrieved from https://cran.r-project.org/package=taRifx

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. *Frontiers in Neuroscience*, *12*. doi:10.3389/fnins.2018.00048

Gupta, A. K. (2000). *Matrix variate distributions.* Chapman & Hall/CRC monographs and surveys in pure and applied mathematics ; Boca Raton u.a.: Chapman & Hall / CRC.

Hallquist, M. N. & Wiley, J. F. (2018). MplusAutomation : An R Package for Facilitating Large-Scale Latent Variable Analyses in M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 621–638. doi:10.1080/10705511.2017.1402334

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. doi:10.1007/BF02289447

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*, 139–147. doi:10.1016/j.biopsycho.2008.04.001

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi:10.1080/10705519909540118

Jennrich, R. I. (2007). *Rotation methods, algorithms, and standard errors.* Erlbaum Mahwah, NJ.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. doi:10.1007/BF02289233

Kaiser, H. F. (1959). Computer Program for Varimax Rotation in Factor Analysis. *Educational and Psychological Measurement*, *19*, 413–420. doi:10.1177/001316445901900314

Kaur, M., Battisti, R. A., Ward, P. B., Ahmed, A., Hickie, I. B., & Hermens, D. F. (2011). MMN/P3a deficits in first episode psychosis: Comparing schizophrenia-spectrum and affective-spectrum subgroups. *Schizophrenia Research*. doi:10.1016/j.schres.2011.03.025

Kayser, J. & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*, 2307–2325. doi:10.1016/S1388-2457(03)00241-4

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.) (T. D. Little, Ed.). Methodology in the social sciences. New York ; London: The Guilford Press.

Kreiss, J. P. & Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, *40*, 357–378. doi:10.1016/j.jkss.2011.08.009

Lastovicka, J. L. & Thamodaran, K. (1991). Common Factor Score Estimates in Multiple Regression Problems. *Journal of Marketing Research*, *28*, 105. doi:10.2307/3172730

Lewis, J. B. & Linzer, D. A. (2005). Estimating regression models in which the dependent variable is based on estimates. *Political Analysis*, *13*, 345–364. doi:10.1093/pan/mpi026

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, Mass.: MIT Press.

MacCallum, R. C. & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, *51*, 201–226. doi:10.1146/annurev.psych.51.1.201

Mai, Y. & Zhang, Z. (2018). Software Packages for Bayesian Multilevel Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *00*, 1–9. doi:10.1080/10705511.2018.1431545

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 737–749. doi:10.1080/10705511.2018.1444993

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. doi:10.1146/annurev-clinpsy-032813-153700

McDonald, R. (1996). Path Analysis with Composite Variables. *Multivariate Behavioral Research*, *31*, 239–270. doi:10.1207/s15327906mbr3102

Möcks, J. (1988a). Decomposing event-related potentials: A new topographic components model. *Biological Psychology*, *26*, 199–215. doi:10.1016/0301-0511(88)90020-8

Möcks, J. (1988b). Topographic components model for event-related potentials and some biophysical considerations. *IEEE Transactions on Biomedical Engineering*, *35*, 482–484. doi:10.1109/10.2119

Möcks, J. & Verleger, R. (1986). Principal component analysis of event-related potentials: A note on misallocation of variance. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *65*, 393–398. doi:10.1016/0168-5597(86)90018-3

Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., & Arnfred, S. M. (2006). Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, *29*, 938–947. doi:10.1016/j.neuroimage.2005.08.005

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). CRC press.

Muthén, L. K. & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Myers, N. D., Ahn, S., & Jin, Y. (2013). Rotation to a Partially Specified Target Matrix in Exploratory Factor Analysis: How Many Targets? *Structural Equation Modeling: A Multidisciplinary Journal*, *20*. doi:10.1080/10705511.2013.742399

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*, 494–505. doi:10.3758/s13428-014-0486-7

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2010). *Generalized Linear Models*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470556986

Papaliagkas, V., Kimiskidis, V., Tsolaki, M., & Anogianakis, G. (2008). Usefulness of event-related potentials in the assessment of mild cognitive impairment. *BMC Neuroscience*, *9*, 1–10. doi:10.1186/1471-2202-9-107

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: A toolbox for hierarchical linear modeling of electroencephalographic data. *Computational Intelligence and Neuroscience*, *2011*. doi:10.1155/2011/831409

Plate, T. & Heiberger, R. (2016). *abind: Combine Multidimensional Arrays*. Retrieved from https://cran.r-project.org/package=abind

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). *Multilevel Structural Equation Modeling*. Elsevier B.V. doi:10.1016/B978-044452044-9/50013-6

Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.

Rosseel, Y. (2012). {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rutherford, A. (2011). *Anova and Ancova: A GLM Approach* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc. doi:10.1002/9781118491683

Scharf, F. & Nestler, S. (2018). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119–136. doi:10.1016/j.ijpsycho.2018.03.019

Schmitt, T. A. & Sass, D. A. (2011). Rotation criteria and hypothesis testing for Exploratory Factor Analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*, 95–113. doi:10.1177/0013164410387348

Skrondal, A. & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–575. doi:10.1007/BF02296196

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–358. doi:10.1017/S0048577201000324

Statisticat & LLC. (2016). *LaplacesDemon: Complete Environment for Bayesian Inference*. Bayesian-Inference.com. Retrieved from https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software

Thomas, S. J., Leeson, P. C. R., Gonsalvez, C. J., & Johnstone, S. J. (2013). Structural equation modelling to assess relationships between event- related potential components and skin conductance in the context of emotional stimuli . *International Journal of Psychophysiology*, *94*, 245–246.

Thomson, G. H. (1935). The definition and measurement of "g" (general intelligence). *Journal of Educational Psychology*, *26*, 241–262. doi:10.1037/h0059873

Thomson, G. H. (1938). Methods of Estimating Mental Factors. *Nature*, *141*, 246–246. doi:10.1038/141246a0

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* Chicago: University of Chicago Press. doi:10.1037/10018-000

Tremblay, A. & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*, 124–139. doi:10.1111/psyp.12299

Verleger, R. & Möcks, J. (1987). Varimax may produce slow-wave-like shapes by merging monotonic trends with other components. *Journal of Psychophysiology*, *1*, 265–270.

Verleger, R., Paulick, C., Möcks, J., Smith, J. L., & Keller, K. (2013). Parafac and go/no-go: Disentangling CNV return from the P3 complex by trilinear component analysis. *International Journal of Psychophysiology*, *87*, 289–300. doi:10.1016/j.ijpsycho.2012.08.003

Wang, K., Begleiter, H., & Porjesz, B. (2000). Trilinear modeling of event-related potentials. *Brain topography*, *12*, 263–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10912734

Wickham, H. (2007). Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, *21*, 1–20. Retrieved from http://www.jstatsoft.org/v21/i12/

Widaman, K. F. (1993). Common Factor Analysis versus Principal Component Analysis: Differential bias in representing model parameters? doi:10.1207/s15327906mbr28031

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Lawrence Erlbaum Associates.

Widaman, K. F. (2018). On Common Factor and Principal Component Representations of Data: Implications for Theory and for Confirmatory Replications. *Structural Equation Modeling*, *00*, 1–19. doi:10.1080/10705511.2018.1478730

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00433. arXiv: 0610-2896

Wood, C. C. & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical neurophysiology*, *59*, 249–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6203715

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park: Sage Publications.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY, USA: State Univ. of New York Press.

Yuan, K.-H. & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, *26*, 177–198. doi:10.1016/S0167-9473(97)00025-X

Zhang, G. (2014). Estimating Standard Errors in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *49*, 339–353. doi:10.1080/00273171.2014.908271

Chapter 4

Should regularization replace simple structure rotation in Exploratory Factor Analysis?

To obey the publisher's copyright restrictions, this chapter contains the post-peer review author version of the manuscript. The published article can be found under the following reference:

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in Exploratory Factor Analysis? *Structural Equation Modeling: A Multidisciplinary Journal.* doi:10.1080/10705511.2018.1558060

Nachweis über Anteile der Co-Autoren, Florian Scharf

Advances in the analysis of event-related potential data with factor analytic methods

---

Nachweis über Anteile der Co-Autoren:

Titel: **Should regularization replace simple structure rotation in Exploratory Factor Analysis?**

Journal: Structural Equation Modeling: A Multidisciplinary Journal

Autoren: Florian Scharf, Steffen Nestler

Anteil Florian Scharf (Erstautor):

- Projektidee
- Konzeption der Simulationsstudie
- Programmierung und Durchführung der Simulationsstudie
- Auswertung der Simulationsstudie
- Schreiben der Publikation

Anteil Steffen Nestler (Senior-Autor):

- Projektidee
- Konzeption der Simulationsstudie
- Schreiben der Publikation

Florian Scharf

Steffen Nestler

Should regularization replace simple structure rotation in Exploratory Factor Analysis?

Florian Scharf and Steffen Nestler

University of Leipzig

Author Note

Florian Scharf, Institute of Psychology, University of Leipzig, Germany. Steffen Nestler, Institute of Psychology, University of Leipzig, Germany. We embrace the values of openness and transparency in science (http://www.researchtransparency.org/). We have therefore published all data necessary to reproduce the reported results and provide reproducible scripts for all data analyses reported in this paper. Correspondence concerning this article should be addressed to Florian Scharf, Neumarkt 9 - 19, 04109 Leipzig, Germany, Email: florian.scharf@uni-leipzig.de, Phone: +49 341 - 9735924.

Abstract

Exploratory factor analysis (EFA) is an important tool when the measurement structure of psychological constructs is uncertain. Typically, factor rotation is applied to obtain interpretable results resembling a simple structure. However, an overwhelming multitude of rotation techniques is available of which none is unequivocally superior. Recently, regularization has been suggested as an alternative to factor rotation. In two simulation studies, we addressed the question if regularized EFA is a suitable alternative for rotated EFA. We compared their performance in recovering pre-defined factor loading patterns with varying amounts of cross-loadings. Elastic net regularized EFA yielded estimates comparable to rotated EFA. For complex loading patterns, both rotated and regularized EFA tended to underestimate cross-loadings and inflate factor correlations but regularized EFA was able to recover loading patterns as long as a subset of items followed a simple structure. We conclude that regularization is a suitable alternative to factor rotation for psychometric applications.

Keywords: Factor Rotation, Regularization, Penalized Maximum Likelihood, Exploratory Factor Analysis, Structural Equation Modelling

Should regularization replace simple structure rotation in Exploratory Factor Analysis?

Exploratory factor analysis (EFA) is one of the most commonly used statistical methods in psychological research. EFA allows researchers to summarize the observed data (e.g., item responses) as a function of a few latent variables (e.g., traits), typically called *factors* (see, e.g., Mulaik, 2010, for a general introduction). After an initial solution has been estimated, *factor rotation* is used in order to obtain a more interpretable solution. A variety of rotation techniques have been proposed for this purpose including, for example, Varimax, Geomin, or Quartimax (see Browne, 2001; Mulaik, 2010, for overviews). However, the multitude of rotation techniques makes it difficult for applied researchers to choose an appropriate technique for their application scenario. Furthermore, research found that rotation techniques differ in their ability to uncover a known population factor structure and that the performance of a specific rotation technique depends on the population pattern itself (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011). As the population factor structure is unknown in practical EFA applications, the choice of the rotation technique is a purely subjective step.

Regularized (or sparse) EFA has been suggested as an alternative to the factor rotation step (e.g., Trendafilov, 2014; Yamamoto, Hirose, & Nagata, 2017). Instead of rotating factor loadings, regularized EFA tries to achieve a more interpretable solution such as a simple structure by penalizing factor loadings and/or factor correlations directly in the estimation step – shrinking non-substantial parameters towards zero. Regularized EFA addresses the subjectivity of the rotation approach to some extent, because the tuning parameters that are used for penalization can be determined in a more objective way, for example, by drawing on information criteria such as the BIC (e.g., Hastie, Tibshirani, & Friedman, 2009; Jacobucci, Grimm, & McArdle, 2016; James, Witten, Hastie, & Tibshirani, 2013).

Despite the growing body of methodological literature on regularized EFA, including illustrative examples of the potential of regularized EFA as a substitute for factor rotation (Trendafilov, 2014; Yamamoto et al., 2017), applications of regularized

EFA are rare. One reason for this might be that applied researchers are lacking sufficient information to judge the usefulness of regularized EFA for their specific use case. Previous research focused on direct comparisons of either rotation techniques (e.g., Schmitt & Sass, 2011) or regularization methods (e.g., Hirose & Konishi, 2012) but extensive direct comparisons of factor rotation and regularization across many data situations are not yet available. The present article aims to fill this gap by comparing the performance of different rotation techniques with different regularization techniques with regard to parameter estimation. In addition, we investigated whether only the factor loadings or both factor loadings and factor correlations should be treated as to-be-regularized parameters in regularized EFA.

In the following, we first describe how the parameters of the EFA model are conventionally estimated by factor rotation techniques. Then, we explain how regularized EFA parameters can be obtained using penalized maximum likelihood estimation with a *least absolute shrinkage and selection operator* (lasso), *ridge*, or *elastic net* (enet) penalty (e.g., Hastie et al., 2009). Thereafter, we present the results of two simulation studies comparing the performance of factor rotation and regularization in the estimation of factor loadings and factor correlations. In addition, we investigated the influence of the sparsity of the factor loading pattern on the recovery of the population parameters.

Factor Rotation in EFA

The EFA model describes the p observed variables as a weighted linear combination of m factors (e.g., Mulaik, 2010):

$$Y = \Lambda\eta + \epsilon \tag{1}$$

where Y is the $p \times n$ matrix of observed variables measured from n observations, η is the $m \times n$ matrix of factor scores, Λ is a $p \times m$ matrix of factor loadings, and ϵ is a $p \times n$ matrix of error terms. The parameters of the EFA model are often estimated with a maximum likelihood (ML) approach. Here, estimates are obtained that minimize the discrepancy between the model-implied covariance matrix of the variables Σ and their observed covariance matrix S (Jöreskog, 1969):

$$F_{ML}(\Sigma, S) = ln|\Sigma| + tr(\Sigma^{-1}S) - ln|S| - n \tag{2}$$

It is well known, that the EFA model is rotationally indeterminate. That is, an infinite set of equally well fitting factor solutions exists for a given data set that may be transformed into each other by a rotation matrix H (e.g., Mulaik, 2010, p. 276):

$$Y = \nu + \Lambda H H^{-1}\eta + \epsilon \tag{3}$$

Here, ΛH are the rotated factor loadings and $H^{-1}\eta$ are the rotated factor scores. Consequently, researchers have to 'choose' a factor solution that describes the observed data in a convenient (i.e., interpretable) way. This is achieved by rotating the initial model, that is, computing an optimal rotation matrix H according to a pre-specified criterion. Most of the common rotation techniques optimize a *simple structure* criterion of the factor loadings requiring that each variable should load highly onto one factor and should have low cross-loadings onto the other factors (Browne, 2001; Thurstone, 1935, 1954).

About 50 rotation techniques have been proposed in the methodological literature (Trendafilov, 2014). The most apparent distinction can be made between orthogonal rotation techniques (e.g., $Varimax$; Kaiser, 1958, 1959), which constrain the factor correlation to zero, and oblique rotation techniques (e.g., *Geomin*; Yates, 1987), which allow the factors to be correlated. Apart from that, rotation techniques differ in the exact criterion that is used to operationalize a simple factor loading pattern (e.g., Browne, 2001), and hence in their tolerance for cross-loadings (Schmitt & Sass, 2011). In the following, we will briefly contrast three common rotation techniques: $Varimax$ rotation, *Geomin* rotation, and $Facparsim$ as a member of the general *Crawson-Ferguson* (CF) rotation family.

$Varimax$ (Kaiser, 1958, 1959) is one of the most widely applied rotation techniques (Fabrigar, Wegener, MacCallum, & Strahan, 1999). It assumes uncorrelated factors and maximizes the variance of the squared loadings:

$$f(\Lambda) = \sum_{i=1}^{p} \lambda_{ij}^4 - \frac{1}{p}\left(\sum_{i=1}^{p} \lambda_{ij}^2\right)^2 \tag{4}$$

An initial *Varimax* rotation is an essential part of the popular oblique *Promax* criterion (Hendrickson & White, 1964).

The *Geomin* rotation was explicitly developed in order to represent more complex loading patterns. It minimizes the (variable-wise) geometric mean of the squared factor loadings (Browne, 2001):

$$f(\Lambda) = \sum_{i=1}^{p} \left[\prod_{j=1}^{m} (\lambda_{ij}^2 + \epsilon)\right]^{\frac{1}{m}} \tag{5}$$

Here, ϵ is an additional parameter to ensure that the rotation criterion is generally differentiable even if one of the factor loadings is exactly zero for each variable. Traditionally, $\epsilon = 0.01$ is used but a slightly higher value of $\epsilon = 0.5$ has been suggested in the literature to better represent more complex factor structures (Marsh et al., 2009; Marsh et al., 2010; Morin, Marsh, & Nagengast, 2013).

The CF rotation criterion considers both variable-wise and factor-wise complexity in the rotation:

$$f(\Lambda) = (1-k)\sum_{i=1}^{p}\sum_{j=1}^{m}\sum_{l\neq j, l=1}^{m} \lambda_{ij}^2\lambda_{il}^2 + k\sum_{j=1}^{m}\sum_{i=1}^{p}\sum_{l\neq i, l=1}^{p} \lambda_{ij}^2\lambda_{lj}^2 \tag{6}$$

The first and second term in this sum quantify the row (i.e., variable) and column (i.e., factor) complexity, respectively, in the factor loading matrix. The weight $k = [0, 1]$ determines which complexity receives more emphasis during factor rotation with higher values indicating more emphasis on factor complexity. Many rotation criteria may be described as special cases of the CF criterion with different values of k (e.g., Sass & Schmitt, 2010).

Due to the different simple structure criteria, rotation techniques differ in their ability to recover population patterns that differ in their amount of cross-loadings.

More specifically, simulation studies showed that orthogonal rotations of oblique factor patterns yield spurious cross-loadings even if the population factor loading pattern is a simple structure (Schmitt & Sass, 2011). Furthermore, oblique rotations tend to yield inflated factor correlations and underestimated cross-loadings in the presence of high cross-loadings in the population factor loading pattern (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011). Thus, factor rotation techniques can achieve a more unique assignment of variables to factors at the cost of less distinct factors (or vice versa), and it is desirable to achieve a reasonable trade-off between factor correlation and cross-loadings.

To summarize, factor rotation is utilized to find a factor solution that is as interpretable as possible according to some simplicity criterion. However, applying rotation techniques has two major drawbacks: First, researchers have to choose among a multitude of rotation techniques that is still growing (e.g., Beauducel, 2018; Ertel, 2011; Jennrich, 2004, 2006; Yamamoto & Jennrich, 2013). In addition, some rotation techniques have tuning parameters (e.g., k and ϵ mentioned above) that also have to be chosen (and that have profound consequences for parameter estimation). Second, the performance of a rotation technique in terms of the suitability of the parameter estimates depends on the true data generating mechanism in the population, especially on the amount of cross-loadings, which is unknown in practice. Therefore, alternative approaches that perform consistently well across a wide range of factor patterns would be preferable – obviating the need for such a subjective choice.

Regularized EFA

The rotation problem in EFA can be re-conceptualized as a variable (or model) selection problem in which a set of indicator variables for each factor needs to be chosen among all indicators (e.g., Hirose & Konishi, 2012; Hirose & Yamamoto, 2014). A lot of work in the SEM literature has been done on efficient model selection via heuristic search algorithms in the context of model modification (e.g., Glymour, Madigan, Pregibon, & Smyth, 1997; Marcoulides, Drezner, & Schumacker, 1998;

Marcoulides & Drezner, 2003; Marcoulides & Ing, 2013; Marcoulides & Falk, 2018). Essentially, these approaches are best subset selection methods that try to go through the space of possible models as efficiently as possible but still apply conventional estimators (such as Maximum Likelihood) to estimate the model parameters. The best model is then selected based, for instance, on the Bayesian information criterion (BIC, Schwarz, 1978). In contrast, regularization strives for a solution in which as many parameters as possible are (close to) zero directly during the estimation of the model with the goal that only few but substantial variables remain in the final model (Hastie et al., 2009).

A vector of parameters in which most of the entries are zero is called *sparse*. With respect to EFA, where the parameters to be estimated are the factor loadings and factor correlations, a perfect simple structure of the factor loadings can be seen as a special case of sparsity, and this is the reason why regularized EFA has been suggested as an alternative to factor rotation (Trendafilov, 2014). However, unlike the simple structure criterion, the sparsity condition does not refer to a specific pattern of the non-zero estimates within the set of parameters (e.g., that cross-loadings are zero). In that sense, a conventional rotated EFA and a regularized EFA both aim for a simple factor loading pattern but the latter does not distinguish between factor and variable complexity (see Hirose & Yamamoto, 2015a; Yamamoto et al., 2017, for more formal treatments of this notion). This property may enable regularized EFA to flexibly recover a larger variety of factor loading patterns than rotated EFA – challenging the predominance of rotation in typical psychometric applications.

Several variations of regularized EFA have been proposed in the literature (e.g., Arruda & Bentler, 2017; Hirose & Konishi, 2012; Hirose & Yamamoto, 2014; Huang, Chen, & Weng, 2017a, 2017b; Jacobucci et al., 2016; Jung & Takane, 2007; Jung & Lee, 2011; Trendafilov & Adachi, 2015; Trendafilov, Fontanella, & Adachi, 2017). Here, we focus on approaches that directly add a penalty term at the ML-estimation stage with the aim of finding a sparse measurement model (Hirose & Yamamoto, 2014; Jacobucci et al., 2016). That is, the parameters of the model are estimated by minimizing a penalized version of the ML fit function (Eq. 2):

$$F_{regEFA}(\Sigma, S) = F_{ML}(\Sigma, S) + \alpha \cdot P(\theta) \quad (7)$$

$$= ln|\Sigma| + tr(\Sigma^{-1}S) - ln|S| - n + \alpha \cdot P(\theta) \quad (8)$$

Here, $P(\theta)$ is a penalty function of a vector of model parameters θ that may, in principle, contain any parameter of the model, that is factor loadings or factor correlations. The tuning parameter α determines the amount of penalty applied during estimation and needs to be determined in a separate step. In general, the penalty term will increase as a function of the number of non-zero parameter estimates so that the estimation procedure prefers models with many low or zero parameter estimates. In that respect, rotation and regularization have similar objectives (see Trendafilov, 2014; Yamamoto et al., 2017, for illustrative examples).

The vector of penalized parameters θ may contain the factor loadings (i.e., $\theta = vec(\Lambda)$) or both the factor loadings and the factor correlations (i.e., $\theta = [vec(\Lambda) \quad \varphi]$, where φ denotes a vector containing all factor correlations)[1]. Furthermore, a variety of different penalty functions have been proposed in the literature including ridge, lasso, and enet penalties. Both ridge (Hoerl & Kennard, 1970) and lasso (Tibshirani, 1996) penalties are based on vector norms of the parameter vectors. Specifically, ridge uses the sum of the squared parameter estimates as penalty term, while lasso penalizes the sum of the absolute values of the parameter estimates:

$$P_{ridge} = \|\theta\|_2 = \sum_i \theta_i^2 \quad (9)$$

$$P_{lasso} = \|\theta\|_1 = \sum_i |\theta_i| \quad (10)$$

Here, $\| \cdot \|$ denotes the respective norm operator and the sum is taken across all

[1]For the sake of completeness, it should be noted that θ must not contain factor loadings and variable residuals at the same time because their strong relationship (the higher the factor loadings, the smaller the variable residuals) would lead to severe estimation problems (Jacobucci et al., 2016).

parameters contained in θ. Both penalties result in a shrinkage of the parameters towards zero but only lasso can shrink the parameter estimates to exactly zero (i.e., for $\alpha \to \infty$), allowing for variable selection (Hastie et al., 2009). Importantly, the variable selection property of regularization also removes the rotational indeterminacy so that regularized EFA solutions are unique (except for reordering of factors and sign switches; Choi, Oehlert, & Zou, 2010). Another consequence of the penalty term is that regularized EFA solutions tend to fit the data slightly worse than rotated EFA solutions.(e.g., Jin, Moustaki, & Yang-Wallentin, 2018; Trendafilov et al., 2017).

The ability to conduct variable selection is an advantage of lasso over ridge penalization. However, lasso regression performs worse than ridge if the number of parameters exceeds the number of observations by far (Zou & Hastie, 2005). For these situations, the enet penalty has been proposed which applies both a lasso and a ridge penalty. That is, enet considers both the sum of the absolute values of the parameter estimates and the sum of the squared parameter estimates:

$$P_{enet} = (1 - \beta)\|\theta\|_1 + \beta\|\theta\|_2 \quad (11)$$

Enet can be seen as a generalization of lasso and ridge. It includes an additional weight parameter β that determines which of the penalties receives more weight. Notably, when $\beta = 1$ or $\beta = 0$, enet is equivalent to ridge and lasso, respectively. When $0 < \beta < 1$, continuously less parameters are shrunken to exactly zero the more β approaches 1 (Hastie et al., 2009).

For the sake of completeness, it should be mentioned that further penalty functions have been proposed with the explicit goal of achieving sparser solutions than factor rotation (Fan & Li, 2001; Hirose & Yamamoto, 2015b; Hirose, 2016; Zhang, 2010). This is especially important for data sets with a very large number of variables (relative to the sample size) such as genome data (e.g., Carvalho et al., 2008), or fMRI data (e.g., Hirose, 2016). However, as outlined in the context of factor rotation, simpler (or sparser) solutions are typically accompanied by inflated factor correlations for psychometric data sets. Considering even sparser solutions would rather aggravate the outlined problems. Therefore, in the present paper, we focus on penalties that do not specifically aim for sparser solutions than factor rotation (i.e., ridge, lasso, and enet).

Apart from the choice of the penalty function, the tuning parameter α heavily influences the estimates of regularized EFA. In general, the parameter estimates and the proportion of non-zero parameters decrease as α increases. In that sense, regularization may be seen as a continuous and objective approach to achieve a simple (or sparse) solution as the tuning parameter is determined utilizing an objective criterion. Typically, the model is estimated over a range of possible values for α, and the set of parameters is chosen that yields the best cross-validated fit as indicated by the root mean squared error or some information criterion (Hastie et al., 2009) such as the BIC. For regularized factor analysis models, the BIC performs well in finding a penalty weight that results in reasonable parameter estimates (Hirose & Yamamoto, 2014; Jacobucci et al., 2016).

In summary, regularized EFA aims for a sparse factor loading matrix with as many zero-elements as possible but without assuming a specific form of simple structure (e.g., that each variable should have at least one zero loading). This makes regularization a potential alternative to factor rotation also for psychometric applications (see also Trendafilov, 2014). Importantly, it should *not* be expected that regularized EFA always performs better than rotated EFA - although there may be conditions under which regularized EFA is generally superior in terms of parameter estimation. Rather, regularized EFA may provide a better compromise between the ability to uncover simple structure *when* it exists with the ability to offer reasonably interpretable results when this is not the case. In that sense, regularized EFA could already be considered a suitable alternative for factor rotation if its estimates are not substantially worse than the best rotated EFA for a given factor loading pattern. Previous research convincingly demonstrated that regularized EFA achieves interpretable solutions for some specific applications (e.g., genome data or popular standard examples for EFA). Extending these studies, we investigated if this observation can be generalized over a wider range of populations inspired by typical psychometric data sets in which we systematically varied the size of the cross-loadings.

The present study

Extensive empirical comparisons of factor rotation techniques for typical psychometric data sets are available (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011) and a number of studies have compared different penalty functions in the context of both exploratory factor analysis and regression (Fan & Li, 2001; Hirose & Yamamoto, 2014, 2015b; Hirose, 2016; Huang et al., 2017a; Zhang, 2010). Occasionally, rotation techniques have been included in simulation studies on regularization but these comparisons were limited to one specific rotation technique per study and considered only a small variety of factor loading patterns (Hirose & Yamamoto, 2015b; Ning & Georgiou, 2011; Trendafilov & Adachi, 2015, 2015).

Hence, more direct and extensive comparisons of factor rotation and regularization on the same data set are necessary to judge the usefulness of regularization for typical psychometric applications. Closing this gap, we report the results of two simulation studies comparing the performance of the described regularization methods with factor rotation. In the first simulation, we compared the asymptotic properties of factor rotation and regularization for large samples, and, in the second study, we investigated whether the results from the large samples also apply to samples sizes that are more realistic in psychological research.

Study 1: Asymptotic performance

In this study, we compared the performance of factor rotation and regularization for large samples. We based our comparison on factor loading patterns for which the performance of common factor rotation techniques is known. In addition, we investigated the performance of regularization and factor rotation on extended factor loading patterns with more items and varying degrees of sparsity. We focused on oblique *CF-Varimax* rotated EFA due to its popularity, and on *Geomin* and *Facparsim* rotated EFAs as comparison techniques because they performed best for complex loading patterns in previous simulations (Asparouhov & Muthén, 2009; Schmitt & Sass, 2011). We estimated regularized EFAs with ridge, lasso, and enet penalties. We either penalized the factor loadings only or both the factor loadings and the inter-factor correlations.

As outlined above, the success of factor rotation depends on the degree to which the population factor loadings follow a simple structure. Regarding the performance of regularization, it is an important property, especially of lasso penalties, that the performance of the regularization depends on the degree to which the true model is actually sparse (Donoho, 2006; Donoho & Stodden, 2006). Only if a sufficient degree of sparsity holds in the population, regularization is likely to recover the correct factor loading pattern. In the present context, the factor loading patterns are sparser the closer they approximate a perfect simple structure. Therefore, we expected that regularization behaves similarly as factor rotation in case of a population pattern that conforms to a simple structure. For more complex factor loading patterns, the cross-loading estimates are shrunken towards zero at the cost of inflated factor correlations. We expected that this effect is reduced if the penalty also considers the inter-factor correlations. In this case, inflated factor correlations would result in a higher penalty term, which in turn should allow the estimation procedure to aim for a better compromise between minimal cross-loadings and minimal factor correlations.

Methods

Simulation Model. We simulated factor loading patterns of varying complexity with 3 factors and 18 (basic conditions) or 36 variables (extended conditions). For the sake of comparability, we adapted the simulation setup that was used in Schmitt and Sass (2011, Tab. 1) comprising of 18 variables that follow a perfect simple pattern, an approximate simple pattern with small cross-loadings (< 0.20) or a complex pattern with substantial cross-loadings of up to 0.40 (standardized loadings). The standardized main loadings varied between 0.63 and 0.75 and the factors were standardized (variance of 1) and substantially correlated (0.40). The residual covariance matrix was a diagonal matrix, that is, the residuals were uncorrelated.

Table 1
Simulated factor loading patterns (standardized) in the replicated conditions

	Perfect Simple Pattern			Approximate Simple Pattern			Complex Pattern		
Variable	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	0.75	0.00	0.00	0.70	0.11	0.14	0.67	0.22	0.13
2	0.75	0.00	0.00	0.70	0.17	0.05	0.68	0.09	0.23
3	0.75	0.00	0.00	0.68	0.16	0.16	0.68	0.27	0.05
4	0.75	0.00	0.00	0.70	0.05	0.17	0.65	0.39	0.09
5	0.75	0.00	0.00	0.72	0.08	0.08	0.64	0.13	0.39
6	0.75	0.00	0.00	0.70	0.11	0.11	0.67	0.18	0.18
7	0.00	0.75	0.00	0.11	0.69	0.17	0.05	0.68	0.27
8	0.00	0.75	0.00	0.05	0.72	0.08	0.25	0.63	0.38
9	0.00	0.75	0.00	0.05	0.72	0.08	0.38	0.63	0.21
10	0.00	0.75	0.00	0.16	0.68	0.16	0.09	0.69	0.18
11	0.00	0.75	0.00	0.08	0.71	0.11	0.05	0.73	0.05
12	0.00	0.75	0.00	0.05	0.71	0.14	0.27	0.67	0.13
13	0.00	0.00	0.75	0.08	0.14	0.70	0.04	0.40	0.66
14	0.00	0.00	0.75	0.14	0.14	0.69	0.38	0.25	0.63
15	0.00	0.00	0.75	0.14	0.11	0.70	0.26	0.18	0.66
16	0.00	0.00	0.75	0.11	0.05	0.71	0.14	0.09	0.70
17	0.00	0.00	0.75	0.16	0.14	0.69	0.22	0.22	0.66
18	0.00	0.00	0.75	0.08	0.05	0.72	0.18	0.09	0.69

Note. The factor correlations were .40 among all factors. These simulation parameters were adapted from Schmitt and Sass (2011, where the same patterns are used but unstandardized loadings are presented). The factor loadings were standardized with respect to the total variable variances (Muthén, 2004, Appendix 3). F· = Factor.

In addition to these basic conditions, we also explored the behavior of factor rotation and regularization for five extended factor loading patterns with 36 variables. These patterns were constructed by concatenating different combinations of the basic patterns (Tab. 2). These extended conditions enabled us to investigate the influence of the number of items and the degree of sparsity on the performance of factor rotation an regularization. For instance, in the condition 'Extended Simple 1' comprised of a

Perfect Simple pattern and an Approximate Simple pattern, that is, the variables 1 to 18 loaded on the 3 factors with 6 main loadings (0.75) per factor and otherwise zero-loadings (Tab. 1, left-most pattern), and the variables 19 to 36 loaded on the 3 factors with small cross-loadings (Tab. 1, middle pattern). In this pattern, both factor and variable complexity are higher than in the basic Perfect Simple pattern but the sparsity is preserved to some extent due to the zero cross-loadings of the first 18 variables. Hence, this condition is optimally suited to investigate differential behavior of rotated and regularized EFA.

Finally, in order to exclude that differences between basic and extended conditions may be attributed to the increased number of variables, we included two conditions (Extended Simple 2, Extended Complex 3) that differed to the respective basic condition only in the number of variables.

Table 2
Simulated factor loading patterns in the extended conditions

Condition	Loading Pattern 1	Loading Pattern 2
Extended Simple 1	Approximate Simple	Perfect Simple
Extended Simple 2	Approximate Simple	Approximate Simple
Extended Complex 1	Complex	Perfect Simple
Extended Complex 2	Complex	Approximate Simple
Extended Complex 3	Complex	Complex

Note. All extended patterns had 36 variables. The basic patterns can be found in Tab. 1.

Procedure. All simulations and analyses were conducted in *R* (Version 3.4.4, R Core Team, 2018). All scripts necessary to reproduce the simulations and analyses are available from the Open Science Framework. For each factor loading pattern, we derived the implied covariance matrix of the variables from the common factor model (e.g., Mulaik, 2010, p. 138). In order to investigate the asymptotic behavior of the discussed approaches, we drew large random samples of $N = 10000$ participants from a continuous multivariate normal distribution using the package *mvtnorm* (Genz et al., 2017).

For each condition, the respective data set was subjected to rotated ML-EFAs and regularized EFAs, extracting 3 factors. ML-EFA was conducted as implemented in the package *psych* (Version 1.7.5 Revelle, 2016) and rotated using oblique *CF-Varimax*, oblique *Geomin* ($\epsilon = 0.01$ or 0.5) or *Facparsim* rotations from the *GPArotation* package (Bernaards & Jennrich, 2005). Regularized EFA was conducted using the package *regSEM*, a general package that estimates regularized structural equation models (regSEMs) allowing the user to select which parameters of the model should be penalized (Jacobucci, 2017). In order to estimate a regularized EFA, we specified three measurement models but no structural model. Each variable was allowed to load on each factor, the factor variances were fixed to 1, and all factor correlations were freely estimated. Both rotated and regularized EFA were conducted on z-standardized data.

We compared the performance of ridge, lasso and enet penalty on either the factor loadings alone ($Ridge_{\Lambda}$, $Lasso_{\Lambda}$, $Enet_{\Lambda}$) or the factor loadings and factor correlations ($Ridge_{\Lambda,\Phi}$, $Lasso_{\Lambda,\Phi}$, $Enet_{\Lambda,\Phi}$). The tuning parameters α and β (for enet, Eq. 11) were automatically chosen so that they minimized the BIC over the respective sample (Jacobucci et al., 2016). For α, a grid of 100 values starting from $\alpha = 0.001$ with a step size of 10^{-5} was used. For all models, we ensured that the final parameter estimate was not at the boundary of the grid (indicating that the parameter space should be enhanced). For β, we tested values between 0.05 and 0.95 with a step size of 0.05.

Dependent Measures. The standardized root mean residual (SRMR) was calculated as measure of fit between the model-implied and the observed correlation matrices following the procedure described, for instance, by Asparouhov and Muthén (2018, section 2.2). Before calculating the dependent measures describing the recovery of the population parameters, factors were inverted if the sum of their loadings was negative (e.g., Asparouhov & Muthén, 2009, Appendix D) and factor alignment was ensured by reordering the factors according to their highest Tucker congruency with the respective factors in the population loading pattern (e.g., Lorenzo-Seva & ten Berge, 2006). The average congruency across all 3 factors also served as measure of similarity between the estimated and the population loadings. In addition, the average bias (across all factors and variables) was calculated separately for main loadings, cross-loadings, and the inter-factor correlations.

To avoid misunderstandings, it should be reiterated that due to the rotational indeterminacy of the EFA model an infinite set of factor loadings and factor correlations has the same fit for a given correlation matrix. Hence, the concept of parameter bias does not apply in the same manner to the EFA model as to other models (e.g., linear regression). Following conventions of previous studies (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011), we define bias as the deviation of the estimated parameters from the data generating parameters. The main purpose of the reported bias measures is to summarize the estimated factor loading patterns in an efficient way, and they should not be understood as deviations from a ground truth.

Results

The main results of this simulation are summarized in Table 3. The model fit as indicated by the SRMR was nearly perfect across all methods but slightly worse for lasso and enet regularized EFA than for rotated or ridge regularized EFA. The estimated factor loadings and factor correlations for each condition are available from the OSF. Except for ridge penalized regularized EFAs, the congruencies and biases of the main loadings indicated that all investigated methods recovered the general factor pattern sufficiently. However, we observed strong differences between the methods with respect to the biases in the cross-loadings and factor correlations. Notably, across all conditions, methods and factors, the bias in the cross-loadings and the bias in the factor correlations were strongly correlated, $r_{spearman} = -0.98$, indicating that the bias in the factor correlation reliably increased the more the cross-loadings were underestimated.

Basic conditions. *Geomin* ($\epsilon = 0.01$) rotation performed very well for conditions with simple structure in the population, yielding unbiased estimates of main loadings, cross-loadings and factor correlations. However, in the presence of substantial cross-loadings, *Geomin* ($\epsilon = 0.01$) underestimated the cross-loadings and overestimated both main-loadings and factor correlations. This pattern was more pronounced the more complex the factor pattern was (Approximate Simple vs. Complex condition). The alternative tuning parameter in *Geomin* ($\epsilon = 0.5$) had profound influences on the

Table 3

Simulation results for all dependent measures as a function of estimation method and simulation condition in Study 1

		Condition							
		Basic			Extended				
Method	Measure	Perfect Simple	Approximate Simple	Complex	Simple 1	Simple 2	Complex 1	Complex 2	Complex 3
Geomin ($\epsilon = 0.01$)	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	1.00	0.98	0.98	1.00	0.98	1.00	0.98	0.98
	Bias main loadings	-0.00	0.08	0.07	0.03	0.09	0.03	0.07	0.07
	Bias cross-loadings	0.01	-0.09	-0.10	-0.04	-0.09	-0.04	-0.09	-0.10
	Bias factor correlation	-0.02	0.24	0.27	0.10	0.24	0.12	0.25	0.27
Geomin ($\epsilon = 0.5$)	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	0.99	1.00	0.99	1.00	1.00	1.00	1.00	0.99
	Bias main loadings	-0.03	0.00	0.02	-0.01	0.01	-0.00	0.01	0.02
	Bias cross-loadings	0.06	-0.01	-0.04	0.02	-0.01	0.00	-0.03	-0.04
	Bias factor correlation	-0.17	0.02	0.14	-0.07	0.02	-0.01	0.08	0.13
Facparsim	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Bias main loadings	-0.03	-0.00	0.01	-0.02	-0.00	-0.01	0.00	0.01
	Bias cross-loadings	0.07	0.01	-0.03	0.04	0.01	0.02	-0.01	-0.03
	Bias factor correlation	-0.20	-0.03	0.09	-0.12	-0.03	-0.06	0.02	0.08
CF-Varimax	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Bias main loadings	-0.03	0.00	0.01	-0.02	-0.00	-0.01	0.01	0.01
	Bias cross-loadings	0.06	-0.00	-0.04	0.04	0.01	0.02	-0.01	-0.03
	Bias factor correlation	-0.18	-0.00	0.11	-0.11	-0.02	-0.05	0.04	0.09
$Ridge_{\Lambda}$	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	0.94	0.90	0.99	0.83	0.84	0.87	0.89	0.97
	Bias main loadings	-0.07	-0.09	-0.01	-0.15	-0.13	-0.11	-0.09	-0.02
	Bias cross-loadings	0.08	0.06	-0.01	0.10	0.08	0.07	0.04	-0.01
	Bias factor correlation	-0.21	-0.04	0.07	-0.12	-0.04	-0.07	0.01	0.07
$Ridge_{\Lambda,\Phi}$	SRMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Congruencies	0.94	0.90	0.99	0.83	0.85	0.87	0.90	0.97
	Bias main loadings	-0.07	-0.09	-0.02	-0.15	-0.13	-0.11	-0.09	-0.02
	Bias cross-loadings	0.10	0.07	0.01	0.11	0.09	0.08	0.05	-0.00
	Bias factor correlation	-0.23	-0.10	0.01	-0.15	-0.07	-0.09	-0.02	0.04
$Lasso_{\Lambda}$	SRMR	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Congruencies	1.00	0.99	0.99	1.00	0.99	1.00	0.99	0.98
	Bias main loadings	-0.00	0.05	0.04	0.00	0.05	-0.00	0.04	0.04
	Bias cross-loadings	0.00	-0.07	-0.07	-0.00	-0.07	-0.01	-0.07	-0.08
	Bias factor correlation	-0.02	0.18	0.22	0.01	0.18	0.01	0.19	0.22
$Lasso_{\Lambda,\Phi}$	SRMR	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Congruencies	1.00	0.99	0.99	1.00	0.99	1.00	0.99	0.99
	Bias main loadings	-0.01	0.03	0.03	-0.00	0.04	-0.00	0.03	0.03
	Bias cross-loadings	0.01	-0.05	-0.06	-0.00	-0.06	-0.00	-0.06	-0.07
	Bias factor correlation	-0.02	0.13	0.19	0.01	0.17	-0.00	0.16	0.20
$Enet_{\Lambda}$	SRMR	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01
	Congruencies	1.00	0.99	0.99	1.00	0.99	1.00	0.99	0.99
	Bias main loadings	-0.00	0.05	0.03	0.00	0.05	0.00	0.04	0.04
	Bias cross-loadings	0.00	-0.07	-0.07	-0.00	-0.06	-0.00	-0.07	-0.07
	Bias factor correlation	-0.01	0.18	0.19	0.01	0.17	0.00	0.20	0.21
$Enet_{\Lambda,\Phi}$	SRMR	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.02
	Congruencies	1.00	0.99	0.99	1.00	0.99	1.00	0.99	0.99
	Bias main loadings	-0.01	0.03	0.02	-0.00	0.03	-0.00	0.03	0.03
	Bias cross-loadings	0.01	-0.05	-0.06	-0.00	-0.05	0.00	-0.05	-0.07
	Bias factor correlation	-0.02	0.13	0.17	0.00	0.14	-0.00	0.14	0.19

Note. The index symbols indicate which parameters were penalized with Λ = Factor loading matrix, and Φ = Factor correlation matrix.

performance of the *Geomin* rotation, resulting in much weaker biases for Approximate Simple and Complex patterns but performing worse for Simple patterns where it introduced spurious cross-loadings and underestimated factor correlations. A similar trend as for *Geomin* ($\epsilon = 0.5$) was observed for *Facparsim* and *CF-Varimax* rotations but both were slightly less biased for Approximate Simple and Complex factor patterns and more biased for Perfect Simple patterns.

Lasso and enet regularization estimates (see Tab. A1 for an overview of the enet β weights) were very similar to *Geomin* estimates. They perfectly recovered the Simple factor loading pattern, and resulted in underestimated cross-loadings and inflated factor correlations for Approximate Simple and Complex patterns. Notably, lasso and enet were less biased compared to *Geomin* ($\epsilon = 0.01$) but more biased than *Geomin* ($\epsilon = 0.5$) or *Facparsim* in the presence of cross-loadings. Ridge penalized EFAs behaved notably different yielding more biased estimates than all other tested methods for Perfect and Approximate Simple patterns. Complex factor patterns, however, were almost perfectly recovered by ridge regularization, especially when the factor correlations were included in the penalty term ($Ridge_{\Lambda,\Phi}$). Overall, the factor correlations tended to be less inflated for Complex factor patterns when they were included in the penalty term. However, the differences between regularization of only the factor loadings or factor loadings and factor correlations were rather small – except for the $Ridge_{\Lambda,\Phi}$.

Extended conditions. In general, we observed similar trends in the extended conditions, that is, the higher the cross-loadings of a respective pattern, the more difficult it was for the majority of the investigated methods to recover the population pattern. Especially, *Geomin* ($\epsilon = 0.01$) rotation yielded severely biased estimates in all extended conditions. As in the basic conditions, cross-loadings were underestimated and factor correlations overestimated, and this pattern was more pronounced the higher the cross-loadings. Notably, we observed that even a set of additional items with perfect simple structure (Extended Simple 1, Extended Complex 1) did not reduce the biases sufficiently for *Geomin* ($\epsilon = 0.01$). *Geomin* ($\epsilon = 0.5$) and *Facparsim* were clearly less biased in the presence of cross-loadings but – as in the basic conditions – suffered from spurious cross-loadings in the simpler conditions (Extended Simple 1).

Unlike all investigated rotation techniques, lasso and enet penalties recovered the factor loading patterns in the extended conditions perfectly if the additional variables followed a perfect simple structure (Extended Simple 1, Extended Complex 1). In all other extended conditions, lasso and enet yielded similar estimates as in the basic Complex condition, that is, cross-loadings and factor correlations were biased to a similar degree. In contrast to enet and lasso, ridge penalties resulted in fairly distorted estimates in the extended conditions. Remarkably, the ridge penalty additionally distorted the main loadings by a substantial amount.[2]

Discussion

In the first simulation study, we compared the asymptotic performances of regularized EFAs and traditional rotated EFAs. For factor rotation, we replicated previous simulation results indicating that the performance of factor rotation depends on the combination of factor rotation technique, rotation parameter (here: ϵ) and population factor loading pattern. In line with previous notions (Morin et al., 2013), a modified *Geomin* ($\epsilon = 0.5$) criterion and $Facparsim$ rotation performed especially well in conditions with moderate to high variable complexity but performed poorly in conditions with simple structure. This was also the case for oblique *CF-Varimax* rotation. The performance of regularized EFA depended largely on the choice of the penalty function: Lasso and enet recovered the population pattern if it contained a sufficient amount of zero-loadings and otherwise yielded similar estimates as factor rotation. Overall ridge penalties were less successful than lasso and enet. While ridge penalties were superior in some selected conditions, they resulted in severe distortions in some other conditions.

In sum, the advantages of ridge for complex loading patterns were – by far – outweighed by the distortions in other conditions and the inability to recover simple structure in the population. Enet and lasso, however, recovered simple structure where

[2]Some readers may wonder if the ridge estimates were simply over-shrunken, explaining the low congruencies. This was *not* the case as indicated by low Pearson correlations between the ridge estimates and the population pattern ($0.65 < r_{Pearson} < .9$).

it existed and were able to handle the extended conditions in which additional simple structure items were appended. These results are in line with previous studies showing that lasso (but not ridge) asymptotically selects the correct subset of variables (here, items as indicators of the factors) if the sparsity assumption holds, that is, if a sufficient proportion of the parameters to be estimated is zero in the population (Donoho, 2006; Donoho & Stodden, 2006). Put simply, lasso and enet strive for the sparsest parameter matrix and do not distinguish between factor and variable complexity – which enables them to recover the extended factor loading patterns in which there was high variability with respect to variable complexity. Across all conditions, an enet penalty on factor loadings and factor correlations showed slightly better performance than a simple lasso penalty, indicating that it combined the strengths of lasso and ridge to some extent.

With respect to the question if regularization can is a suitable alternative to factor rotation, the present results are promising because, even in the conditions where the sparsity assumption was violated, enet was able to match up with factor rotation techniques – outperforming the traditional *Geomin* ($\epsilon = 0.01$) rotation in every single of the tested conditions. Compared to the modified *Geomin* ($\epsilon = 0.5$) and $Facparsim$ rotations, enet performed worse for patterns with moderate cross-loadings (Approximate Simple) and comparably for patterns with high cross-loadings (Complex). However, it should be considered that both rotations achieved their partial superiority (similarly to ridge regularization) at the cost of an inability to recover the loadings of simple structure factor patterns. In practice (i.e., without a known ground truth), an informed choice of the rotation criterion is not possible. From that perspective, enet regularized EFA has reasonable all-round properties without the necessity of additional subjective choices.

Although these results were convincing in favor of regularization, it should be acknowledged that they were obtained from unrealistically large samples. This is especially relevant for regularized EFA because the penalty is considered directly in the estimation step, and estimation performance largely depends on sample size. In order to conclude that regularization is a suitable alternative to factor rotation, it needs to be established whether these results also hold for more practically common samples sizes. Apart from that, an open question is how factor rotation and regularization compare with respect to the stability (i.e., standard errors) of the parameter estimates. We addressed these questions in the second simulation study.

Study 2: Small-sample performance

In this simulation, we compared the performances of factor rotation and regularization for more realistic sample sizes ($N = 100$, $N = 200$). In order to achieve a feasible simulation time, we focused on the basic conditions from Schmitt and Sass (2011) and the extended conditions with partial simple structure. These conditions were chosen in order to investigate if the performance advantages of regularized EFA from Study 1 are preserved in realistic samples. In addition, we only used the two most successful regularization methods ($Enet_{\Lambda}$ & $Enet_{\Lambda,\Phi}$) from Study 1. These were compared with both versions of the *Geomin* rotation and *CF-Varimax rotation*. We evaluated the average recovery of the factor loading pattern and their empirical standard errors across samples.

Methods

Simulation Model. The same simulation model was used as in the first simulation but we reduced the number of conditions to keep the simulation feasible. Specifically, we included only the basic conditions and the extended conditions with partial simple structure (i.e., Extended Simple 1 and Extended Complex 1; Tab. 2).

Procedure. The procedure in Study 2 differed from Study 1 in two aspects: First, instead of simulating one large sample, we drew $N_{rep} = 500$ random samples from a multivariate normal distribution with either $N = 100$ or $N = 200$ observations per sample. Second, we only applied *Geomin* ($\epsilon = 0.01$ or 0.5) and *CF-Varimax* rotated ML-EFA and enet penalized regularized EFA. For the penalty weight α, a grid of 10 values starting from $\alpha = 0.001$ with a step size of 10^{-4} was used. For the enet weight β, we tested values between 0 and 1 with a step size of 0.1, that is, enet was allowed to result in pure lasso or ridge penalties if this optimized the sample BIC.

Dependent Measures. For each sample, sign and order indeterminacies of EFA estimates were taken into account and the same dependent measures as in Study 1 were calculated. We report means and standard deviations across all samples for all measures.

Results

The main results of the simulation are summarized in Table 4. The average estimated factor loadings and factor correlations for each condition are available from the OSF. ML-EFA converged normally in all samples. Regularized EFA converged normally in all but 2 samples with $N = 100$ in which $Enet_{\Lambda}$ penalized solutions were improper (factor correlations > 1). For these samples, the solution with the tuning parameters (α, β) entered the results that yielded the next best BIC.

Overall, the model-implied covariance matrices fit the observed covariance matrices of the data very well as indicated by the SRMR. The fit was slightly worse for smaller samples and for regularized EFA compared to rotated EFA. The congruencies and biases of the main loadings indicated that all investigated methods recovered the general factor pattern sufficiently. Factor recovery and stability were marginally better for $N = 200$ than for $N = 100$. For the factor loadings, the stability of the estimates did not differ substantially between rotated and regularized solutions. For the factor correlations, *Geomin* ($\epsilon = 0.5$) and *CF-Varimax* yielded smaller standard errors than the other methods. Notably, unlike in Study 1, none of the tested methods was able to perfectly recover the Perfect Simple pattern but rather yielded underestimated factor correlations and spurious cross-loadings. Across all conditions and methods, the biases of factor correlations and cross-loadings were almost perfectly correlated, $r_{spearman} = -0.98$.

Table 4

Simulation results for all dependent measures as a function of estimation method and simulation condition in Study 2

N	Method	Measure	Perfect Simple	Approximate Simple	Complex	Extended Simple 1	Extended Complex 1
100	*Geomin* ($\epsilon = 0.01$)	SRMR	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)
		Congruencies	0.99 (0.01)	0.97 (0.01)	0.96 (0.04)	0.98 (0.00)	0.98 (0.00)
		Bias main loadings	-0.01 (0.02)	0.05 (0.02)	0.05 (0.05)	-0.01 (0.02)	0.03 (0.01)
		Bias cross-loadings	0.02 (0.01)	-0.07 (0.01)	-0.08 (0.03)	-0.01 (0.01)	-0.04 (0.01)
		Bias factor correlation	-0.06 (0.07)	0.17 (0.08)	0.21 (0.13)	0.07 (0.07)	0.10 (0.05)
	Geomin ($\epsilon = 0.5$)	SRMR	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)
		Congruencies	0.98 (0.01)	0.99 (0.01)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.03 (0.02)	0.00 (0.02)	0.02 (0.02)	0.01 (0.02)	-0.00 (0.01)
		Bias cross-loadings	0.06 (0.01)	-0.00 (0.01)	-0.04 (0.01)	0.01 (0.01)	0.00 (0.01)
		Bias factor correlation	-0.17 (0.05)	0.00 (0.06)	0.12 (0.05)	-0.08 (0.05)	-0.03 (0.04)
	CF-Varimax	SRMR	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)
		Congruencies	0.98 (0.01)	0.99 (0.01)	0.99 (0.00)	0.98 (0.00)	0.99 (0.00)
		Bias main loadings	-0.03 (0.02)	-0.00 (0.02)	0.01 (0.02)	0.02 (0.02)	-0.01 (0.01)
		Bias cross-loadings	0.06 (0.01)	0.00 (0.01)	-0.03 (0.01)	0.02 (0.01)	0.02 (0.01)
		Bias factor correlation	-0.18 (0.05)	-0.02 (0.06)	0.10 (0.05)	-0.11 (0.05)	-0.06 (0.04)
	$Enet_{\Lambda}$	SRMR	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)
		Congruencies	0.98 (0.01)	0.98 (0.01)	0.97 (0.01)	0.99 (0.00)	0.99 (0.01)
		Bias main loadings	-0.03 (0.02)	0.02 (0.02)	0.03 (0.02)	-0.00 (0.02)	-0.00 (0.01)
		Bias cross-loadings	0.04 (0.01)	-0.04 (0.02)	-0.07 (0.02)	-0.00 (0.01)	-0.01 (0.01)
		Bias factor correlation	-0.11 (0.07)	0.10 (0.08)	0.19 (0.07)	-0.01 (0.08)	0.01 (0.06)
	$Enet_{\Lambda,\Phi}$	SRMR	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.03 (0.00)
		Congruencies	0.98 (0.02)	0.97 (0.03)	0.98 (0.02)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.03 (0.03)	-0.00 (0.03)	0.01 (0.02)	-0.00 (0.02)	-0.01 (0.01)
		Bias cross-loadings	0.05 (0.02)	-0.02 (0.02)	-0.05 (0.03)	0.00 (0.01)	-0.00 (0.01)
		Bias factor correlation	-0.15 (0.08)	0.01 (0.12)	0.12 (0.10)	-0.04 (0.07)	-0.02 (0.06)
200	*Geomin* ($\epsilon = 0.01$)	SRMR	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)	0.03 (0.00)	0.02 (0.00)
		Congruencies	0.99 (0.00)	0.98 (0.00)	0.97 (0.01)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.01 (0.01)	0.06 (0.01)	0.06 (0.01)	-0.01 (0.01)	0.03 (0.01)
		Bias cross-loadings	0.01 (0.00)	-0.08 (0.01)	-0.10 (0.01)	-0.02 (0.00)	-0.04 (0.01)
		Bias factor correlation	-0.04 (0.07)	0.22 (0.05)	0.24 (0.06)	0.07 (0.04)	0.13 (0.05)
	Geomin ($\epsilon = 0.5$)	SRMR	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)	0.03 (0.00)	0.02 (0.00)
		Congruencies	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.03 (0.01)	-0.00 (0.01)	0.02 (0.01)	0.01 (0.01)	-0.01 (0.01)
		Bias cross-loadings	0.06 (0.01)	-0.00 (0.01)	-0.04 (0.01)	0.01 (0.01)	0.01 (0.01)
		Bias factor correlation	-0.17 (0.05)	0.03 (0.04)	0.12 (0.04)	-0.09 (0.03)	-0.01 (0.04)
	CF-Varimax	SRMR	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)	0.03 (0.00)	0.02 (0.00)
		Congruencies	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.04 (0.01)	-0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	-0.01 (0.01)
		Bias cross-loadings	0.06 (0.01)	0.01 (0.01)	-0.04 (0.01)	0.02 (0.01)	0.02 (0.01)
		Bias factor correlation	-0.18 (0.04)	0.01 (0.04)	0.10 (0.04)	-0.12 (0.03)	-0.05 (0.03)
	$Enet_{\Lambda}$	SRMR	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)
		Congruencies	0.99 (0.00)	0.98 (0.01)	0.98 (0.03)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.02 (0.01)	0.03 (0.01)	0.03 (0.02)	-0.01 (0.01)	-0.00 (0.01)
		Bias cross-loadings	0.03 (0.01)	-0.05 (0.01)	-0.07 (0.01)	-0.01 (0.01)	-0.01 (0.01)
		Bias factor correlation	-0.09 (0.06)	0.16 (0.06)	0.18 (0.06)	-0.01 (0.04)	0.03 (0.06)
	$Enet_{\Lambda,\Phi}$	SRMR	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)
		Congruencies	0.99 (0.00)	0.97 (0.05)	0.98 (0.01)	0.99 (0.00)	0.99 (0.00)
		Bias main loadings	-0.03 (0.01)	0.00 (0.03)	0.02 (0.01)	-0.00 (0.01)	-0.01 (0.01)
		Bias cross-loadings	0.04 (0.01)	-0.02 (0.02)	-0.06 (0.01)	-0.00 (0.01)	-0.00 (0.01)
		Bias factor correlation	-0.12 (0.07)	0.05 (0.15)	0.14 (0.07)	-0.03 (0.05)	0.00 (0.05)

Note. The index symbols indicate which parameters were penalized with Λ = Factor loading matrix, and Φ = Factor correlation matrix. The numbers outside the parentheses give the averages across all samples of the respective measure, the numbers in parentheses give the standard deviation across all samples of the respective measure.

Both *Geomin* and *CF-Varimax* rotations resulted in underestimated cross-loadings and inflated cross-loadings for complex factor loading patterns. As in Study 1, *Geomin* ($\epsilon = 0.5$) and *CF-Varimax* perfectly recovered the Approximate Simple pattern whereas *Geomin* ($\epsilon = 0.01$) yielded similar (but slightly weaker) distortions as in the Complex condition. The enet regularization behaved very similar to the *Geomin* rotations. The inclusion of the factor correlation in the penalty term was more influential than in Study 1. Specifically, $Enet_{\Lambda,\Phi}$ like *Geomin* ($\epsilon = 0.5$) yielded less distorted estimates than $Enet_{\Lambda}$ in the presence of cross-loadings (Approximate Simple & Complex) but performed worse for Perfect Simple patterns. In the extended conditions, regularized EFA was superior to rotated EFA as indicated by (nearly) unbiased estimates (especially for the factor correlation). Altogether, the performance of factor rotation and regularization were very similar with respect to both accuracy and stability of the factor solutions.

Discussion

In this simulation, we investigated whether regularization yields comparable results as factor rotation for realistic sample sizes. Overall, the results of Study 1 generalized quite well to the small sample case and all investigated methods recovered the factor loading pattern with reasonable accuracy (Tucker's congruencies > 0.95; Lorenzo-Seva & ten Berge, 2006) and yielded reasonable fit (SRMR < 0.05 Bentler & Yuan, 1999). In contrast to Study 1, none of the investigated methods was able to recover Perfect Simple patterns without biases. This was almost negligible for *Geomin* ($\epsilon = 0.01$) but all other methods underestimated the factor correlations to a considerable extent. As in study 1, only regularized EFA was able to recover the parameters in the extended conditions.

Despite additional small sample biases, the present results support the notion that (enet) regularization may be used to estimate EFA without a separate rotation step. Differences between *Geomin* rotated and enet regularized estimates with respect to accuracy and stability were rather small. As in Study 1, both methods tended to oversimplify loading patterns with high cross-loadings, resulting in inflated factor correlations. We note that the general size of the biases was rather small, especially for the factor loadings, and would arguably not lead to fundamentally different interpretations of the factor loading pattern (e.g., a different selection of items). With respect to factor correlation bias, we note that interpretations of factor correlations should acknowledge the trade-off between the size of the cross-loadings and the size of the factor correlation that is inherent to all oblique factor analysis methods. In sum, we conclude regularization is a suitable alternative to the traditional factor rotation approach even in the case of small samples.

General Discussion

In two simulation studies, we compared the performance of factor rotation and regularization in recovering pre-defined factor loading patterns that resembled typical psychometric data sets. Both with respect to asymptotic (Study 1) and finite sample performance (Study 2), regularization resulted in similar estimates as factor rotation. In line with previous notions (Morin et al., 2013), *Geomin* rotation with an increased rotation parameter $\epsilon = 0.5$ was superior for factor loading patterns with substantial cross-loadings but was unable to recover perfect simple structures. An enet penalty on both factor loadings and factor correlations showed the best overall performance among the investigated regularization methods and provided reasonable balance between the ability to recover simple structure, if it exists, and the ability to handle complex loading patterns.

We set out to investigate whether regularization is a suitable alternative to simple structure rotation in EFA. With respect to the estimation performance, our results confirm that regularization is a viable alternative to factor rotation. This is not to say that regularization is always the better approach to estimate the EFA parameters but, across all conditions in our simulations, it performed very well compared to the arguably best among the commonly used rotation techniques (cf. Asparouhov & Muthén, 2009; Schmitt & Sass, 2011). In addition, regularization was able to recover the EFA parameters when only a subset of the items followed a simple structure – where factor rotation failed to do so (cf. Extended Simple/Complex 1 in Study 1). Thus, for typical psychometric data sets it can be expected that the results of regularized EFA match the results of simple structure rotated EFA very well.

Despite the promising performance of regularized EFA, the method is not without limitations: First, the penalization approach implemented in regSEM is very simplistic because all penalized parameters are treated equally, no matter whether they are factor loadings, structural coefficients or factor (co-)variances (or correlations). This could be a drawback because the number of factor loadings always exceeds the number of factor correlations or structural parameters; hence, the factor loadings will have a stronger relative influence on the estimation process (cf. Jacobucci et al., 2016; Jacobucci, 2017, for technical details). Second, the lasso and enet penalty (just as factor rotation) oversimplified complex loading patterns. It is well-known that lasso can result in overshrinkage of parameter estimates (here: cross-loadings), therefore, alternative methods for obtaining the final parameters may be considered (see Jacobucci et al., 2016, for a discussion). Lastly, our simulations were limited to the cases where all assumptions of ML-EFA and regSEM were fulfilled. Further research is needed to assess the sensitivity of regSEM to violations of distributional assumptions and model misspecification (e.g., correlated residuals). Especially the use of Likert scales may have profound consequences on the estimation performance (DiStefano, 2002; DiStefano & Morgan, 2014). Moreover, in the light of the increasing availability of large data sets (e.g., Kosinski, Wang, Lakkaraju, & Leskovec, 2016), future investigations should also consider conditions in which the number of variables exceeds the number of observations.

Taken together, the present and previous research (e.g., Trendafilov, 2014; Yamamoto et al., 2017) suggest that, regarding estimation performance, there is not much to lose when replacing factor rotation with regularization but potential gains are also rather small (except for situations as in our extended conditions with partial simple structure). Therefore, applied readers may wonder why they should consider replacing well-established rotation procedures with regularization. We think that regularization has advantages both at a pragmatic and at a conceptual level: First, from a pragmatic perspective, a general use of regularization over factor rotation considerably alleviates the subjectivity of the analyses. Admittedly, researchers must still choose a penalty function just as they have to choose a rotation technique – which cannot be optimally done without a known ground truth – but the elastic net may provide a reasonable default choice and at least the tuning parameter can be determined in an objective and (nearly) automatized way.

At a conceptual level, regularization offers a generalization that subsumes EFA and confirmatory factor analysis (CFA). These methods are often considered separate methods that have different purposes. Using the penalized likelihood approach, however, both are simply two extremes in the space of possible models - differing only in which parameters enter the penalty term $P(\theta)$ in Equation 8. EFA, on the one hand, has measurement models where all paths are allowed and all paths are considered to-be-regularized. CFA, on the other hand, has measurement models where only the theoretically motivated paths are estimated, all other paths are constrained to zero, and no parameter enters the penalty term. As a generalized method that contains both EFA and CFA as special cases, regularization allows all possible variations in between these two extremes. In particular, it has been considered to specify the theoretically motivated main loading paths but to not include them into the penalty term and only regularize all cross-loadings. Such a *semi-confirmatory* approach allows researchers to specify a model that considers there prior beliefs about the factor structure but that does not completely depend on the validity of these beliefs (Huang et al., 2017b).

Beyond basic factor analysis applications, the present findings have implications for the more generalized Exploratory Structural Equation modeling (ESEM) approach in which factor rotation is an important step as well (Asparouhov & Muthén, 2009). ESEMs extent EFA by a structural model of the latent variables. As we operationalized regularized EFA from a regSEM perspective, regularized EFA can be easily extended to a regularized ESEM by (enet) penalizing the factor loadings and correlations of the exploratory factors. Such a regularized ESEM may be used to address similar research questions as ESEM (see also Huang et al., 2017a). Our results on factor rotation and regularization in the context of EFA should in principle apply for a comparison of ESEM and regularized ESEM as well. That is, the more complex the exploratory measurement model, the more rotation and regularization will underestimate the cross-loadings and overestimate the factor correlations. Consequently, the structural parameters may be distorted as well (Mai, Zhang, & Wen, 2018). Nevertheless, future research should investigate this notion empirically and investigate how additional regularization of the structural parameters affects the solutions.

Considering this connection, some of the central limitations of ESEMs may be solved in the regSEM framework. For instance, in ESEM, it is mandatory that the relationships between all or none of the exploratory factors with a given predictor are specified (otherwise, the rotation procedure is not valid, see Asparouhov & Muthén, 2009). RegSEM is much more flexible in that respect, allowing researchers to choose which parameters of the model should be penalized and also add restrictions such as equality constraints to the model. Consequently, regSEM obviates the need for work-arounds to ESEM limitations such as the ESEM within Confirmatory Factor Analysis approach (Morin et al., 2013). In this context, it would, for instance, be interesting to extend regSEM to the multi-level framework (e.g., Asparouhov & Muthén, 2012; Rabe-Hesketh, Skrondal, & Zheng, 2007). All in all, regSEM offers a consistent translation of the rotation problem into an estimation problem – allowing for a unified framework of both confirmatory and exploratory techniques.

Future research should directly compare the performance of regSEM with other methods of semi-automatic model specification such as specification search in order to develop recommendations which method is appropriate for which purposes. In this context, it should be noted that regSEM and specification search are closely related to Bayesian Structural Equation Modeling (BSEM, e.g., Muthén & Asparouhov, 2011) because best subset selection and common penalty functions may be seen as a special case of BSEM with specific priors (Hastie et al., 2009; Jacobucci & Grimm, 2018).

Apart from providing a unifying theoretical framework, the connection to BSEM offers a range of possibilities for improvements because different priors (i.e., penalty functions) could be placed on different parameters. For instance, one could use different priors on the factor loadings and factor correlations, respectively, in order to achieve a better balance between cross-loadings and factor correlations for complex loading patterns.

Conclusion

Regularization is an estimation method for complex statistical models with increasing popularity among social scientists. In two simulation studies, we compared the estimates of the traditional rotated EFA approach and regularized EFA for realistic factor loading patterns with varying complexity. Regularized EFA performed very similar to common factor rotation techniques in the majority of the considered conditions, indicating that regularization is a suitable alternative to the traditional rotation approach. Although regularized EFA was not unequivocally the best method across all conditions, the increased objectivity and the relation of the underlying regSEM to wider statistical frameworks such as ESEM and BSEM make it a valuable tool to be considered by social scientists.

Acknowledgements

We thank Jana Pförtner for her assistance in preparing the simulation study and for proof-reading an earlier draft of the manuscript. We also thank Richard Rau for his valuable comments on an earlier draft of this manuscript.

Appendix

Elastic net weights β for Study 1

Table A1

Elastic net weights β for all conditions in Study 1

Condition	$enet_{\Lambda}$	$enet_{\Lambda,\Phi}$
Perfect Simple	0.40	0.15
Approximate Simple	0.15	0.40
Complex	0.10	0.15
Extended Simple 1	0.20	0.65
Extended Simple 2	0.80	0.65
Extended Complex 1	0.30	0.70
Extended Complex 2	0.15	0.75
Extended Complex 3	0.65	0.10

Note. The index symbols indicate which parameters were penalized with Λ = Factor loading matrix, and Φ = Factor correlation matrix.

References

Arruda, E. H. & Bentler, P. M. (2017). A Regularized GLS for Structural Equation Modeling. *Structural Equation Modeling*, *24*, 657–665. doi:10.1080/10705511.2017.1318392

Asparouhov, T. & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. doi:10.1080/10705510903008204

Asparouhov, T. & Muthén, B. (2012). General random effect latent variable modeling : Random subjects, items, contexts, and parameters. Retrieved from http://www.statmodel.com/download/NCME12.pdf

Asparouhov, T. & Muthén, B. (2018). SRMR in Mplus. Retrieved from http://www.statmodel.com/download/SRMR2.pdf

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. *Journal of Neuroscience Methods*, *295*, 20–36. doi:10.1016/j.jneumeth.2017.11.011

Bentler, P. M. & Yuan, K.-h. (1999). Structural Equation Modeling with Small Samples : Test Statistics. *Multivariate Behavioral Research1*, *34*, 181–197. doi:10.1207/S15327906Mb340203

Bernaards, C. A. & Jennrich, R. I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Browne, M. W. (2001). An overview of analytic rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR360105

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., & West, M. (2008). High-Dimensional Sparse Factor Modelling : Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, *103*, 1438–1456. doi:10.1198/016214508000000869

Choi, J., Oehlert, G., & Zou, H. (2010). A penalized maximum likelihood approach to sparse factor analysis. *Statistics and Its Interface*, *3*, 429–436. doi:10.4310/SII.2010.v3.n4.a1

DiStefano, C. (2002). The Impact of Categorization With Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 327–346. doi:10.1207/S15328007SEM0903_2

DiStefano, C. & Morgan, G. B. (2014). A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling*, *21*, 425–438. doi:10.1080/10705511.2014.915373

Donoho, D. & Stodden, V. (2006). Breakdown Point of Model Selection When the Number of Variables Exceeds the Number of Observations. In *The 2006 ieee international joint conference on neural network proceedings* (pp. 1916–1921). IEEE. doi:10.1109/IJCNN.2006.246934

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, *59*, 797–829. doi:10.1002/cpa.20132. arXiv: 0912.3599

Ertel, S. (2011). Exploratory factor analysis revealing complex structure. *Personality and Individual Differences*, *50*, 196–200. doi:10.1016/j.paid.2010.09.026

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272. arXiv: 99 [1082-989X]

Fan, J. & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties Variable Selection via Nonconcave Penalized Likelihood and its Oracle Prope. *Source Journal of the American Statistical Association*, *96*, 1348–1360. doi:10.1198/016214501753382273

Genz, A., Bretz, F., Tetsuhisa, M., Xuefei, M., Leisch, F., Scheipl, F., & Hothorn, T. (2017). mvtnorm: Multivariate normal and t distributions. Retrieved from http://cran.r-project.org/package=mvtnorm

Glymour, C., Madigan, D., Pregibon, D., & Smyth, P. (1997). Statistical themes and lessons for data mining. *Data Mining and Knowledge Discovery*, *1*, 11–28. doi:10.1023/A:1009773905005

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. *Elements*, *1*, 337–387. doi:10.1007/b94608

Hendrickson, A. E. & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70. doi:10.1111/j.2044-8317.1964.tb00244.x

Hirose, K. (2016). Simple structure estimation via prenet penalization. arXiv: 1607.01145. Retrieved from http://arxiv.org/abs/1607.01145

Hirose, K. & Konishi, S. (2012). Variable selection via the weighted group lasso for factor analysis models. *Canadian Journal of Statistics*, *40*, 345–361. doi:10.1002/cjs.11129

Hirose, K. & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, *79*, 120–132. doi:10.1016/j.csda.2014.05.011. arXiv: 1302.5475

Hirose, K. & Yamamoto, M. (2015a). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, *25*, 863–875. doi:10.1007/s11222-014-9458-0. arXiv: 1205.5868

Hirose, K. & Yamamoto, M. (2015b). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, *25*, 863–875. doi:10.1007/s11222-014-9458-0. arXiv: 1205.5868

Hoerl, A. E. & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*, 55. doi:10.2307/1267351

Huang, P.-H., Chen, H., & Weng, L.-J. (2017a). A Penalized Likelihood Method for Structural Equation Modeling. *Psychometrika*, *82*, 329–354. doi:10.1007/s11336-017-9566-9

Huang, P.-H., Chen, H., & Weng, L.-J. (2017b). A Penalized Likelihood Method for Structural Equation Modeling. *Psychometrika*, *82*, 329–354. doi:10.1007/s11336-017-9566-9

Jacobucci, R. (2017). regsem: Regularized Structural Equation Modeling, 1–13. arXiv: 1703.08489. Retrieved from http://arxiv.org/abs/1703.08489

Jacobucci, R. & Grimm, K. J. (2018). Comparison of Frequentist and Bayesian Regularization in Structural Equation Modeling. *Structural Equation Modeling*, *00*, 1–11. doi:10.1080/10705511.2017.1410822

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793. arXiv: 15334406

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer Texts in Statistics. New York, NY: Springer New York. doi:10.1007/978-1-4614-7138-7

Jennrich, R. I. (2004). Derivative free gradient projection algorithms for rotation. *Psychometrika*, *69*, 475–480. doi:10.1007/BF02295647

Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, *71*, 173–191. doi:10.1007/s11336-003-1136-B

Jin, S., Moustaki, I., & Yang-Wallentin, F. (2018). Approximated Penalized Maximum Likelihood for Exploratory Factor Analysis: An Orthogonal Case. *Psychometrika*, *83*, 628–649. doi:10.1007/s11336-018-9623-z

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202. doi:10.1007/BF02289343

Jung, S. & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior research methods*, *43*, 701–9. doi:10.3758/s13428-011-0077-9

Jung, S. & Takane, Y. (2007). Regularized common factor analysis. *New Trends in Psychometrics*, *1*, 1–10.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. doi:10.1007/BF02289233

Kaiser, H. F. (1959). Computer Program for Varimax Rotation in Factor Analysis. *Educational and Psychological Measurement*, *19*, 413–420. doi:10.1177/001316445901900314

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, *21*, 493–506. doi:10.1037/met0000105. arXiv: arXiv:1011.1669v3

Lorenzo-Seva, U. & ten Berge, J. M. F. (2006). Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology*, *2*, 57–64. doi:10.1027/1614-2241.2.2.57

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 737–749. doi:10.1080/10705511.2018.1444993

Marcoulides, G. A. & Drezner, Z. (2003). Model Specification Searches Using Ant Colony Optimization Algorithms. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 154–164. doi:10.1207/S15328007SEM1001_8

Marcoulides, G. A., Drezner, Z., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using tabu search. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*, 365–376. doi:10.1080/10705519809540112

Marcoulides, G. A. & Ing, M. (2013). The Use of Generalizability Theory in Language Assessment. In *The companion to language assessment* (pp. 1207–1223). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118411360.wbcla014

Marcoulides, K. M. & Falk, C. F. (2018). Model Specification Searches in Structural Equation Modeling with R. *Structural Equation Modeling: A Multidisciplinary Journal*, *00*, 1–8. doi:10.1080/10705511.2017.1409074

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. doi:10.1037/a0019227

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 439–476. doi:10.1080/10705510903008220

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling: an introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.). Charlotte, NC: Information Age Publishing, Inc.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). CRC press.

Muthén, B. O. (2004). *Mplus Technical Appendices*. Muthén & Muthén. Los Angeles, CA. Retrieved from https://www.statmodel.com/download/techappen.pdf

Muthén, B. & Asparouhov, T. (2011). Bayesian SEM : A more flexible representation of substantive theory Web tables. *Psychological methods*, *17*, 313–335. doi:10.1037/a0026802

Ning, L. & Georgiou, T. T. (2011). Sparse factor analysis via likelihood and L1-regularization. *Proceedings of the IEEE Conference on Decision and Control*, 5188–5192. doi:10.1109/CDC.2011.6161415

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). *Multilevel Structural Equation Modeling*. Elsevier B.V. doi:10.1016/B978-044452044-9/50013-6

Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.

Sass, D. A. & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within Exploratory Factor Analysis. *Multivariate Behavioral Research*, *45*, 73–103. doi:10.1080/00273170903504810

Schmitt, T. A. & Sass, D. A. (2011). Rotation criteria and hypothesis testing for Exploratory Factor Analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*, 95–113. doi:10.1177/0013164410387348

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* Chicago: University of Chicago Press. doi:10.1037/10018-000

Thurstone, L. L. (1954). An analytical method for simple structure. *Psychometrika*, *19*, 173–182. doi:10.1007/BF02289182

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, *29*, 431–454. doi:10.1007/s00180-013-0434-5

Trendafilov, N. T. & Adachi, K. (2015). Sparse Versus Simple Structure Loadings. *Psychometrika*, *80*, 776–790. doi:10.1007/s11336-014-9416-y

Trendafilov, N. T., Fontanella, S., & Adachi, K. (2017). Sparse Exploratory Factor Analysis. *Psychometrika.* doi:10.1007/s11336-017-9575-8

Yamamoto, M., Hirose, K., & Nagata, H. (2017). Graphical tool of sparse factor analysis. *Behaviormetrika*, *44*, 229–250. doi:10.1007/s41237-016-0007-3

Yamamoto, M. & Jennrich, R. I. (2013). A cluster-based factor rotation. *British Journal of Mathematical and Statistical Psychology*, *66*, 488–502. doi:10.1111/bmsp.12007

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* Albany, NY, USA: State Univ. of New York Press.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*, 894–942. doi:10.1214/09-AOS729. arXiv: 1002.4734

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*, 301–320. doi:10.1111/j.1467-9868.2005.00503.x

Chapter 5

General Discussion

5.1 Summary of the main results

The aim of the present thesis was to explore how exploratory factor analytic methods can be made less prone to variance misallocation. These efforts resulted in a series of three publications in which variance misallocation in EFA was described as a consequence of the properties of ERP data (Chapter 2), ESEM was proposed as an extension of EFA that acknowledges the structure of ERP data sets (Chapter 3), and regularized estimation was suggested as an alternative to simple structure rotation with desirable properties when perfect simple structure cannot be expected in the population (Chapter 4). In the following, it will be discussed how these contributions extent the literature on variance misallocation in temporal EFA for ERP data.

In chapter 2, it was shown that variance misallocation can be the result of applying orthogonal factor rotation to an actually oblique factor pattern (orthogonality bias) and/or the result of simple structure rotation which biases the estimated factor loadings towards the optimum of its mathematical simplicity criterion (rotation bias). Most importantly, chapter 2 provided formal support for previous recommendations of oblique over orthogonal rotations in temporal EFA for ERP data (Dien, 1998; Dien et al., 2005). An analytical decomposition of the variance-covariance matrix of the latent factors revealed that it can be written as a sum of contributions due to electrodes, participants, and condition effects. Given that no more than two orthogonal factors are physiologically possible (Dien, 2010), and that the number of factors is typically found to be 8 and more for ERP data sets (e.g., Barry, De Blasio, Fogarty, & Karamacoska, 2016), it is highly unlikely for the (co-)variance contributions to sum up to zero and orthogonal rotation should be generally avoided. These results cleared remaining doubts about the general appropriateness of oblique rotation in temporal EFA for ERP data (Dien, 2006; Kayser & Tenke, 2003, 2006).

Another important consequence of the commingled (co-)variance sources (i.e., participants, electrodes, and conditions) is that the factor covariances may not be interpreted substantially without accounting for the commingled contributions. To resolve this problem, in chapter 3, it was proposed to replace EFA with an ESEM in which electrode site, condition, and their interactions are specified as predictors in the structural model. To the best of the author's knowledge, it has not been proposed to apply ESEM to ERP data previously. A simulation study confirmed that ESEM is capable of separating the (co-)variance contributions. In addition, it was shown that variance misallocation can also occur as a consequence of the factor scoring procedure that is necessary as an intermediate step in EFA but not in ESEM approach - although this possibility was considered negligible before (Dien & Frishkoff, 2005). In sum, ESEM has the potential to become a powerful and flexible tool for ERP researchers that is considerably less prone to variance misallocation.

Whereas ESEM solves the problems related to the commingled (co-)variance contributions in EFA, the rotation step remains an important part of the data analytic procedure, and therefore the choice of the rotation technique still has profound consequences. Even for the rather simplistic factor patterns presented in chapter 3, considerable orthogonality biases occurred when the factors were only mildly correlated. In the absence of a known ground truth, it is therefore recommend to estimate an oblique factor loading pattern. However, oblique simple structure rotation techniques are prone to certain biases when the temporal overlap is high (Dien, 1998; Dien et al., 2005; Sass & Schmitt, 2010; Schmitt & Sass, 2011). More specifically, the cross-loadings are underestimated at the cost of inflated factor correlation estimates. Therefore, an alternative to factor rotation that is more tolerant towards temporal overlap was explored in 4.

Chapter 4 investigated if regularized estimation is a suitable alternative to simple structure rotation. Elastic net regularized estimation was found to yield reasonable results irrespective whether the population factor loading pattern was simple or complex. Although the presented simulations focused on typical psychometric applications, the results suggest that regularization is also a suitable alternative to factor rotation for the analysis of ERP data. In this context, especially the presented extended conditions with partial simple structure are of interest because conditions with partial temporal overlap can be expected to occur regularly with ERP data sets. In addition, considering the factor correlation a to-be-regularized parameter offers the possibility to counterweigh the influence of temporal overlap on the estimated factor solution to some extent.

To sum up, the work presented in this thesis identified, first, the presence of multiple sources of (co-)variance, second, the factor scoring step and, third, high temporal overlap of the factors as major causes of variance misallocation in EFA for ERP data. ESEM was proposed as a remedy to the first and second point, and, regarding the third point, regularized estimation was proposed as a potential substitute for factor rotation that is better able to cope with the high prevalence of temporal overlap.

5.2 Implications and future research questions

The presented results imply that a combination of the ESEM approach and regularized estimation may provide a data analytic framework for ERP data that is less prone to all three causes of variance misallocation. Recently, *regularized structural equation modeling* (regSEM) has been proposed as a method that offers such a combination (Jacobucci et al., 2016). In the following, a temporal regSEM for ERP data will be outlined in more detail and discussed in the light of alternative suggestions to reduce the rotation bias. Finally, further research questions and potential extensions of ESEM/regSEM are discussed.

More specifically, an ESEM can be conceptualized as a regSEM by using the same structural model as in the ESEM and specifying a measurement model in which all factors are allowed to load on all observed variables. Crucially, as explained in more detail in chapter 4, all factor loadings enter the penalty term during estimation to achieve a sparse measurement model. How can regSEM be applied to ERP data? Building on the findings regarding ESEM in chapter 3, a structural model should be specified in which the indicator variables for electrode sites and conditions as well as their interactions predict the latent factors. The factors should be allowed to correlate because even after controlling for the factor topography, the biases resulting from a violated orthogonality assumption are profound (Chapter 3). Furthermore, all factors are allowed to load on each sampling point, and all factor loadings are considered to-be-regularized.[1] In addition, it may be considered to include the factor correlations into the penalty term as well to counteract the inflation of factor correlation estimates due to temporal overlap (Chapters 2 & 3). The findings in chapter 4 suggest that an elastic net penalty could provide a reasonable default choice for the penalty function. A regSEM specified in this way both properly considers the structure of ERP data sets and it should be less prone to variance misallocation than simple structure rotation. Despite the promising results reported here, a direct empirical test of this notion is a necessary task for future research.

It may be argued that the proposed regSEM addresses the problem of rotation bias only indirectly because it is maximally agnostic about the expected time courses of the latent factors. Alternatively, it was suggested to develop ERP-specific rotation criteria that either make use of a priori knowledge of condition effects (Beauducel & Leue, 2015), or that include assumptions regarding the time courses of the factors that are motivated directly from ERP research (Beauducel, 2018). Given that a priori knowledge of condition effects is rarely available, the latter approach is arguably more generally applicable than the former. Specifically, assuming that ERPs are *transient* voltage deflections, Beauducel (2018) proposed ERP-specific rotation techniques that take a two-step approach. First, an initial Varimax-rotated solution is found. Second, a rotation target is derived either by fixing successive loadings with small slopes to zero (Event-related orthogonal partial Procrustes rotation, EPP rotation) or by fitting an optimal Gaussian shape to the Varimax solution (Gaussian event-related Procrustes rotation, GEP rotation). The rotation target is then supplied to a target rotation algorithm yielding the final rotated solution. Both EPP and GEP rotation aim at reducing typical distortions of the factor loading estimates that have been observed in the literature on variance misallocation (Wood & McCarthy, 1984). Beauducel (2018) found that both EPP and GEP rotation were able to recover factor loading patterns that included slow-wave potentials (i.e., factors with non-zero loadings spread over the whole epoch) which are known to be a challenging ground for simple structure rotation techniques (Verleger & Möcks, 1987).

[1] In regSEM, the initial model may be over-parameterized. As long as a sufficient number of parameters is penalized, identification can still be achieved because some of parameters are shrunken to zero (Jacobucci, 2017; Jacobucci et al., 2016).

ERP-specific rotation criteria can easily be combined with the ESEM approach presented in chapter 3, offering an alternative approach to reduce the rotation bias within the ESEM framework. A direct comparison of ESEM with simple structure rotation, ESEM with ERP-specific rotation, and regSEM is required to explore which of the methods performs best under a variety of conditions. In order to judge in how far the methods are generally useful, it is crucial for such a comparison to include conditions that are challenging enough. For instance, on the one hand, slow-wave potentials may be challenging to recover for regSEM (just as for simple structure rotation) because the ideal factor solution may not be sparse any more. On the other hand, ERP-specific rotation may be challenged by factors with bi-phasic time courses because it assumes that the targeted time course shape is approximately correct (e.g., mono-phasic, transient factors for GEP rotation). These examples illustrate the importance of a representative choice of the simulation conditions because otherwise no generalizable conclusions can be drawn.

Apart from the question which of the methods proposed here has the best relative performance, an important question is how these methods perform when compared with other frequently used decomposition techniques such as ICA, multimode PCAs (Möcks, 1988), or temporospatial PCA (i.e., applying a spatial PCA to the results of a temporal PCA; Dien, 2010). Some efforts have been made in that direction (Bugli & Lambert, 2007; Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012; Dien, Khoe, & Mangun, 2007; Makeig et al., 1999; Verleger, Paulick, Möcks, Smith, & Keller, 2013), but these studies were often limited to real data examples making it hard to judge the generalizability (but see Groppe et al., 2008, for a notable exception). In this context, it is important to investigate the mathematical relationships between the underlying models in order to establish generalizable conclusions about their relative strengths and weaknesses. For instance, it would be interesting to learn under which conditions different methods can be expected to give (approximately) equivalent results.

Finally, it would be interesting to explore in how far some of the assumptions of the temporal EFA/ESEM approach can be relaxed. Arguably the most restrictive assumption is that the factor loading patterns are fixed across participants and conditions.[2] Especially, the consequences of latency-jitter (i.e., time-shifted factor loadings patterns between participants and conditions) have been investigated by previous research (Donchin, 1978; Möcks, 1986), showing that latency-jitter results in additional factors in the solution that have a certain shape. However, it is hard to distinguish whether genuine factors from additional factors due to latency-jitter in empirical applications. Therefore, an extension of ESEM that allows for variation in the factor loading patterns (see Marsh et al., 2017, for a solution in the context of confirmatory factor analysis) or even allows to attribute differences in factor loading patterns to other variables (De Roover, Timmerman, & Ceulemans, 2017) would be very useful for ERP researchers (e.g., Barry et al., 2016).

5.3 Limitations

Despite the promising results presented above, some important limitations of factor analytic approaches (i.e., EFA, ESEM, regSEM) in the context of ERP data have to be considered. The main limitation of all factor analytic approaches in the context of ERP data is that they do not take any biophysiological considerations into account but are solely data-driven analytic approaches. For this reason, some authors have argued that factor analytic methods should be generally avoided for ERP data (Groppe et al., 2008). However, whereas it is correct that the lack of a direct physiological interpretation should be kept in mind when using factor analytic methods, it does not mean that they are not useful. Beauducel et al. (2000) showed that ERPs quantified by factor analytic methods were much more reliable than ERPs quantified by the traditional measures described in section 1.2. It has also been shown that the traditional measures are more prone to variance misallocation than factor analytic quantifications because they are essentially naïve factor scoring methods (Beauducel & Debener, 2003; Donchin, 1978). Finally, some studies suggest that source localization can be improved when it is based on a

[2] Due the principles of volume conduction, the assumption of fixed time courses at all electrode sites (within each participant and condition) is physically plausible. (e.g., Nunez & Srinivasan, 2006).

factor analytic solution (Dien, Spencer, & Donchin, 2003), and that the improvement is larger when variance misallocation is minimized through the choice of an appropriate rotation criterion (Dien, 2010).

These examples clearly show that factor analytic solutions can have benefits for ERP-specific analyses such as source localization, and it seems reasonable to expect that these benefits are even more pronounced when variance misallocation is further reduced (Dien, 2010). In the present studies, such hypotheses could not be tested because all simulation studies reported here generated data from factor analytic models with pre-specified parameters rather than from source dipoles placed in head models. The reason for preferring a model-based simulation approach was that it allowed the investigation of biases in the model parameters, and unbiased estimation of model parameters is an important prerequisite for any real data application. In order to judge whether ESEM/regSEM approaches are generally useful for ERP data analyses, the simulation studies presented here must be complemented with simulation studies based on physiologically plausible head models (e.g., Dien et al., 2005), and with real data applications that demonstrate that ESEM/regSEM are able to find well-established effects.

Another limitation concerns the focus on temporal factor analytic approaches in which the sampling points are considered the observed variables to be analyzed. Hence, no conclusions can be drawn regarding extension of spatial EFA approaches in which the electrode sites are considered as observed variables. Unlike a temporal EFA, a temporal ESEM or regSEM cannot simply be translated into a spatial ESEM/regSEM by rearranging the data matrix subjected to the analysis. Rather, the structural model needs to be adapted as well. As a spatial approach does not assume fixed time courses of the factors across participants and conditions, it is not straightforward how the sampling points should be treated in such an analysis. Apart from that, spatial factors are not sparse and characterized by even greater overlap than temporal factors (Dien et al., 2007). Consequently, spatial analysis approaches have fundamentally different methodological requirements, and cannot be easily subsumed under the same research agenda as temporal analysis approaches.

Finally, some limitations that arise from the simulation design in the presented studies should be acknowledged. First, the considered factor loading patterns were very simplistic with only two factors. Typically, 8 or more factors can be expected for ERP data sets (e.g., Dien et al., 2005). This simplification was made to focus on the main principles behind variance misallocation but it may have created a too optimistic impression of the capability of factor analytic methods as very severe cases of variance misallocation have been demonstrated in the literature (e.g., Beauducel, 2018). Second, the number of factors was mostly assumed to be correct, concealing the effects of over- and underextraction of factors on the results. It is plausible to assume that especially factors with high temporal and/or spatial overlap may be challenging to detect as separate factors in common factor extraction methods such as parallel analysis (Horn, 1965). In this sense, it remains an open debate how to decide on the number of extracted factors in ERP applications (see also Dien, 2006; Dien et al., 2005; Kayser & Tenke, 2003). Third, it was assumed that the correct factor model is specified, for instance, that the error terms are uncorrelated. For time series such as ERP data, it is more common to show a certain degree of autocorrelation in the noise – leading to one or more additional factors that capture these autocorrelations (Dien, 2018). While simulation studies with real EEG-background noise did not find any hints on such distortions (Dien et al., 2005), it cannot be ruled out that noise factor have an impact on the factor solution.

5.4 Conclusion

This thesis investigated how variance misallocation can be avoided in applications of temporal EFA to ERP data. The presence of multiple sources of (co-)variance, the factor scoring step, and high temporal overlap of the factors were identified as major causes of variance misallocation in EFA for ERP data. It was shown that ESEM is capable of separating the (co-)variance sources and that it avoids biases due to factor scoring. Further, regularized estimation was shown to be a suitable alternative for factor rotation that is able to recover factor loading patterns in which only a subset of the variables follow a simple structure. Based on these results, regSEMs and ESEMs with ERP-specific rotation have been proposed as promising extensions of the EFA approach that might be less prone to variance misallocation. Future research should provide a direct comparison of regSEM and ESEM, and conduct simulation studies with more physiologically motivated data generation algorithms.

Chapter 6

References

Achim, A., & Marcantoni, W. (1997). Principal component analysis of event-related potentials: Misallocation of variance revisited. *Psychophysiology*, *34*, 597–606. doi:10.1111/j.1469-8986.1997.tb01746.x

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. doi:10.1080/10705510903008204

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. doi:10.1016/j.ijpsycho.2016.06.003

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. *Journal of Neuroscience Methods*, *295*, 20–36. doi:10.1016/j.jneumeth.2017.11.011

Beauducel, A., & Debener, S. (2003). Misallocation of variance in event-related potentials: Simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. *Journal of Neuroscience Methods*, *124*, 103–112. doi:10.1016/S0165-0270(02)00381-3

Beauducel, A., Debener, S., Brocke, B., & Kayser, J. (2000). On the reliability of augmenting/ reducing: Peak amplitudes and principal component analysis of auditory evoked potentials. *Journal of Psychophysiology*, *14*, 226–240.

Beauducel, A., & Leue, A. (2015). Controlling for experimental effects in event-related potentials by means of principal component rotation. *Journal of Neuroscience Methods*, *239*, 139–147. doi:10.1016/j.jneumeth.2014.10.008

Browne, M. W. (2001). An overview of analytic rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR360105

Bugli, C., & Lambert, P. (2007). Comparison between principal component analysis and independent component analysis in electroencephalograms modelling. *Biometrical Journal*, *49*, 312–327. doi:10.1002/bimj.200510285

De Roover, K., Timmerman, M. E., & Ceulemans, E. (2017). How to detect which variables are causing differences in component structure among different groups. *Behavior Research Methods*, *49*, 216–229. doi:10.3758/s13428-015-0687-8

De Winter, J. C. F., & Dodou, D. (2014). Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations. *Communications in Statistics - Simulation and Computation*, *0918*, 1–39. doi:10.1080/03610918.2013.862274

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research Methods*, *43*, 258–268. doi:10.3758/s13428-010-0024-1

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE*, *7*. doi:10.1371/journal.pone.0030135

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55. doi:10.1023/A:1022218503558

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clinical Neurophysiology*, *117*, 699–702. doi:10.1016/j.clinph.2005.09.029

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi:10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying Principal Components Analysis to event-related potentials: a tutorial. *Developmental Neuropsychology*, *37*, 497–517. doi:10.1080/87565641.2012.697503

Dien, J. (2016). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology*. doi:10.1016/j.ijpsycho.2016.09.006

Dien, J. (2018). ERP PCA Toolkit 2.66 Tutorial. College Park: University of Maryland. Retrieved from http://sourceforge.net/projects/erppcatoolkit

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*, 1808–1825. doi:10.1016/j.clinph.2004.11.025

Dien, J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–208). Cambridge, MA: MIT Press.

Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax vs. infomax rotations. *Human Brain Mapping*, *28*, 742–763. doi:10.1002/hbm.20304

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Brain research. Cognitive brain research*, *17*, 637–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14561451

Donchin, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering*, *5*. doi:10.3389/fneng.2012.00014

Fogarty, J. S., Barry, R. J., De Blasio, F. M., & Steiner, G. Z. (2018). ERP components and behavior in the auditory equiprobable go/no-go task: Inhibition in young adults. *Psychophysiology*, e13065. doi:10.1111/psyp.13065

Fuchs, A. H., & Milar, K. S. (2003). Psychology as a Science. In *Handbook of psychology*. doi:10.1002/0471264385.wei0101

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x. arXiv: NIHMS150003

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*, 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x. arXiv: NIHMS150003

Groppe, D. M., Makeig, S., Kutas, M., & Diego, S. (2008). Independent Component Analysis of Event-Related Potentials, 1–44.

Hirose, K., & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, *79*, 120–132. doi:10.1016/j.csda.2014.05.011. arXiv: 1302.5475

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. doi:10.1007/BF02289447

Huang, P.-H., Chen, H., & Weng, L.-J. (2017). A Penalized Likelihood Method for Structural Equation Modeling. *Psychometrika*, *82*, 329–354. doi:10.1007/s11336-017-9566-9

Jackson, A. F., & Bolger, D. J. (2014). The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology*, *51*, n/a–n/a. doi:10.1111/psyp.12283

Jacobucci, R. (2017). regsem: Regularized Structural Equation Modeling, 1–13. arXiv: 1703.08489. Retrieved from http://arxiv.org/abs/1703.08489

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793. arXiv: 15334406

Kappenman, E. S., & Luck, S. J. (2012). ERP Components: The Ups and Downs of Brainwave Recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The oxford handbook of event-related potential components* (Chap. 1). New York, NY: Oxford University Press.

Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*, 2307–2325. doi:10.1016/S1388-2457(03)00241-4

Kayser, J., & Tenke, C. E. (2006). Consensus on PCA for ERP data, and sensibility of unrestricted solutions. *Clinical Neurophysiology*, *117*, 703–707. doi:10.1016/j.clinph.2005.11.015

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, Mass.: MIT Press.

Luck, S. J., & Gaspelin, N. (2016). How to Get Statistically Significant Effects in Any ERP Experiment (and Why You Shouldn't). *In Press, Psychophysiology 1*, *4424*, 1–44. doi:10.1111/psyp.12639

Maganti, R. K., & Rutecki, P. (2013). EEG and Epilepsy Monitoring. *CONTINUUM: Lifelong Learning in Neurology*, *19*, 598–622. doi:10.1212/01.CON.0000431378.51935.d8

Makeig, S., Westerfield, M., Jung, T. P., Covington, J., Townsend, J., Sejnowski, T. J., & Courchesne, E. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *19*, 2665–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10087080

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2017). What to do When Scalar Invariance Fails: The Extended Alignment Method for Multi-Group Factor Analysis Comparison of Latent Means Across Many Groups. *Psychological Methods*, *23*, 524–545. doi:10.1037/met0000113

McDonald, R. (1996). Path Analysis with Composite Variables. *Multivariate Behavioral Research*, *31*, 239–270. doi:10.1207/s15327906mbr3102

Möcks, J. (1986). The Influence of Latency Jitter in Principal Component Analysis of Event-Related Potentials. doi:10.1111/j.1469-8986.1986.tb00659.x

Möcks, J. (1988). Topographic components model for event-related potentials and some biophysical considerations. *IEEE Transactions on Biomedical Engineering*, *35*, 482–484. doi:10.1109/10.2119

Möcks, J., & Verleger, R. (1986). Principal component analysis of event-related potentials: A note on misallocation of variance. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *65*, 393–398. doi:10.1016/0168-5597(86)90018-3

Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., & Arnfred, S. M. (2006). Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, *29*, 938–947. doi:10.1016/j.neuroimage.2005.08.005

Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, *26*, 1041–1054. doi:10.1016/j.mri.2008.01.011

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain : the neurophysics of EEG* (2. ed.). Oxford u.a.: Oxford Univ. Press. Retrieved from https://katalog.ub.uni-leipzig.de/Record/0000771752

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within Exploratory Factor Analysis. *Multivariate Behavioral Research*, *45*, 73–103. doi:10.1080/00273170903504810

Scharf, F., & Nestler, S. (2018a). Exploratory structural equation modeling for event-related potential data—An all-in-one approach? *Psychophysiology*, e13303. doi:10.1111/psyp.13303

Scharf, F., & Nestler, S. (2018b). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119–136. doi:10.1016/j.ijpsycho.2018.03.019

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in Exploratory Factor Analysis? *Structural Equation Modeling: A Multidisciplinary Journal*. doi:10.1080/10705511.2018.1558060

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for Exploratory Factor Analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*, 95–113. doi:10.1177/0013164410387348

Slotnick, S. D. (2005). Source localization. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 149–166). Cambridge, MA: MIT Press.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–358. doi:10.1017/S0048577201000324

Sur, S., & Sinha, V. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*, 70. doi:10.4103/0972-6748.57865

Thompson, R. F., & Zola, S. M. (2003). Biological Psychology. In *Handbook of psychology*. doi:10.1002/0471264385.wei0103

Thomson, G. H. (1935). The definition and measurement of "g" (general intelligence). *Journal of Educational Psychology*, *26*, 241–262. doi:10.1037/h0059873

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* doi:10.1037/10018-000

Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, *29*, 431–454. doi:10.1007/s00180-013-0434-5

Trendafilov, N. T., & Adachi, K. (2015). Sparse Versus Simple Structure Loadings. *Psychometrika*, *80*, 776–790. doi:10.1007/s11336-014-9416-y

Trendafilov, N. T., Fontanella, S., & Adachi, K. (2017). Sparse Exploratory Factor Analysis. *Psychometrika*. doi:10.1007/s11336-017-9575-8

van der Linden, W. J. (2009). Conceptual Issues in Response-Time Modeling. *Journal of Educational Measurement*, *46*, 247–272. doi:10.1111/j.1745-3984.2009.00080.x

Verleger, R., & Möcks, J. (1987). Varimax may produce slow-wave-like shapes by merging monotonic trends with other components. *Journal of Psychophysiology*, *1*, 265–270.

Verleger, R., Paulick, C., Möcks, J., Smith, J. L., & Keller, K. (2013). Parafac and go/no-go: Disentangling CNV return from the P3 complex by trilinear component analysis. *International Journal of Psychophysiology*, *87*, 289–300. doi:10.1016/j.ijpsycho.2012.08.003

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Lawrence Erlbaum Associates.

Widaman, K. F. (2018). On Common Factor and Principal Component Representations of Data: Implications for Theory and for Confirmatory Replications. *Structural Equation Modeling*, *00*, 1–19. doi:10.1080/10705511.2018.1478730

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, *113*, 767–791. doi:10.1016/S1388-2457(02)00057-3

Wood, C. C., & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical neurophysiology*, *59*, 249–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6203715

Florian Scharf

Advances in the analysis of event-related potential data with factor analytic methods

Dissertation
submitted to the Faculty of Life Sciences of the University of Leipzig

Leipzig, 23.01.2019

171 pages, 72 references, 10 figures, 14 tables

Researchers are often interested in comparing brain activity between experimental contexts. Event-related potentials (ERPs) are a common electrophysiological measure of brain activity that is time-locked to an event (e.g., a stimulus presented to the participant). A variety of decomposition methods has been used for ERP data among them temporal exploratory factor analysis (EFA). Essentially, temporal EFA decomposes the ERP waveform into a set of latent factors where the factor loadings reflect the time courses of the latent factors, and the amplitudes are represented by the factor scores.

An important methodological concern is to ensure the estimates of the condition effects are unbiased and the term *variance misallocation* has been introduced in reference to the case of biased estimates. The aim of the present thesis was to explore how exploratory factor analytic methods can be made less prone to variance misallocation. These efforts resulted in a series of three publications in which variance misallocation in EFA was described as a consequence of the properties of ERP data, ESEM was proposed as an extension of EFA that acknowledges the structure of ERP data sets, and regularized estimation was suggested as an alternative to simple structure rotation with desirable properties.

The presence of multiple sources of (co-)variance, the factor scoring step, and high temporal overlap of the factors were identified as major causes of variance misallocation in EFA for ERP data. It was shown that ESEM is capable of separating the (co-)variance sources and that it avoids biases due to factor scoring. Further, regularized estimation was shown to be a suitable alternative for factor rotation that is able to recover factor loading patterns in which only a subset of the variables follow a simple structure. Based on these results, regSEMs and ESEMs with ERP-specific rotation have been proposed as promising extensions of the EFA approach that might be less prone to variance misallocation. Future research should provide a direct comparison of regSEM and ESEM, and conduct simulation studies with more physiologically motivated data generation algorithms.

Florian Scharf

Advances in the analysis of event-related potential data with factor analytic methods

Dissertation
submitted to the Faculty of Life Sciences of the University of Leipzig

Leipzig, 23.01.2019

Introduction

Researchers are often interested in comparing brain activity between experimental contexts. Event-related potentials (ERPs) are a common electrophysiological measure of brain activity that is time-locked to an event (e.g., a stimulus presented to the participant; Luck, 2014). ERPs are computed from a continuous electroencephalogram (EEG) that is recorded at multiple electrode sites (e.g., 64 or 128) from the participant's scalp with a high sampling rate (e.g., 500 Hz). For each participant, electrode site and event type, the continuous EEG signal is cut into epochs around the events of interest (e.g., stimuli in several experimental conditions) and averaged across repetitions of the same event to improve the signal-to-noise ratio. The resulting ERP data set, consisting of the averaged waveforms, can be arranged in a 4-dimensional hypermatrix with the dimensions sampling points $\times$ participants $\times$ electrodes $\times$ event type.

The ERP is typically described in terms of amplitude, polarity, latency, and topography (i.e., the distribution of the voltages across all electrode sites) around the time points of minimal and maximal voltage deflection. ERP researchers may be interested in differences between different events in any (or all) of these four features with the ultimate goal to attribute differential brain activity to differences in psychological processes (e.g., differential states of attention). Here, only the case where researchers are interested in amplitude differences is considered. Traditionally, the amplitudes of the ERP peaks for each participant, condition, and electrode site are quantified for subsequent statistical analyses using simple peak amplitudes (i.e., the local minima and maxima of the voltage), mean voltages in time windows around the grand average peaks, or using a variety of area under the curve measures (Luck, 2014, chapter 9). Alternatively, statistical tests for condition effects may be done separately for each sampling point and electrode site (Groppe, Urbach, & Kutas, 2011a, 2011b). Functional interpretations of the ERPs are then made based on the results of the statistical analyses, typically an analysis of variance (ANOVA).

The statistical analysis of ERP data sets with this traditional approach is problematic for at least two reasons. First, the electric potential recorded from the scalp is a 2D projection of the 3D source activity in the brain. This makes it hard to attribute differences in the ERP waveforms to the underlying source signals. Second, due to their high dimensionality, the statistical analysis of ERP data suffers from a massive multivariate comparison problem that can only be solved at the cost of a considerable reduction in statistical power. Taken together, these problems result in an interpretation problem. For instance, when a significant condition effect (in a certain time range) is observed at two different electrode sites, it is not clear if this result reflects a condition effect on a single source signal that projects to both electrode sites or a condition effect on two separate sources.

A variety of decomposition methods has been used for ERP data among them temporal exploratory factor analysis (EFA; Dien, 2012; Donchin, 1978). Essentially, temporal EFA decomposes the ERP waveform into a set of latent factors where the factor loadings reflect the time courses of the latent factors, and the amplitudes are represented by the factor scores. EFA addresses the first challenge because the decomposition into factors does not rely on visible peaks in the waveform but on statistical properties of the data. EFA also addresses the second challenge because statistical analyses of condition effects can be conducted on the factor scores, condensing the information from all sampling points into a single score for each factor, participant, electrode, and condition. In this context, it is crucial that the estimation of experimental effects is not biased by the preceding EFA because otherwise functional interpretations of the factors might be misguided. The term *variance misallocation* has been introduced in reference to the case where variance is incorrectly attributed to factors, resulting in biased condition effect estimates (Wood & McCarthy, 1984).

The goal of the present dissertation was to investigate how the risk of variance misallocation can be minimized in applications of factor analytic methods to ERP data. In a series of three publications, the determinants of the occurrence of variance misallocation are identified (Scharf & Nestler, 2018b), and recently proposed improvements to EFA approaches are investigated that can considerably reduce the risk of variance misallocation (Scharf & Nestler, 2018a, 2019).

Study 1

In Scharf and Nestler (2018b), the principles behind variance misallocation were investigated by means of an analytic decomposition of the factor (co-)variance matrix and a Monte Carlo simulation. The study set out from the fact that ERP data sets differ from psychometric data sets, for which EFA was originally intended, in at least two ways: First, the observations in the rows of an ERP data matrix are not independent and exchangeable. Rather, they are well structured and some observations are more strongly correlated with each other than others because they stem from the same electrode site, the same participant, and/or the same event type, and this fact cannot be acknowledged in the EFA model. Second, latent factors extracted from ERP data are likely to have a considerable temporal overlap (i.e., a considerable amount of cross-loadings), and this can hardly be influenced by researchers themselves. The study was concerned with the consequences of these two properties for the estimation performance and interpretability of the EFA parameters.

It was shown that variance misallocation can be the result of inappropriately applying orthogonal factor rotation (orthogonality bias) and/or the result of simple structure rotation which biases the estimated factor loadings towards the optimum of its mathematical simplicity criterion (rotation bias). Most importantly, the presented considerations provided formal support for previous recommendations of oblique over orthogonal rotations in temporal EFA for ERP data (Dien, 1998; Dien, Beal, & Berg, 2005). An analytical decomposition of the variance-covariance matrix of the latent factor revealed that it can be written as a sum of contributions due to electrodes, participants, and condition effects. Given that no more than two orthogonal factors are physiologically possible (Dien, 2010), and that the number of factors is typically 8 and more for ERP data sets (e.g., Barry, De Blasio, Fogarty, & Karamacoska, 2016), it is highly unlikely for the (co-)variance contributions to sum up to zero and orthogonal rotation should be generally avoided. These results cleared remaining doubts about the general appropriateness of oblique rotation in temporal EFA for ERP data (Dien, 2006; Kayser & Tenke, 2003, 2006).

Study 2

Addressing the consequences of the neglected structure of ERP data sets, in Scharf and Nestler (2018a), exploratory structural equation modeling (ESEM) is proposed as an alternative to EFA that can properly acknowledge the structure of ERP data sets, for instance, providing substantively interpretable factor correlation estimates. ESEM expands EFA by a structural model in which predictors of the latent variables can be specified (Asparouhov & Muthén, 2009). Specifically, it was proposed to replace EFA with an ESEM in which electrode site, condition, and their interactions are specified as predictors in the structural model. A simulation study confirmed that ESEM is capable of separating the (co-)variance contributions. In addition, it was shown that variance misallocation can also occur as a consequence of the factor scoring procedure that is necessary as an intermediate step in the EFA but not in the ESEM approach.

Study 3

The results from Study 1 emphasized the importance of the rotation step in EFA for the occurrence of variance misallocation. As a rotation step is also an essential part of ESEM, ESEM suffers from biases due to factor rotation as well. Recently, regularized (or sparse) estimation of factor models has been proposed as a substitute for factor rotation that is able to provide good factor solutions even for some conditions under which rotated EFA does not (see Trendafilov, 2014, for a review). Whereas several different regularized factor analysis methods were proposed, no extensive comparison of the performance of regularization and simple structure rotation has been available. The simulation study by Scharf and Nestler (2019) closed this gap, comparing the performance of simple structure rotation and regularization for a wide range of factor loading patterns.

The results showed that elastic net regularized estimation is a suitable alternative to factor rotation. It yielded reasonable results irrespective whether the population factor loading pattern was simple or complex. Although the presented simulations focused on typical psychometric applications, the results are relevant for the analysis of ERP data as well. Especially the results from a set of conditions with partial simple structure are of interest for ERP data applications. In these conditions, half of the observed variables followed a simple structure whereas the other half did not. Such conditions with partial temporal overlap can be expected to occur regularly with ERP data sets. The promising performance of regularized estimation under such conditions makes it a suitable candidate for ERP applications as well.

Implications and conclusion

The presented results imply that a combination of the ESEM approach and regularized estimation may provide a data analytic framework for ERP data that is less prone to all three causes of variance misallocation. Recently, regularized structural equation modeling (regSEM) has been proposed offering such a combination (Jacobucci, Grimm, & McArdle, 2016). More specifically, an ESEM can be conceptualized as a regSEM by using the same structural model as in the ESEM and specifying a measurement model in which all factors are allowed to load on all observed variables, and in which all factor loadings are regularized during the estimation. Alternatively, it was suggested to develop ERP-specific rotation criteria that include assumptions regarding plausible time courses of the factors that are motivated directly from ERP research (Beauducel, 2018). ERP-specific rotation criteria can easily be combined with the ESEM approach, offering an alternative approach to reduce the rotation bias within the ESEM framework. A direct comparison of ESEM with simple structure rotation, ESEM with ERP-specific rotation, and regSEM is required to explore which of the methods performs best under a variety of conditions.

To conclude, the present dissertation investigated how variance misallocation can be avoided in applications of temporal EFA to ERP data. The presence of multiple sources of (co-)variance, the factor scoring step, and high temporal overlap of the factors were identified as major causes of variance misallocation in EFA for ERP data. It was shown that ESEM is capable of separating the (co-)variance sources and that it avoids biases due to factor scoring. Further, regularized estimation was shown to be a suitable alternative for factor rotation that is able to recover factor loading patterns in which only a subset of the variables follow a simple structure. Based on these results, regSEMs and ESEMs with ERP-specific rotation have been proposed as promising extensions of the EFA approach that might be less prone to variance misallocation. Future research should provide a direct comparison of regSEM and ESEM, and conduct simulation studies with physiologically motivated data generation algorithms.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. doi:10.1080/10705510903008204

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. doi:10.1016/j.ijpsycho.2016.06.003

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. *Journal of Neuroscience Methods*, *295*, 20–36. doi:10.1016/j.jneumeth.2017.11.011

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55. doi:10.1023/A:1022218503558

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clinical Neurophysiology*, *117*, 699–702. doi:10.1016/j.clinph.2005.09.029

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi:10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying Principal Components Analysis to event-related potentials: a tutorial. *Developmental Neuropsychology*, *37*, 497–517. doi:10.1080/87565641.2012.697503

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*, 1808–1825. doi:10.1016/j.clinph.2004.11.025

Donchin, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*, 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793

Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*, 2307–2325. doi:10.1016/S1388-2457(03)00241-4

Kayser, J., & Tenke, C. E. (2006). Consensus on PCA for ERP data, and sensibility of unrestricted solutions. *Clinical Neurophysiology*, *117*, 703–707. doi:10.1016/j.clinph.2005.11.015

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, Mass.: MIT Press.

Scharf, F., & Nestler, S. (2018a). Exploratory structural equation modeling for event-related potential data—An all-in-one approach? *Psychophysiology*, e13303. doi:10.1111/psyp.13303

Scharf, F., & Nestler, S. (2018b). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119–136. doi:10.1016/j.ijpsycho.2018.03.019

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in Exploratory Factor Analysis? *Structural Equation Modeling: A Multidisciplinary Journal*. doi:10.1080/10705511.2018.1558060

Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, *29*, 431–454. doi:10.1007/s00180-013-0434-5

Wood, C. C., & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical neurophysiology*, *59*, 249–60.

Florian Scharf

Advances in the analysis of event-related potential data with factor analytic methods

Dissertation
submitted to the Faculty of Life Sciences of the University of Leipzig

Leipzig, 23.01.2019

Einleitung

Forscher sind häufig daran interessiert die Hirnaktivität zwischen verschiedenen experimentellen Kontexten zu vergleichen. Ereignis-korrelierte Potentiale (EKPs) sind ein oft verwendetes elektrophysiologisches Maß der Hirnaktivität, die zeitlich gekoppelt an ein Ereignis auftritt (z.B. an einen Stimulus, der dem Probanden präsentiert wird; Luck, 2014). EKPs werden aus einem Elektroenzephalogramm (EEG) gewonnen, welches kontinuierlich von vielen auf dem Kopf des Probanden platzierten Elektroden (z.B. 64 oder 128) mit hoher Sampling-Rate (z.B. 500 Hz) aufgezeichnet wurde. Das kontinuierliche EEG-Signal wird in Epochen um die interessierenden Ereignisse (z.B. Stimuli in verschiedenen experimentellen Bedingungen) herum ausgeschnitten und separat für alle Probanden, Elektroden und Ereignistypen über alle Wiederholungen des gleichen Ereignisses hinweg gemittelt, um das Signal-Rausch-Verhältnis zu verbessern. Der daraus resultierende EKP-Datensatz, welcher die mittleren Spannungsverläufe enthält, kann in einer 4-dimensionalen Hypermatrix mit den Dimensionen Messzeitpunkte $\times$ Probanden $\times$ Elektroden $\times$ Ereignistypen arrangiert werden.

EKPs werden typischerweise anhand der Amplitude, Polarität, Latenz und Topographie (d.h., der Spannungsverteilung über die Elektroden hinweg) der Messzeitpunkte nahe der minimalen und maximalen Spannungswerte beschrieben. Klassischerweise interessieren sich EKP-Forscher für Unterschiede zwischen Ereignistypen in einer dieser Eigenschaften mit dem Ziel Unterschiede in der Hirnaktivität auf Unterschiede in psychologischen Prozessen (z.B. unterschiedliche Aufmerksamkeitszustände) zurückzuführen. Hier soll ausschließlich der Fall betrachtet werden, bei dem Wissenschaftler an Amplitudenunterschieden interessiert sind.

Typischerweise werden die Amplituden der EKP-Gipfel pro Proband, Bedingung und Elektrode für die anschließende statistische Analyse quantifiziert, indem man die Gipfelamplituden (d.h., die lokalen Minima und Maxima des Spannungszeitverlaufs), die mittlere Spannung in einem Zeitfenster um die Gipfel des *Grand Averages* oder ein Maß für die Fläche unter der Kurve verwendet (Luck, 2014, Kapitel 9). Alternativ können statistische Tests einzeln für jeden Messzeitpunkt und jede Elektrode durchgeführt werden (Groppe, Urbach & Kutas, 2011a, 2011b). Eine funktionelle Interpretation des EKPs erfolgt dann auf Basis der Ergebnisse der statistischen Analysen, zumeist Varianzanalysen.

Die statistische Analyse von EKP-Datensätzen mit diesem klassischen Ansatz ist unter mindestens zwei Gesichtspunkten problematisch. Erstens ist das elektrische Potenzial auf der Schädeloberfläche eine 2-dimensionale Projektion der Quellaktivität im 3-dimensionalen Raum. Dadurch ist es schwierig, die EKP-Verläufe auf die zugrundeliegenden Quellsignale zurückzuführen. Zweitens führt die Hochdimensionalität der EKP-Daten zu einem massiven multiplen Testproblem für die statistischen Analysen, dem nur auf Kosten der Teststärke begegnet werden kann. Zusammen genommen resultiert daraus auch ein Interpretationsproblem. Beispielsweise bleibt es unklar, ob ein statistisch signifikanter Bedingungseffekt an zwei Elektroden im selben Zeitraum das Ergebnis eines Bedingungsunterschiedes in einer einzigen zugrunde liegenden Quelle oder eines Bedingungsunterschiedes verschiedener Quellen ist.

Eine Reihe von Dekompositionsmethoden wurde bereits auf EKP-Daten angewendet, darunter temporale explorative Faktoranalysen (EFA; Dien, 2012; Donchin, 1978). Im Wesentlichen zerlegt eine temporale EFA die EKP-Zeitverläufe in eine Anzahl an latenten Faktoren, wobei die Faktorladungen die Zeitverläufe der latenten Faktoren widerspiegeln und die Faktorwerte die abbilden. Die EFA adressiert das Überlappungsproblem, weil die Dekomposition nicht auf den sichtbaren Gipfeln des EKPs basiert, sondern auf statistischen Eigenschaften der Daten. Auch das multiple Testproblem wird in der EFA reduziert, weil Inferenzstatistik für die Bedingungseffekte anhand der Faktorwerte betrieben werden kann, wobei die Informationen von allen Messzeitpunkten in je einem einzelnen Wert pro Faktor, Proband, Elektrode und Bedingung zusammengefasst werden. In diesem Zusammenhang ist es wichtig, dass die geschätzten Bedingungseffekte erwartungstreu sind und nicht durch die vorangehende EFA verzerrt werden, weil andernfalls auch die funktionelle Interpretation der Faktoren beeinträchtigt wäre. Der Begriff *Varianzmissallokation* wurde in Bezug auf den Fall eingeführt, in dem Varianz falsch auf die Faktoren verteilt wird, sodass auch die Bedingungseffekte nicht erwartungstreu sind (Wood & McCarthy, 1984).

Das Ziel der vorliegenden Dissertation war es zu untersuchen, wie das Risiko von Varianzmissallokation in der Anwendung von faktoranalytischen Methoden auf EKP-Daten minimiert werden kann. In einer Reihe von drei Publikationen wurden die Determinanten des Auftretens von Varianzmissallokation herausgearbeitet (Scharf & Nestler, 2018b) und kürzlich vorgeschlagene Weiterentwicklungen von EFA-Ansätzen im Hinblick darauf untersucht, ob sie das Risiko von Varianzmissallokation verringern (Scharf & Nestler, 2018a, 2019).

Studie 1

In Scharf und Nestler (2018b) wurden die Prinzipien hinter Varianzmissallokation mittels einer analytischen Zerlegung der Varianz-Kovarianz-Matrix der latenten Faktoren und einer Monte Carlo Simulation untersucht. Ausgangspunkt war die Tatsache, dass EKP-Datensätze sich von psychometrischen Datensätzen, für welche EFA ursprünglich entwickelt wurde, auf mindestens zwei Arten unterscheiden: Zum einen sind die Beobachtungen in den Zeilen einer EKP-Datenmatrix nicht unabhängig und austauschbar, sondern stark strukturiert. Einige Beobachtungen korrelieren stärker miteinander als andere, weil sie von derselben Elektrodenposition, vom selben Probanden und/oder vom selben Ereignistyp stammen. Dies kann innerhalb des EFA-Modells nicht berücksichtigt werden. Zum anderen zeichnen sich die Faktoren, die bei EKP-Daten extrahiert werden durch substantielle zeitliche Überlappung aus (d.h., durch substantielle Kreuzladungen), was kaum beeinflusst werden kann. Die Studie beschäftigte sich mit den Konsequenzen beider Eigenschaften für die Performanz der Schätzungen sowie die Interpretierbarkeit der EFA-Parameter.

Es zeigte sich, dass Varianzmissallokation infolge der nicht gerechtfertigten Verwendung orthogonaler Rotationen (Orthogonalitätsbias) und/oder infolge der Faktorrotation auftreten kann, welche die geschätzten Faktorladungen hin zum Optimum ihres Einfachstrukturkriteriums verzerrt (Rotationsbias). Vor allem lieferten die dargestellten Überlegungen eine formale Begründung für die in einigen Studien gefundene Überlegenheit obliquer Rotationsmethoden bei temporaler EFA für EKP-Daten (Dien, 1998; Dien, Beal & Berg, 2005). Die analytische Zerlegung der Varianz-Kovarianz-Matrix der latenten Faktoren zeigte, dass diese als Summe von Beiträgen der Elektroden, Probanden und der Bedingungsunterschiede dargestellt werden kann. Berücksichtigt man die Tatsachen, dass physiologisch maximal zwei orthogonale Faktoren möglich sind (Dien, 2010) und dass die Zahl der Faktoren bei typischen EKP-Datensätzen eher acht oder mehr beträgt (z.B. Barry, De Blasio, Fogarty & Karamacoska, 2016), ist es äußerst unwahrscheinlich, dass sich die Beiträge zu null aufsummieren und orthogonale Rotation sollte generell vermieden werden. Diese Befunde räumten letzte Zweifel an der generellen Angemessenheit obliquer Rotation bei Anwendung temporaler EFA auf EKP-Daten aus (Dien, 2006; Kayser & Tenke, 2003, 2006).

Studie 2

In Scharf und Nestler (2018a) wurden die Konsequenzen der vernachlässigten Struktur von EKP-Datensätzen adressiert. Das explorative Strukturgleichungsmodell (ESEM) wurde als Alternative zur EFA vorgeschlagen, bei dem die Struktur von EKP-Daten angemessen berücksichtigt werden kann, wodurch es zum Beispiel inhaltlich interpretierbare Faktorkorrelationsparameter liefern kann. ESEM erweitert EFA um ein Strukturmodell, in dem Prädiktoren der Faktoren spezifiziert werden können (Asparouhov & Muthén, 2009). Konkret wurde vorgeschlagen, EFA durch ein ESEM zu ersetzen, in dem die Elektrodenposition, die Bedingungseffekte sowie deren Interaktionen als Prädiktoren ins Strukturmodell aufgenommen werden. Eine Simulationsstudie bestätigte, dass ESEM in der Lage ist, die einzelnen Beiträge zur Faktor(ko)varianz zu trennen. Darüberhinaus wurde gezeigt, dass Varianzmissallokation auch als Folge der Faktorwertbestimmung auftreten kann, welche ein Zwischenschritt der EFA-Methode, aber nicht des ESEMs ist.

Studie 3

Die Ergebnisse aus Studie 1 unterstrichen die Bedeutung des Rotationsschrittes in der EFA für das Auftreten von Varianzmissallokation. Da Faktorrotation auch essentieller Bestandteil des ESEMs ist, ist dieses ebenfalls anfällig für den Rotationsbias. Kürzlich wurde regularisierte Schätzung des Faktormodells als Alternative zur Rotation vorgeschlagen, welche in der Lage ist, unter allgemeineren Bedingungen gute Ergebnisse zu erzielen als Faktorrotation (Trendafilov, 2014). Während zahlreiche Varianten regularisierter Faktoranalysen vorgeschlagen wurden, fehlte es an ausführlichen Vergleichen zwischen regularisierten und rotierten Faktoranalysen. In Scharf und Nestler (2019) wurde diese Lücke geschlossen, indem die Performanz von Rotation und Regularisierung für eine breite Auswahl an Populationsfaktorladungsmustern verglichen wurde.

Die Ergebnisse zeigten, dass eine *elastic net* regularisierte Schätzung eine geeignete Alternative zur Faktorrotation darstellt, weil sie brauchbare Ergebnisse lieferte, unabhängig davon, ob das Populationsladungsmuster eher einfach oder komplex war. Obwohl die präsentierte Simulationsstudie auf typische psychometrische Anwendungen fokussierte, sind die Ergebnisse auch im Kontext der Analyse von EKP-Daten relevant. Insbesondere zeigte sich, dass Regularisierung Vorteile bietet, wenn nur ein Teil der Variablen einer Einfachstruktur folgt. Derartige Bedingungen können bei EKP-Daten regelmäßig auftreten, sodass Regularisierung auch in diesem Kontext in Betracht gezogen werden sollte.

Implikationen

Die vorgestellten Ergebnisse implizieren, dass eine Kombination aus dem ESEM-Ansatz und regularisierter Schätzung eine Analysemethode bereitstellen könnte, die weniger anfällig für alle drei Ursachen von Varianzmissallokation ist. Kürzlich wurden regularisierte Strukturgleichungsmodelle (regSEM) vorgestellt, welche so eine Kombination verfügbar machen (Jacobucci, Grimm & McArdle, 2016). Konkret kann ein ESEM als ein regSEM aufgefasst werden, bei dem dasselbe Strukturmodell spezifiziert wurde wie im ESEM, aber ein Messmodell angenommen wird, in dem alle Faktoren auf alle Variablen laden, und alle Faktorladungen in den Strafterm der regularisierten Schätzung eingehen. Alternativ wurde vorgeschlagen, EKP-spezifische Rotationskriterien zu entwickeln, welche Annahmen über plausible Zeitverläufe der Faktoren machen, welche direkt aus der EKP-Forschung abgeleitet sind (Beauducel, 2018). EKP-spezifische Rotation kann ohne Weiteres mit dem ESEM-Ansatz verwendet werden, sodass der Rotationsbias auch im Rahmen des ESEMs reduziert werden könnte. Ein direkter Vergleich zwischen ESEM mit Rotation zur Einfachstruktur, ESEM mit EKP-spezifischer Rotation und regSEM ist notwendig, um herauszufinden, welche Methode über eine breite Spanne von Bedingungen am besten funktioniert.

Conclusio

Die vorliegende Dissertation untersuchte, wie Varianzmissallokation vermieden werden kann, wenn temporale EFA auf EKP-Daten angewendet wird. Die Existenz mehrerer Varianzquellen, der Schritt der Faktorwertbestimmung und die starke zeitliche Überlappung der Faktoren wurden als Hauptgründe für Varianzmissallokation identifiziert. Es wurde gezeigt, dass ein ESEM-Ansatz die Varianzquellen trennen kann und dass er Verzerrungen durch Faktorwertbestimmung vermeidet. Weiterhin wurde gezeigt, dass regularisierte Schätzung eine geeignete Alternative zur Faktor-rotation ist, die auch dann noch brauchbare Ergebnisse liefert, wenn nur ein Teil der Variablen einer Einfachstruktur folgt. Basierend auf diesen Ergebnissen wurden regSEMs und ESEMs mit EKP-spezifischer Rotation als vielversprechende Erweiterungen des EFA-Ansatzes vorgestellt. Zukünftige Forschung sollte einen direkten Vergleich von regSEM und ESEM sowie weitere Simulationsstudien mit physiologisch motivierter Datenerzeugung durchführen.

Referenzen

Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. doi:10.1080/10705510903008204

Barry, R. J., De Blasio, F. M., Fogarty, J. S. & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. doi:10.1016/j.ijpsycho.2016.06.003

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. *Journal of Neuroscience Methods*, *295*, 20–36. doi:10.1016/j.jneumeth.2017.11.011

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, *11*, 43–55. doi:10.1023/A:1022218503558

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clinical Neurophysiology*, *117*, 699–702. doi:10.1016/j.clinph.2005.09.029

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi:10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying Principal Components Analysis to event-related potentials: a tutorial. *Developmental Neuropsychology*, *37*, 497–517. doi:10.1080/87565641.2012.697503

Dien, J., Beal, D. J. & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*, 1808–1825. doi:10.1016/j.clinph.2004.11.025

Donchin, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Groppe, D. M., Urbach, T. P. & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P. & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*, 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x

Jacobucci, R., Grimm, K. J. & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793

Kayser, J. & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*, 2307–2325. doi:10.1016/S1388-2457(03)00241-4

Kayser, J. & Tenke, C. E. (2006). Consensus on PCA for ERP data, and sensibility of unrestricted solutions. *Clinical Neurophysiology*, *117*, 703–707. doi:10.1016/j.clinph.2005.11.015

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2. Aufl.). Cambridge, Mass.: MIT Press.

Scharf, F. & Nestler, S. (2018a). Exploratory structural equation modeling for event-related potential data—An all-in-one approach? *Psychophysiology*, e13303. doi:10.1111/psyp.13303

Scharf, F. & Nestler, S. (2018b). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119–136. doi:10.1016/j.ijpsycho.2018.03.019

Scharf, F. & Nestler, S. (2019). Should regularization replace simple structure rotation in Exploratory Factor Analysis? *Structural Equation Modeling: A Multidisciplinary Journal*. doi:10.1080/10705511.2018.1558060

Trendafilov, N. T. (2014). From simple structure to sparse components: A review. *Computational Statistics*, *29*, 431–454. doi:10.1007/s00180-013-0434-5

Wood, C. C. & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical neurophysiology*, *59*, 249–60.

Florian Scharf

Curriculum Vitae

January 2019

Windmühlenstraße 18
04107 Leipzig, Germany
0160 9917 9426
florian@scharf-net.de

Education

2014 – 2016	M.Sc. Psychology	University of Leipzig
2011 – 2014	B.Sc. Psychology	University of Leipzig
2009 – 2010	B.Sc. Chemistry (studies not completed)	University of Leipzig
2009	Abitur [higher education entrance qualification]	Gymnasium Brandis

Academic and Work Experience

10/2016 – present	**Research and teaching fellow**, **Doctoral Student**, University of Leipzig, Chair for Psychological Methods
01/2017 – present	**Guest researcher**, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig
01/2018 – 02/2018	**DAAD research stay** hosted by Dr. Urte Roeber and Prof. Robert O'Shea as part of a DAAD joint research cooperation granted to Prof. Erich Schröger and Prof. Robert O'Shea, Murdoch University
03/2017 – 12/2017	**Research fellow**, University of Leipzig, Chair for Cognitive & Biological Psychology, in partial replacement of Assistant Professor Dr. Sabine Grimm
09/2012 – 09/2016	**Student teaching assistant**, University of Leipzig, Chair for Experimental Psychology & Methods and Chair for Psychological Methods (from 04/2016)
03/2015 – 07/2016	**Student research & teaching assistant**, University of Leipzig, Chair for Cognitive & Biological Psychology
09/2014	**Research stay**, Hungarian Academy of Sciences, Research Centre for Natural Sciences, Institute of Cognitive Neuroscience and Psychology
01/2014 – 03/2015	**Student research assistant**, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Otto-Hahn-Group "Neural Bases of Intonation in Speech and Music"
08/2013 – 10/2013	**Internship**, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Otto-Hahn-Group "Neural Bases of Intonation in Speech and Music"
07/2012 – 12/2013	**Student research assistant**, Leipzig University of Applied Sciences, Project TOPINUS

Awards and Funding

2018	**Presentation award** for best junior scientist presentation for the talk "Orthogonal versus Oblique Rotation in Temporal EFA for Event-related Potentials" presented at the 8th European Congress of Methodology in Jena
2016	**Poster award** for the best poster prepared in the context of the "Projektmodul" [Student research project]
04/2015 – 09/2016	**Deutschlandstipendium**, German national scholarship for high-achieving and committed students
2015	**Wilhelm-Wundt Teaching Award** granted by the Fachschaftsrat Psychologie [student representatives] of the University of Leipzig

Professional Memberships

German Society for Psychology (DGPs): Section Quantitative Methods and Evaluation

Journal Reviewer (Ad-hoc)

Psychophysiology, International Journal of Statistics and Probability

Publications

Peer-reviewed articles

Scharf, F., & Nestler, S. (in press). Should regularization replace simple structure rotation in Exploratory Factor Analysis? *Structural Equation Modeling: A Multidisciplinary Journal*. DOI: 10.1080/10705511.2018.1558060

Scharf, F., & Nestler, S. (2018). Exploratory structural equation modeling for event-related potential data – an all-in-one approach? *Psychophysiology*. DOI: 10.1111/psyp.13303.

Wetzel, N., Scharf, F. & Widmann, A. (2018). Can't ignore – distraction by task-irrelevant sounds in early and middle childhood. *Child Development*. DOI: 10.1111/cdev.13109

Scharf, F., & Nestler, S. (2018). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology*, *128*, 119-136. DOI: 10.1016/j.ijpsycho.2018.03.019.

Bianco, R., Novembre, G., Keller, P. E., Kim, S.-G., Scharf, F., Friederici, A. D., et al. (2016). Neural networks for harmonic structure in music perception and action. *NeuroImage*, *142*, 454-464. DOI: 10.1016/j.neuroimage.2016.08.025.

Bianco, R., Novembre, G., Keller, P. E., Scharf, F., Friederici, A. D., Villringer, A., et al. (2016). Syntax in action has priority over movement selection in piano playing: An ERP study. *Journal of Cognitive Neuroscience*, *28*(1), 41-54. DOI: 10.1162/jocn_a_00873

Conference Posters

Scharf, F., & Nestler, S. (2018). *Temporal EFA for ERP data generally results in correlated factors* Poster presented at 51st meeting of the German Psychological Society (DGPs), Frankfurt, Germany.

Scharf, F., & Nestler, S. (2018). *Can regularization replace simple structure rotation in Exploratory Factor Analysis?* Poster presented at ZIB [Centre for International Student Assessment] Summer Academy 2018, Kiel, Germany. Scharf, F., & Nestler, S. (2018). *Temporal EFA for ERP data generally results in correlated factors* Poster presented at 8th IMPRS NeuroCom Summer School, Leipzig, Germany.

Scharf, F., Pförtner, J., & Nestler, S. (2017). *Loading recovery in Maximum Likelihood and Bayesian Exploratory Factor Analysis*. Poster presented at 13th Conference of the Expert Group of Methods and Evaluation of the German Psychological Society, Tübingen, Germany.

Scharf, F., Müller, D., Marzecová, A., Schröger, E., & Grimm, S. (2017). *Can repetition priming explain sequential effects in temporal preparation*. Poster presented at 59th Conference of Experimental Psychologists (TEAP), Dresden, Germany.

Bianco, R., Novembre, G., Keller, P. E., Scharf, F., Friederici, A. D., Villringer, A., et al. (2015). *Musical syntax in perception and action: An fMRI study.* Poster presented at 21st Annual Meeting of the Organization for Human Brain Mapping (OHBM), Honolulu, HI, USA.

Conference Talks

Scharf, F., & Nestler, S. (2018). *Orthogonal versus Oblique Rotation in Temporal EFA for Event-related Potentials Analysis*. Talk presented presented at the 8th European Congress of Methodology in Jena, Germany.

Software Developement

Contributor to *lavaan* [R-package for structural equation modeling]

Teaching & Supervision

Seminars, Tutorials & Workshops

2018 **Seminar: Evaluation und Forschungsmethodik [Statistical analyses using R (Advanced course)]**
Statistics exercise for master students

2018 **Steuerung psychologischer Experimente [Programming psychological experiments]**
MATLAB and Psychophysics Toolbox course for master students

2018 **Seminar: Perzeptive Prozesse [Perception]**
Master course about the state of psycho-physiological research on perception

2016 – 2018 **Seminar: Computer gestützte Datenanalyse in R [Statistical analyses using R]**
Statistics exercise for 1st year bachelor students

2017 – 2018 **Introduction to Big Data (with Steffen Nestler)**
2-day workshop about Big Data analyses

2017 **Auswertung und Interpretation von Daten aus Hirnstrommessungen [Analysis and interpretation of EEG data]**
Introduction to the event-related potential technique for master students

2016 **Introduction to R**
2-day workshop for beginners

2016 **Tutorial: Statistik für Fortgeschrittene [Advanced Statistics]**
Accompanying tutorial for master students

2016 **Steuerung psychologischer Experimente [Programming psychological experiments]**
Tutor assistant during practical sessions for master students

2015 **Auswertung und Interpretation von Daten aus Hirnstrommessungen [Analysis and interpretation of EEG data]**
Tutor assistant during practical sessions for master students

2015 **Empiriepraktikum: Station EEG-Ableitung [Experimental practicals: EEG recording]**
Teaching assistant during laboratory sessions for bachelor students

2013 – 2015 **Tutorial: Multivariate Statistik [Multivariate Statistics]**
Accompanying tutorial for master students

2013 – 2015 **Tutorial: Multivariate Statistik [Multivariate Statistics]**
Accompanying tutorial for bachelor students

2012 – 2016 **Tutorial: Deskriptive & Inferenzstatistik [Descriptive & Inferential Statistics]**
Accompanying tutorial for bachelor students

Bachelor theses

2018 **Katja Moldt, B.Sc. Psychology, University of Leipzig**
[The auditory MMN to predictable pitch deviants – a replication study]
2018 **Juliane Feiler, B.Sc. Psychology, University of Leipzig**
[Reliability and stability of the auditory MMN elicited by pitch deviants across different SOAs]
2017 **Jana Pförtner, B.Sc. Psychology, University of Leipzig**
[Comparing loading and correlation estimates of EFA and PCA – a replication study]

Interns

2018 **Lena Scholz**

Selbstständigkeitserklärung

gemäß § 8(2) der Promotionsordnung

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde, und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Die Anteile meines Co-Autors Steffen Nestler an den in den Kapiteln 2, 3 und 4 erbrachten wissenschaftlichen Leistungen sind jeweils aufgeführt. Ich versichere hiermit außerdem, dass ich der alleinige Urheber der Kapitel 1 und 5 bin.

Ich versichere, dass die vorliegende Arbeit in gleicher oder in ähnlicher Form keiner anderen wissenschaftlichen Einrichtung zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde.

Es haben keine früheren erfolglosen Promotionsversuche stattgefunden.

Die Promotionsordnung der Fakultät für Lebenswissenschaften der Universität Leipzig vom 29.04.2015 ist mir bekannt und ich erkenne diese an.

Leipzig, den 23.01.2019

Florian Scharf