Regular Article

Configural frequency trees

Wolfgang Wiedermann[1] (iD), Keith C. Herman[1], Wendy Reinke[1] and Alexander von Eye[2]

[1]Department of Educational, School & Counseling Psychology and Missouri Prevention Science Institute, University of Missouri, Columbia, USA and [2]Department of Psychology, Michigan State University, East Lansing, MI, USA

Abstract

Although variable-oriented analyses are dominant in developmental psychopathology, researchers have championed a person-oriented approach that focuses on the individual as a totality. This view has methodological implications and various person-oriented methods have been developed to test person-oriented hypotheses. Configural frequency analysis (CFA) has been identified as a prime method for a person-oriented analysis of categorical data. CFA searches for configurations in cross-classifications and asks whether the number of observed cases is larger (CFA type) or smaller (CFA antitype) than expected under a probability model. The present study introduces a combination of CFA and model-based recursive partitioning (MOB) to test for type/antitype heterogeneity in the population. MOB CFA is well suited to detect complex moderation processes and can distinguish between subpopulation and population types/antitypes. Model specifications are discussed for first-order CFA and prediction CFA. Results from two simulation studies suggest that MOB CFA is able to detect moderation processes with high accuracy. Two empirical examples are given from school mental health research for illustrative purposes. The first example evaluates heterogeneity in student behavior types/antitypes, the second example focuses on the effect of a teacher classroom management intervention on student behavior. An implementation of the approach is provided in R.

The holistic-interactionistic view (e.g., Bergman & Magnusson, 1997; Magnusson, 2001) on individual functioning has influenced several developmental theories that study adaption and maladaption in individuals (Cairns, 1979; Cairns, Elder, & Costello, 1996; Cicchetti & Schneider-Rosen, 1986; Ford & Lerner, 1992; Kagan, 1994; Magnusson, 1985; Magnusson, Stattin, Damon, & Lerner, 1998; Magnusson & Allen, 1983). This view implies that when studying interrelated factors of individual behavior only in isolation (i.e., ignoring their potentially complex interactions), the factors may lose their meaning and consequence for individual behavior. Developmental psychopathology researchers have greatly benefited from incorporating person-oriented and ideographic approaches in their (variable-orientation dominated) methodological toolbox (Daukantaitė, Lundh, & Wångby-Lundh, 2019; Greenberg, Speltz, Deklyen, & Jones, 2001; Keller, Spieker, & Gilchrist, 2005; Molenaar, 2010; Mun, Bates, & Vaschillo, 2010; Sterba & Bauer, 2010; von Eye, 2010; von Eye & Bergman, 2003). The *person-oriented approach* (Bergman & Magnusson, 1997) emphasizes the individual as an integrated psychological, biological, and social organism and focuses on individual characteristics of persons, their dynamic development over time, and their variation across

contexts. This view has methodological implications which have been discussed by, for example, Bergman (2001), Bergman, Magnusson, and El Khouri (2003), and Molenaar and Campbell (2009), as it requires moving from a variable-oriented "*aggregate, then analyze*" to a person-oriented "*analyze, then aggregate*" perspective (cf., Rose, 2015). Over the course, various statistical methods have been identified as being ideally suited to evaluate person-oriented research questions (see, e.g., Bergman & Magnusson, 1997; Sterba & Bauer, 2010; von Eye, Bergman, & Hsieh, 2015; for a discussion of recent advances in person-oriented statistical methods see Wiedermann, Bergman, & von Eye, 2016).

One of the characteristic features of the person-oriented approach is that research questions are asked in *person terms* and operationalized as well as analyzed in *terms of patterns* of events or values of variables (Magnusson & Stattin, 1998). In the domain of categorical variables, *configural frequency analysis* (CFA; Lienert, 1968; von Eye, 1990) is therefore a prime method for evaluating hypotheses that are compatible with the person-oriented approach (von Eye, 2010). CFA has been successfully applied in a number of fields. We present two examples.

In a study on violent offenders, Hildebrand Karlén, Nilsson, Wallinius, Billstedt, and Hofvander (2020) used exploratory CFA, in combination with variable-oriented analyses, to identify potential configurations of early-onset externalizing problems in adolescence that are predictive of problem aggregation patterns in early adulthood in criminal behavior, substance abuse, and psychiatric comorbidity. Two types resulted. The first describes patterns of physical assault, threats and robbery with later sexual delinquency and arson. The second describes patterns of physical

assault, threats, and robbery, which are later followed by murder, sexual delinquency, and arson. This result supports Stattin and Magnusson's (1996) theory according to which patterns of externalizing problems tend to gravitate to clusters of behavior problems, and that concentration of several problem areas is prognostic for adjustment problems in adulthood.

The second example, which leads us directly to the topic of this article, can be found in psychophysiological research (Müller, Netter, & von Eye, 1997). The authors performed a two-group CFA in which they examined bivariate catecholamine response curves in normo- and hypertensive patients under stress. Results suggested in the form of types and antitypes that hypertensives respond to stress with decreases in their high tonic adrenaline values. In a two-group comparison, the authors found that hypertensives and normotensives also differ in that maladaptive patterns of response to stress are more frequent in hypertensives than in normotensives – a statement that cannot be justified by standard CFA; configural group comparison is required.

Configural frequency modeling has been extended in a number of ways, including CFA for longitudinal data (von Eye, Mun, & Bogat, 2008) and mediational processes (von Eye, Mun, & Mair, 2009). Although both of these models are person-oriented in nature, they still assume that identified extreme patterns (so-called CFA types and antitypes; Lienert, 1968) hold for the entire population under study. The study of subpopulations, for example, through the inclusion of moderator variables, constitutes a continuing methodological challenge in configural frequency modeling. In the presence of variability of effects – that is, in the presence of moderation – models of CFA have been developed by von Eye, Mair, and Mun (2010). In addition, decision tree techniques have been incorporated in configural frequency modeling by Lautsch and Ninke (2000) as well as Melcher, Lautsch, and Schmutz (2012). These authors suggested to combine chi-squared automatic interaction detectors (CHAID; Kass, 1980) with prediction CFA (P-CFA; Lienert & Krauth, 1973). In their approach, CHAID is used to detect heterogenous configurations with respect to the occurrence of an event of interest (the outcome variable), that is, a binary outcome variable is partitioned into several subgroups and covariates (partitioning variables) are used to characterize the subgroups. In a second step, P-CFA is then used to validate potential extreme cells as CFA types or antitypes (to be defined in detail below). More recently, von Eye, Wiedermann, and von Weber (2019) proposed configural and log-linear models for the analysis of decision tree structures. In essence, the authors make use of special contrasts corresponding to the decision made in a tree, and the so-called Schuster transformation (von Eye, Schuster, & Rogers, 1998; Wiedermann & von Eye, 2020b) is used to guarantee that specified contrasts can be interpreted as intended. In a second step, CFA is used to identify extreme patterns that represent the sequence of decisions. For an application see Stemmler, Heine, and Wallner (2019).

The present study continues the development of methods for the analysis of moderation effects in configural frequency modeling by making use of model-based recursive partitioning (MOB; Zeileis, Hothorn, & Hornik, 2008). Based on recursively partitioned generalized linear models (GLMs) (Rusch & Zeileis, 2013), we present the specification of *CFA trees*. The proposed approach differs from the previous attempts to combine tree-based methods and CFA. Instead of using tree methods to determine configurations of interest, we use MOB to recursively partition the CFA model itself. The MOB CFA algorithm results in a regression tree. Each terminal node of the tree is defined by a separate CFA model with potentially varying type/antitype patterns.

While the MOB CFA algorithm can be applied to a large family of CFA models (the method can be applied as long as the base model can be written as a log-linear model), in the present article, we present applications for first-order CFA (Lienert, 1968) and P-CFA (Lienert & Krauth, 1973) in the presence of complex moderation. The moderation process is described in the form of a tree leading to subgroups for which unique CFA pattern results are possible. Subgroup-specific CFA type/antitype patterns open the door to identify *population types/antitypes* (i.e., type/antitype patterns that hold for the entire set of data) and *subpopulation types/antitypes* (i.e., patterns that are specific to certain subgroups in the population). We use two empirical data examples from the field of school mental health research for illustrative purposes. In the first example, first-order MOB CFA is applied to detect (sub)population types/antitypes in student problem behavior patterns. In the second example, we apply MOB P-CFA in the context of evaluating the intervention effect of a teacher classroom management program on students' internalization problems. To demonstrate the statistical performance of the method, we present the results from two Monte Carlo simulation studies suggesting that CFA trees can be detected with sufficient statistical power in particular in the context of large samples and large variability of effects.

The article is structured as follows: First, we introduce CFA and discuss the steps taken when analyzing a cross-classification table with respect to the occurrence of configurations. Second, we provide a nontechnical introduction into recursive partitioning (MOB). Third, we discuss MOB in the context of a Poisson GLM (McCullagh & Nelder, 1989) to estimate a first-order CFA, present results of a simulation study, and illustrate the method using data from school mental health. Fourth, we present CFA tree models for P-CFA, present simulation results, and give an empirical example. Fifth, we discuss sensitivity analysis methods to evaluate the stability of CFA model trees and demonstrate their application using the two data examples. Finally, the article closes with a discussion on the implications of the findings for school mental health, potential extensions, and potential limitations of the presented approach.

An Overview of Configural Frequency Analysis

CFA (Lienert, 1968; von Eye, 1990; von Eye et al., 2010; von Eye & Gutiérrez Peña, 2004) is a widely applied method to analyze configurations of categorical variables. Here, configurations are characterized as ensembles of categories that define the cells of cross-classification tables (Lienert, 1968). In configural frequency modeling one asks whether the number of times the configurations occur in the data can be expected under a specific model, the so-called *base model* (von Eye, 2004). The orientation of deviations from expectancy classifies the deviating cell as either a CFA type (i.e., the configuration occurs more often than expected under the base model) or a CFA antitype (i.e., the configuration occurs less often than expected under the base model). Conducting CFA consists of four steps: (a) specification of a CFA base model, (b) selection of a significance procedure for type/antitype detection, (c) performing CFA under protection of the nominal significance level (protection is needed because of multiple testing), and (d) interpretation of the types/antitypes. In the following paragraphs, we give a summary of each of the steps.

Step 1: Specification of CFA base model

The CFA base model is a theory-driven probability model used to estimate expected frequencies of the configurations. Here, it is

important to note that the base model contains all effects that are *not of interest*. If this model is rejected, the effects of interest (i.e., the effects that are not part of the base model) are needed to explain the data. Therefore, CFA identifies sectors in the data space where effects are particularly noticeable. Such data regions are expressed in terms of CFA types and antitypes. The majority of CFA base models can be written as log-linear models $\log(E[m]) = \mathbf{X}\, \lambda$ with $E[m]$ being the expected frequencies, $\mathbf{X}$ is the design matrix, and λ is the vector of model parameters (for an exception see, for example, symmetry models discussed in von Eye, 2002). Base models can be global (e.g., first-order CFA), regional (e.g., *k*-group CFA), they can incorporate special effects (e.g., CFA with covariates), and they can come in the form of multistep models (e.g., mediation CFA; for a classification of CFA base model see von Eye & Wiedermann, 2021). Instead of searching for the most parsimonious model to explain variable associations (a perspective usually taken in log-linear modeling; Christensen, 1997), CFA identifies configurations that contradict the base model the strongest. Three characteristics are of central importance when formulating CFA base models: (a) interpretability, (b) parsimony, and (c) the sampling scheme. *Interpretability* refers to the fact that every base model must be specified in a way that there exists only one reason why CFA types and antitypes can occur; this reason is the effect of interest that is not part of the CFA base model. *Parsimony* implies that the number of parameters of the base model must be as small as possible while, at the same time, as complete as needed to guarantee that the emerging types/antitypes can be interpreted as intended. Finally, decisions concerning base models need to take into account the *sampling scheme* of the study (von Eye & Schuster, 1998). Here, one distinguishes between Poisson, multinomial, and product-multinomial sampling. Under Poisson sampling, neither the sample size nor the frequencies are determined a priori by the researcher. In the case of multinomial sampling, the sample size is fixed a priori, and under product-multinomial sampling, the number of observations in variable categories are determined a priori. The sampling scheme has implications for the selection of base models and significance tests discussed in the next section.

Before we move on with the second step of configural frequency modeling, selection of significance tests, we want to emphasize one important feature of log-linear model specification in the context of recursive partitioning. Application of MOB requires that the log-linear model of interest is specified on the individual level. That is, instead of regressing aggregated cell counts on the design matrix, one regresses a vector m_i (with $i = 1, \ldots, n$ indexing the participants), with $m_i = 1$ if the configuration is observed for subject i and $m_i = 0$ otherwise (also called the *expanded-data method*; Hunt, Friesen, Sama, Ryan, & Milton, 2015). The reason for this is that often continuous splitting variables (e.g., age) are included in addition to categorical splitting variables (such as gender). When all splitting variables are categorical, individual level data are no longer required. Here, however, we treat the more general individual-level case. An individual-level log-linear model can be written as

$$\log(E[m_i]) = \mathbf{X}_i\, \lambda$$

with $i = 1, \ldots, n$ referring the ith subject; $\mathbf{X}_i$ represents the subject-level design matrix consisting of $2^J \times n$ rows (with J being the number of binary variables and n representing the

sample size), and λ being the subject-level vector of parameters. Case-wise and standard aggregated log-linear models give similar parameter and standard error estimates (except for the intercept parameter which depends on the sample size; Christensen, 1997). Example specifications of (case-wise) subject-level design matrices are given below. Case-wise definitions of design matrices have been used, for example, in the context of modeling continuous subject-specific effects of choice behavior in the context of log-linear Bradley Terry models (Hatzinger & Dittrich, 2012) and in modeling covariate effects on inter-rater agreement (Hunt et al., 2015). In the present context, we make use of the expanded-data method to describe CFA base models on the level of the individual.

Step 2: Selection of significance tests

Significance tests in CFA evaluate the CFA null hypothesis $H_0: m_k = \hat{m}_k$ where m_k are the observed cell frequencies and $\hat{m}_k = E[m_k]$ denote the expected cell frequencies (k indexes the configurations and E indicates expectancy) estimated under some CFA base model. CFA types emerge when $m_k > \hat{m}_k$ and CFA antitypes emerge when $m_k < \hat{m}_k$. If the null hypothesis is retained, configuration k is neither a type nor an antitype. In the present study, we focus on the binomial test as one of the first tests that have been suggested in the context of CFA (Lienert & Krauth, 1973). The test has the advantage that it can be applied under any sampling scheme. Further, the test is nonparametric and exact (thus, there is no need to assume that a test statistic is sufficiently close to the sampling distribution). The binomial test, however, comes with lower power compared to some alternatives such as the local Pearson χ^2 test or von Eye and Mair's (2008) procedure based on standardized Pearson residuals. The power disadvantage of the binomial test disappears for large sample scenarios (von Eye & Wiedermann, 2021) which are arguably preferable when searching for population subgroups that differ qualitatively from each other.

Step 3: Performing CFA under alpha protection

To determine whether a cell significantly differs from expectation, one needs to select a nominal significance level α. Because significance tests are repeatedly performed under the same null hypothesis, α protection methods are required to avoid inflated Type I error rates (i.e., inflated rates of false positive results). The issue is further exacerbated by the fact that, in CFA, significance tests depend on each other. For example, von Weber, Lautsch, and von Eye (2003) demonstrated that, when analyzing a simple 2×2 cross-classification with a first-order CFA, the four significance tests are completely dependent upon one another.

Several approaches are available for α adjustment. Bonferroni α correction, that is, using a nominal significance level of $\alpha^* = \alpha/k$ with k indicating the number of significance tests performed, is widely accepted for this endeavor. However, several alternative α adjustment techniques (see, e.g., Dunnett & Tamhane, 1992; Hochberg, 1988; Holland & Copenhaver, 1987; Olejnik, Li, Supattathum, & Huberty, 1997) can be used which are known to be more powerful than Bonferroni adjustment. In the following applications of CFA, we make use of the more conservative Bonferroni procedure with the understanding that the performance of MOB CFA can potentially be improved by selecting more liberal α adjustment techniques.

Step 4: Interpretation of CFA types/antitypes

Following von Eye et al. (2010), the interpretation of CFA types/antitypes is based on four sources of information. The first of these is the meaning of the variable categories that define a configuration. The second source are the implications of the applied base model. For example, when the base model only includes the main effects of the variables that span the cross-classification (a so-called first-order CFA), CFA types/antitypes show where in the data space variable interactions are located (or most visible). When the base model excludes the relations between predictor and outcome variables, CFA types/antitypes indicate the predictor configurations that are particularly strongly related to specific configurations of the outcome variable(s). The third source of information is the sampling scheme (e.g., Poisson, multinomial, product-multinomial). The sampling scheme places constraints on the specification of a base model which have implications for the hypotheses that can be tested in the base model. The fourth source is information that is carried by variables that are not part of those that span the cross-classification under study (e.g., explicitly distinguishing between classification indicators used to separate independent variables). This information can be used, for instance, by including variables as covariates or moderator variables. Covariates have the potential that CFA types/antitypes emerge or disappear. Moderators can identify type/antitype patterns that differ across groups. In the present approach, these separate independent variables come in the form of splitting variables used to partition the data. In the following section, we introduce principles of recursive partitioning in the context of log-linear CFA base models.

Model-based Recursive Partitioning

Tree-based methods, valuable alternatives to standard parametric methods, have extensively been discussed in the past (see, e.g., Breiman, Friedman, Olshen, & Stone, 1984; Hothorn, Hornik, & Zeileis, 2006; Quinlan, 1993; Morgan & Sonquist, 1963; Strobl, Malley, & Tutz, 2009; Zhang & Singer, 2010). In conventional classification and regression tree (CART) and CHAID algorithms, it is the aim to recursively partition the covariate space to identify subgroups with different values of an outcome variable. In contrast, MOB (Zeileis et al., 2008) uses parameters of a predefined parametric model as the basis for recursive partitioning. The MOB algorithm partitions a set of predefined covariates (so-called splitting variables) by evaluating parameter instabilities of a parametric model. More specifically, a parametric model is formulated to represent an empirical question of interest. In the present study, the parametric model of interest is a log-linear model that defines the meaning of CFA types/antitypes, should they exist. This log-linear model is then fed into the MOB algorithm to evaluate whether covariates (splitting variables) exist which alter the parameters of the log-linear model. Making use of the GLM framework, we specify the log-linear model as Poisson regression. That is, let Y be the count response of interest with the mean $E[Y] = \mu$ (E being the expected value operator) and $\mathbf{X}$ defines a matrix collecting independent variables . The mean of Y is modeled using a linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ using $g(\mu) = \eta$ with g referring to the link function and $\boldsymbol{\beta}$ defines the vector of regression coefficients. In the context of log-linear modeling, one often replaces $\boldsymbol{\beta}$ with the parameter vector λ indicating the effects associated with contrasts in the design matrix $\mathbf{X}$. We can write the log-linear model in the form of a Poisson regression model using a log-link,

$$\log(\mu) = \mathbf{X}\,\boldsymbol{\lambda}.$$

Identifying a parameter instability in λ with respect to a partitioning covariate implies that subgroup-specific conditional effects exist in the data. The (score-based) generalized M-fluctuation test (Zeileis & Hornik, 2007) is used to assess the stability of (log-linear) model parameters along selected splitting variables. To protect the global significance level, score tests are used with Bonferroni α correction. The type of M-fluctuation test depends on the measurement level of the splitting variable. For continuous variables, the *supLM* statistic (Andrews, 1993) is employed; for categorical splitting variables χ^2 statistics by Hjort and Koning (2002) are applied. A parameter instability is captured by fitting separate models for each subgroup. To split the sample along the selected covariate, the cut-point with the largest model fit improvement is used. The algorithm terminates in end nodes when no parameter instabilities are found or when the size of the subsample is smaller than an a priori selected minimum (Rusch & Zeileis, 2013; Zeileis et al., 2008). To avoid overfitting (in particular, when sample sizes are very large), post-pruning strategies based on information criteria (such as, e.g., the Akaike information criterion; AIC) have been suggested to prune back splits that do not improve model fit (Zeileis et al., 2008). Each end node consists of a local log-linear model

$$\log(\mu_g) = \mathbf{X}_g\,\boldsymbol{\lambda}_g$$

with g indexing the detected subgroups ($g = 1, \ldots, G$). Subsequently, these G local log-linear models are treated as subgroup-specific CFA base models. Before we move on and present model tree specifications for first-order CFA (as one of the most important CFA models) and P-CFA (in which one distinguishes between predictor and outcome variables), we briefly introduce a graphical visualization of the MOB CFA end nodes.

Mosaic displays as data visualization

Regression trees offer a representation of potentially complex moderation mechanisms in the form of accessible tree structures. In addition, depending on the models displayed in the end nodes, differences in model parameters can be visualized using graphical methods as well. In the context of configural frequency modeling, partitions found via MOB CFA can be represented as n-way contingency tables. Thus, mosaic plots (or mosaic displays; Hartigan & Kleiner, 1984; Friendly, 1994, 1995; for an example see Figure 5) constitute an attractive graphical method to summarize partition-specific CFA results (for a discussion of the use of mosaic displays in the CFA context see Mun, von Eye, Fitzgerald, & Zucker, 2001).

Mosaic plots have become the primary method to visualize frequencies of a n-way contingency tables and to analyze structures of log-linear models (Friendly, 1994, 1995). In mosaic plots, cell frequencies are represented by tiles (rectangles) that are proportional to the cell frequencies of the cross-table. The idea is best introduced using a simple two-way table. The expected frequencies under independence can be displayed as tiles whose widths are proportional to the total frequency in each column (the first variable), and whose heights are proportional to the total frequency in each row (the second variable). Then, the area of the tile is proportional to the cell frequency. Specifically, a unit

area is divided into bars with width equal to the marginal frequencies (probabilities) of the first variable, and these bars are divided into tiles with heights being equal to the conditional frequencies (probabilities) of the second variable. The resulting area reflects the total quantity (cell frequencies) in the cross-table of the variables. When two variables are independent, the tiles of the mosaic plot will align (i.e., all tiles in each row have the same height). Not aligning tiles indicate dependency structures between the two variables. Several pattern detection approaches can be imposed on the mosaic plot through making use of coloring and shading of tiles (Friendly, 1994).

In the present context, feasibility of mosaic plots depends on the number of variables involved and the complexity of the base model. Here, the maximum number of variables to meaningfully use mosaic displays simply depends on the readers' familiarity with interpreting mosaic plots. When the model is too complex to be visualized in an accessible way, simple CFA summary tables can be used instead of mosaic plots. In the first empirical example given below, we use CFA summary tables, in the second example, we give an example using mosaic displays. Although readers may be required to familiarize themselves with decoding statistical information from mosaic displays first, this type of visualization is ideally suited to display MOB CFA differences across partitions. In its simplest form, mosaic plots can be generated for the *g*th contingency table. In addition, each partition-specific mosaic plot can be shaded according to the status of the configuration as "no deviation" (marked gray in the plots given in Figures 5 and 6), type (marked green), or antitype (marked red), that is, the pattern detection method is the presence/absence of types/antitypes itself. However, more complex representations that take into account fit information of partition-specific log-linear models can be used as well (cf. Friendly, 1994, 1995).

First-order CFA Trees

The first CFA tree model discussed here builds on the first-order CFA base model. This model constitutes the classic version of CFA and was originally proposed by Lienert (1968). The model only takes main effects of all variables into account and, thus, assumes independence of variables. If CFA types/antitypes exist, they can only emerge due to the presence of interactions (of any order) among variables. For J variables (X_j, $j = 1, …, J$), the first-order base model can be written as

$$\log(E[m]) = \lambda + \sum_{j=1}^{J} \lambda^{X_j}$$

with λ being the intercept and λ^{X_j} being the main effect parameter for X_j. For $J = 3$ binary variables (i.e., in case of a $2 \times 2 \times 2$ cross contingency table), the effect coded design matrix $\mathbf{X}$ for this model is

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix}.$$

The first column of the design matrix represents the intercept, the second column specifies the main effect for the first variable (X_1), and the third and fourth column give the main effects for the second and third variable (X_2 and X_3). Each column contrasts the two categories of each variable (e.g., the presence of a symptom is coded with 1 and the absence of the symptom is indicated by −1).

In CFA trees, we assume that parameters of the base model can vary across subgroups (partitions) of the data. Formally, the first-order MOB CFA model for $g = 1, …, G$ subgroups and $j = 1, …, J$ indicators is given by

$$\log(E[m]_g) = \lambda_g + \sum_{j=1}^{J} \lambda_g^{X_j}$$

with λ_g being the subgroup-specific intercept and $\lambda_g^{X_j}$ specifying the subgroup-specific main effects. Varying log-linear model parameters across subgroups open the door to the presence of subgroup-specific CFA types and antitypes. Splitting variables of CFA trees can be conceptualized as moderators. Tree estimation is based on a reformulation of the log-linear model as a Poisson count regression model (often referred to as the "Poisson trick") making use of the fact that Poisson counts are jointly multinomially distributed (Agresti, 2002). To apply the MOB algorithm, one requires a case-wise design matrix for the Poisson regression model. To give an example, consider the case of two individuals, three binary classification variables (X_1–X_3; 0 = symptom absence, 1 = symptom present), and two continuous splitting variables U and V. For Person 1, suppose we observe $X_1 = X_2 = X_3 = 0$ (i.e., the configuration {0 0 0}) as well as $U = u_1$ and $V = v_1$. In contrast, for Person 2, we obtain the configuration {1 0 0} together with $U = u_2$ and $V = v_2$. The corresponding case-wise design matrix for the two participants then takes the form

Person	m	X_1	X_2	X_3	U	V
1	1	0	0	0	u_1	v_1
1	0	1	0	0	u_1	v_1
1	0	0	1	0	u_1	v_1
1	0	1	1	0	u_1	v_1
1	0	0	0	1	u_1	v_1
1	0	1	0	1	u_1	v_1
1	0	0	1	1	u_1	v_1
1	0	1	1	1	u_1	v_1
2	0	0	0	0	u_2	v_2
2	1	1	0	0	u_2	v_2
2	0	0	1	0	u_2	v_2
2	0	1	1	0	u_2	v_2
2	0	0	0	1	u_2	v_2
2	0	1	0	1	u_2	v_2
2	0	0	1	1	u_2	v_2
2	0	1	1	1	u_2	v_2

where m gives the observed frequencies for all possible patterns based on the three X indicators (the intercept is implied). Because {0 0 0} was observed for the first person, we use $m = 1$ for this particular configuration, and $m = 0$ for the remaining seven configurations. Similarly, for Person 2, we use $m = 1$ for {1 0 0} and $m = 0$ otherwise. For both participants, values of the splitting values U and V (u_1, u_2, v_1, and v_2) are each repeated eight times according to the number of configurations. The resulting design matrix contains 16 rows (eight possible patterns for each of two participants). More generally, for n participants and J binary indicators, the number of rows of the case-wise design matrix is given by $n \cdot 2^J$. In the corresponding MOB Poisson regression model, m is treated as the response, the dummy variables for X_1–X_3 are used as independent variables (effect coding could be applied as well), and U and V serve as splitting variables. Within MOB, generalized M-fluctuation tests (Zeileis & Hornik, 2007) are used to assess stability of model parameters for the indicators X_1, X_2, and X_3 along the splitting variables U and V. The resulting partitions can be interpreted as terminal nodes of a CFA tree containing separate first-order CFA models with varying parameters and (potentially) partition-specific types and antitypes. In the following section, we illustrate the performance of first-order MOB CFA using Monte Carlo simulations.

Simulation Study I

To illustrate the performance of MOB CFA in the context of a first-order base model, a simulation study was performed using the R statistical programing environment (R Core Team, 2020). The data generating mechanism is summarized in Figure 1. Four subgroups based on three relevant partitioning variables, Z_1, Z_2, and Z_5, among a set of n_{cov} covariates were generated. The number of covariates (Z_1, …, Z_{ncov}) was either $n_{cov} = 5$ or 10. Following Dusseldorp and Van Mechelen (2014), Fokkema, Smits, Zeileis, Hothorn, and Kelderman (2018), and van Wie, Li, and Wiedermann (2019), all covariates were randomly drawn from a multivariate normal distribution with means for Z_1, Z_2, Z_4, and Z_5 of 10, 30, −40, and 70. The means of the remaining covariates (Z_3 and depending on the value of n_{cov}, Z_6, …, Z_{10}) were randomly generated from the uniform distribution on the interval [−70, 70]. The standard deviation of all covariates was set to 10 and pairwise correlations of all covariates were fixed at 0.3. Following Schrepp (2006), simulated configurations were based on a set of dichotomous indicators, for example, representing the absence or presence of symptoms that occur randomly in the underlying population (0 = symptom absent, 1 = symptom present). In the simulation, the number of indicators was set to 5 and the probabilities of observing the symptoms were randomly drawn from a uniform distribution with a minimum of 0.05 and a maximum of 0.15. To simulate the occurrence of CFA types, in the next step, observed (individual-level) symptom configurations were replaced with predefined type configurations with probabilities $\delta = 0.20$, 0.25, 0.30, and 0.35. The predefined type configurations varied across subgroups. In Subgroup 1 (individuals for which $Z_1 \leq 17$ and $Z_2 \leq 30$), the configuration {1 1 1 1 1} constitutes a CFA type, in Subgroups 2 ($Z_1 > 17$ and $Z_2 \leq 30$) and 3 ($Z_2 > 30$ and $Z_5 \leq 63$) no types occurs, and in Subgroup 4 ($Z_2 > 30$ and $Z_5 > 63$) the configuration {0 1 1 1 1} is a CFA type. In other words, in one subgroup, a CFA type was characterized by the occurrence of all five symptoms, and in a second subgroup, a CFA type is characterized by the absence of the first symptom and the presence of the remaining four symptoms.
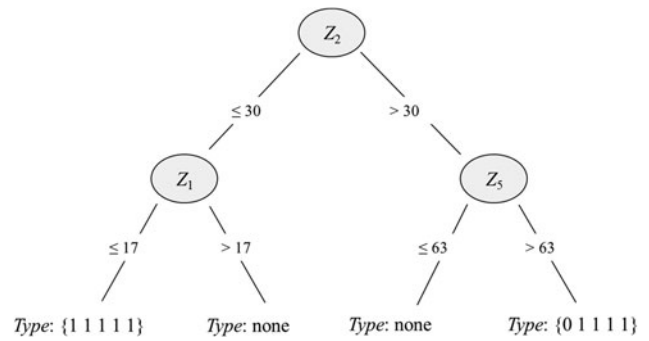

Figure 1. Data-generating mechanism of simulated first-order configural frequency analysis (CFA) trees based on three splitting variables Z_1, Z_2, and Z_5.

Sample sizes were set to $n = 500$, 1,000, and 1,500. The simulation design was fully crossed resulting in 2 (number of covariates) × 3 (sample size) × 4 (effect size δ) = 24 simulation conditions. For each simulation condition, 100 samples were generated.

For each generated sample, we computed the subject-specific (effect coded) design matrix of a first-order CFA for the five binary indicators to explain the observed configuration frequencies. MOB was used to identify (potentially) better fitting local first-order CFA models. Bonferroni-corrected parameter stability tests were performed using a nominal significance level of 5%. Regression trees were estimated with a maximal depth of three, limiting the number of terminal nodes to eight. The function GLM tree of the R package *partykit* (Hothorn & Zeileis, 2015) was used for MOB estimation. Next, for each subgroup identified through MOB, we performed a separate first-order CFA with Bonferroni-corrected binomial tests (implemented in the R package *confreq*, Stemmler & Heine, 2017) to evaluate the presence of types and antitypes.

Results

Figure 2 summarizes the performance of the MOB CFA algorithm to identify the correct tree structure. We focus on the size and the overall accuracy of the estimated trees. Following Fokkema et al. (2018) and van Wie et al. (2019), we defined an accurately recovered tree as one that (a) consists of seven nodes in total (three splitting and four terminal nodes), (b) uses Z_2 as the first splitting variable with a value of 30 ± 5 (i.e., ± half the population standard deviations of the partitioning covariates), (c) identifies Z_1 as the next split on the left with 17 ± 5, and (d) uses Z_5 as the final splitting variable on the right side of the tree with a value of 63 ± 5.

The upper panel of Figure 2 gives the average tree size as a function of the number of covariates, sample size, and the probability to observe a subgroup-specific CFA type (δ). For large sample sizes ($n = 1,500$), the average tree size is close to the true one, independent of the number of covariates and the level of δ. For smaller sample sizes, the average tree size tends to be smaller, in particular, when the probability of CFA types is small as well. For example, for $n = 1,000$ and $\delta = 0.20$, we obtain an average tree size of 6.77 ($SD = 0.70$). For $n = 500$ and $\delta = 0.20$, the average tree size declines to 3.11 nodes on average ($SD = 1.10$). Similarly, a large number of covariates also slightly decreases the average tree size. This effect is, however, mitigated for larger samples. For $n = 500$, 1,000, and 1,500, we observe an average tree sizes of 4.38 ($SD = 1.57$), 6.52 ($SD = 0.98$), and 6.98 ($SD = 0.22$) if $n_{cov} = 5$, and 4.15 ($SD = 1.49$), 6.38 ($SD = 1.15$),
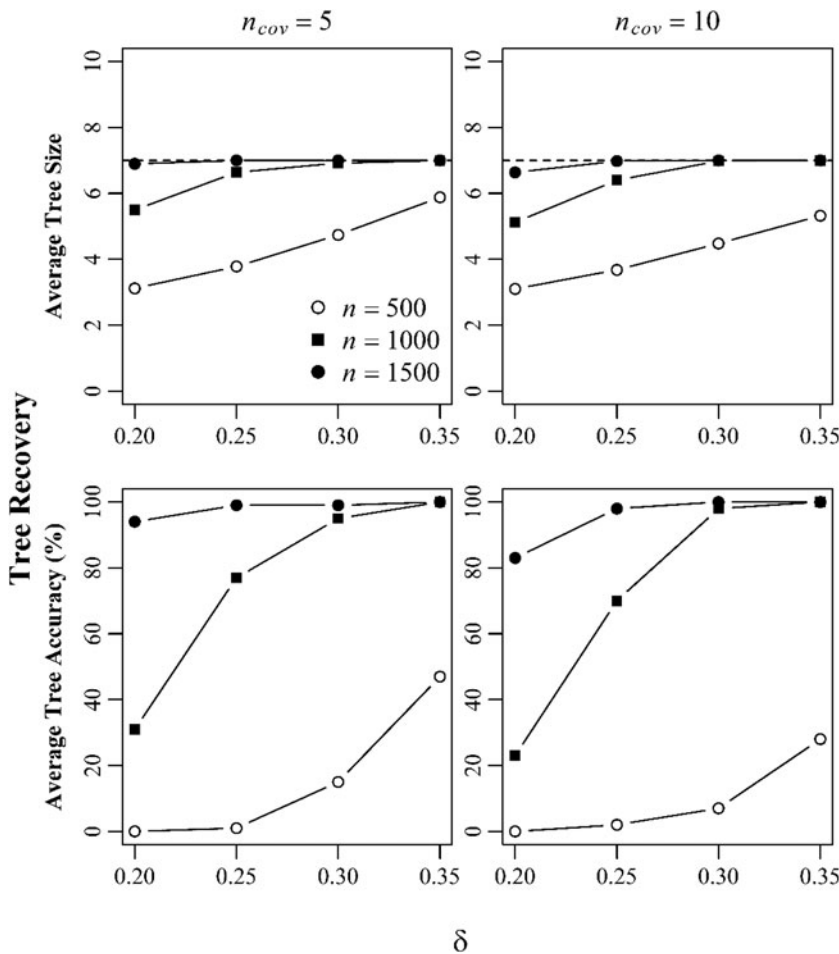
Figure 2. Average tree size and accuracy of recursively partitioned first-order configural frequency analysis (CFA) as a function of sample size (n), number of covariates (n_{cov}), and probability of CFA types (δ).

and 6.91 ($SD = 0.47$) if the number of covariates increases to $n_{cov} = 10$.

Next, we focus on the tree accuracy (cf. Figure 2 lower panel). As expected, accuracy heavily depends on sample size and the probability of CFA types (δ). For $n = 1,500$, for example, tree accuracy is close to 100% across all level of δ, except for $n_{cov} = 10$ and $\delta = 0.20$ (however, tree accuracy is still larger than 80% in this data condition). For $n = 1,000$, larger effect sizes ($\delta > 0.25$) are needed to achieve adequate accuracy. In contrast, tree accuracy tends to be low for small samples ($n = 500$) and slightly decreases with the number of covariates (again the latter effect is mitigated by larger samples). Overall, we can conclude that the MOB CFA is able to identify the true data-generating model, however, larger samples are generally preferable when studying CFA tree structures.

Finally, we focus on the percentage of correctly identified CFA types per subgroup. Here, we focus on cases in which MOB CFA correctly returned four subgroups. In Subgroup 1, the percentage of detecting the configuration {1 1 1 1 1} as a CFA type ranged from 96.2% to 100% across all simulation conditions. In Subgroup 4, the configuration {0 1 1 1 1} was detected in 100% of the cases. Here, the case of a small sample size ($n = 500$), a small effect size ($\delta = 0.20$), and a large number of covariates ($n_{cov} = 10$) constitutes an exception. In this extreme case, MOB CFA correctly identified the four subgroups in only one simulated sample (in 70% of the simulated samples, a two-group solution was preferred) and the binomial test for the configuration {0 1 1 1 1} was nonsignificant in this sample. Overall, however, we

can conclude that MOB CFA is able to determine subgroup-specific CFA types with high probability.

Empirical Example 1

Next, we illustrate the application of first-order MOB CFA using a real-world example. The data we use come from a large-scale randomized controlled trial (RCT) that evaluates the impact of the Incredible Years Teacher Classroom Management (IY TCM) program on student social–emotional, disruptive behavior, and academic outcomes (cf. Reinke, Herman, & Dong, 2018). The IY TCM program is designed to promote effective classroom management practices for preschool and early elementary teachers and focuses on promoting students' prosocial skills, increasing parents' involvement, and positive teacher–parent relationships (Webster-Stratton, Reid, & Hammond, 2004). The original study focused on evaluating the effectiveness of the IY TCM program in students in kindergarten to third grade from nine schools in a school district in the Midwestern part of the United States. In the following re-analysis, we focus on students' behavior at study baseline using data obtained from 1,630 students (104 teachers) with valid baseline and post-treatment measures (Wiedermann, Reinke, & Herman, 2020).

Student behavior, that is concentration problems (C), disruptive behavior (D), emotional dysregulation (E), internalization problems (I), and family problems (F), was assessed with the *Teacher Observation of Classroom Adaptation-Checklist*

Table 1. Demographic statistics. Values in parentheses give percentages

	Control (n = 815)	Treatment (n = 815)	Total (n = 1,630)
Concentr. Probl.	86 (10.6)	85 (10.4)	171 (10.5)
Disruptive Behav.	79 (9.7)	88 (10.8)	167 (10.2)
Emot. Dysregul.	79 (9.7)	95 (11.7)	174 (10.7)
Internal. Probl.	83 (10.2)	81 (9.9)	164 (10.1)
Family Probl.	127 (15.6)	103 (12.6)	230 (14.1)
Race: Black	608 (74.6)	619 (76)	1,227 (75.3)
Gender: Female	403 (49.4)	391 (48)	794 (48.7)
Grade			
K	232 (28.5)	231 (28.3)	463 (28.4)
1st	215 (26.4)	224 (27.5)	439 (26.9)
2nd	249 (30.6)	157 (19.3)	406 (24.9)
3rd	119 (14.6)	203 (24.9)	322 (19.8)
School			
(1)	76 (9.3)	84 (10.3)	160 (9.8)
(2)	95 (11.7)	76 (9.3)	171 (10.5)
(3)	83 (10.2)	106 (13.0)	189 (11.6)
(4)	110 (13.5)	99 (12.1)	209 (12.8)
(5)	95 (11.7)	77 (9.4)	172 (10.6)
(6)	56 (6.9)	71 (8.7)	127 (7.8)
(7)	105 (12.9)	118 (14.5)	223 (13.7)
(8)	97 (11.9)	72 (8.8)	169 (10.4)
(9)	98 (12.0)	112 (13.7)	210 (12.9)

(TOCA-C; Koth, Bradshaw, & Leaf, 2009). Internal consistency (evaluated via Cronbach's alpha) for the subscales ranged from 0.70 to 0.94. To evaluate potential patterns of students' problematic behavior, we dichotomized subscale-specific mean composite scores. Problematic behavior was defined as subscale scores at or beyond the 90th sample percentile (0 = TOCA-C subscale score <90th percentile, 1 = TOCA-C subscale score ≥90th percentile). Binary problem behavior indicators were analyzed using MOB CFA with a first-order base model, assuming independence between problem behavior indicators. Gender (0 = male, 1 = female), grade (K-3rd grade), race (1 = black, 0 = others), and school membership were used as splitting variables. Descriptive statistics of all variables for the overall sample and by treatment status are given in Table 1.

Before using recursive partitioning to find subgroups of problem behavior patterns, we estimate a first-order CFA for the total sample without considering splitting variables as a benchmark. The LR goodness of fit test rejects the model of independence for the total sample ($\chi^2(26) = 809.27$, $p < .001$). Therefore, we expect types and/or antitypes to occur. Table 2 gives the expected frequencies, p values of cell-wise binomial tests, and CFA decisions based on a Bonferroni-corrected significance level of 0.05/ 32 = 0.00156. Overall, we observe 10 types and seven antitypes. We refrain from a detailed interpretation of these 17 extreme cells. Instead, we highlight some important aspects of the results

and use them as a benchmark for comparison with results obtained from first-order MOB CFA. Some aspects of the results in Table 2 are quite typical for CFA. For example, the largest cell describing the absence of any problem behavior {0 0 0 0 0} constitutes a CFA type. Here, 901.3 cases (55.3% of the sample) were expected but 1,097 (67.3%) were observed. The second largest cell, configuration {1 0 0 0 1} describing students with severe concentration problems and family issues, also constitutes a CFA type; about 17.7 (1.2%) were expected and 98 (6.0%) were observed. Interestingly, the third and fourth largest cells in the cross-table determine CFA antitypes. These are the configurations {0 0 0 0 1} (i.e., severe family problems only; expected to be observed in about 148.5 cases [9.1%], but only observed for 74 students [4.5%]) and {0 0 0 1 0} (i.e., internalization problems only; expected for about 101.3 students [6.2%], however, only observed for 73 students [4.5%]). Other important CFA types are the configurations {0 1 1 0 0} and {0 1 1 1 0}. The first one describes students with severe disruptive behavior as well as severe emotional dysregulation which occurred in 47 cases (2.9%). However, according to the independence model, one would expect only 12.5 cases (0.8%). The second configuration describes students with severe disruptive behavior, emotional dysregulation, and internalization problems. Here, only 1.4 cases (0.09%) were expected, but 24 students (1.5%) were observed.

Figure 3 gives the estimated MOB CFA tree using similar specifications as in Simulation Study 1. Three of the four covariates were identified as important splitting variables: school membership, race, and gender. First, MOB separates three schools (Schools 4, 5, and 6) from the remaining six (Schools 1, 2, 3, 7, 8, and 9). Further, in the three separated schools, students' race constitutes an important moderating factor. In contrast, for the remaining six schools, gender instead of race proves to be an important influence. The estimated tree results in four subgroups (terminal nodes). Subgroup 1 consists of non-black students in Schools 4, 5, 6 (n = 146, 9.0%), Subgroup 2 are black students in the same schools (n = 362, 22.2%), Subgroup 3 describes male students in the Schools 1, 2, 3, 7, 8, and 9 (n = 554, 34.0%), and the last subgroup consists of female students in one of the six schools (n = 568, 34.8%). Each terminal node consists of a separate first-order CFA base model. Bonferroni-corrected binomial tests were used to test whether types or antitypes are present per subgroup. The LR goodness of fit test rejects the model of independence in all four subgroups (LR χ^2 statistics ranged from 75.98 to 337.78 with 26 degrees of freedom and all ps < .001). Therefore, as expected, types and antitypes emerge in every subgroup and, more important, type/antitype patterns vary across subgroups where some patterns can be observed in every subgroup and some patterns are unique to specific subgroups. In the following, we interpret some of the results presented in Figure 3 with respect to the occurrence of types/ antitypes at the population and subpopulation levels. The type pattern {0 0 0 0 0}, describing the absence of problem behavior, occurs in every subgroup indicating that this is a *population type* (i.e., one that is observed across all partitions of the sample). Similarly, the antitype pattern {1 0 0 0 0} is also observed in every subgroup and can, thus, be conceptualized as a *population antitype*. The second largest type for the total sample, {1 0 0 0 1}, is specific to the Subgroups 1 and 2 (i.e., students in the Schools 4, 5, and 6), and, therefore, describes a *subpopulation type* (i.e., a type that occurs in specific partitions of the data). In other words, the occurrence of severe concentration problems

Table 2. First-order configural frequency analysis (CFA) results for the total sample ($n = 1{,}630$)

C	D	E	I	F	m	$E[m]$	p value	Decision
0	0	0	0	0	1,097	901.33	0.00000	Type
1	0	0	0	0	38	107.49	0.00000	Antitype
0	1	0	0	0	48	104.05	0.00000	Antitype
1	1	0	0	0	3	12.41	0.00162	–
0	0	1	0	0	27	108.18	0.00000	Antitype
1	0	1	0	0	0	12.9	0.00003	Antitype
0	1	1	0	0	47	12.49	0.00000	Type
1	1	1	0	0	7	1.49	0.00088	Type
0	0	0	1	0	73	101.31	0.00146	Antitype
1	0	0	1	0	4	12.08	0.00703	–
0	1	0	1	0	6	11.69	0.05362	–
1	1	0	1	0	0	1.39	0.40639	–
0	0	1	1	0	20	12.16	0.02353	–
1	0	1	1	0	2	1.45	0.42542	–
0	1	1	1	0	24	1.4	0.00000	Type
1	1	1	1	0	4	0.17	0.00003	Type
0	0	0	0	1	74	148.51	0.00000	Antitype
1	0	0	0	1	98	17.71	0.00000	Type
0	1	0	0	1	2	17.14	0.00001	Antitype
1	1	0	0	1	0	2.04	0.39392	–
0	0	1	0	1	10	17.82	0.03240	–
1	0	1	0	1	1	2.13	0.37284	–
0	1	1	0	1	9	2.06	0.00029	Type
1	1	1	0	1	5	0.25	0.00001	Type
0	0	0	1	1	7	16.69	0.00637	–
1	0	0	1	1	4	1.99	0.14109	–
0	1	0	1	1	1	1.93	0.42601	–
1	1	0	1	1	1	0.23	0.20532	–
0	0	1	1	1	6	2	0.01662	–
1	0	1	1	1	2	0.24	0.02437	–
0	1	1	1	1	8	0.23	0.00000	Type
1	1	1	1	1	2	0.03	0.00037	Type

Note. C = concentration problems, D = disruptive behavior, E = emotional dysregulation, I = internalization problems, F = family problems, m = observed frequencies, $E[m]$ = expected frequencies. Bonferroni-corrected $\alpha = 0.00156$.

and family issues are specific to the three schools described above. The largest antitype according to the CFA results based on the total sample – family issues only ({0 0 0 0 1}) – is, in fact, specific to Subgroups 3 and 4 and constitutes a *subpopulation antitype*. In other words, family issues only occur with lower probability than expected in the remaining six schools. This effect does not depend on students' gender.

Prediction CFA Trees

P-CFA has been proposed by Lienert and Krauth (1973) and extended by von Eye and Bogat (2005) and von Eye, Mair, and

Bogat (2005). The main difference between standard CFA and P-CFA concerns the status of the variables involved. In standard CFA (e.g., the first-order CFA discussed above) all variables are treated as exploratory variables. In contrast, in P-CFA one explicitly distinguishes between predictor (independent) and outcome (dependent) variables with the goal to identify predictor patterns that are associated with certain outcomes. Outcomes can come in the form of specific single events (e.g., receiving a clinical diagnosis) or in the form of specific patterns associated with multiple outcome variables (e.g., showing a clinically relevant behavior pattern). Similar to standard CFA, P-CFA does not ask whether predictor variables are related to outcome variables (these types of
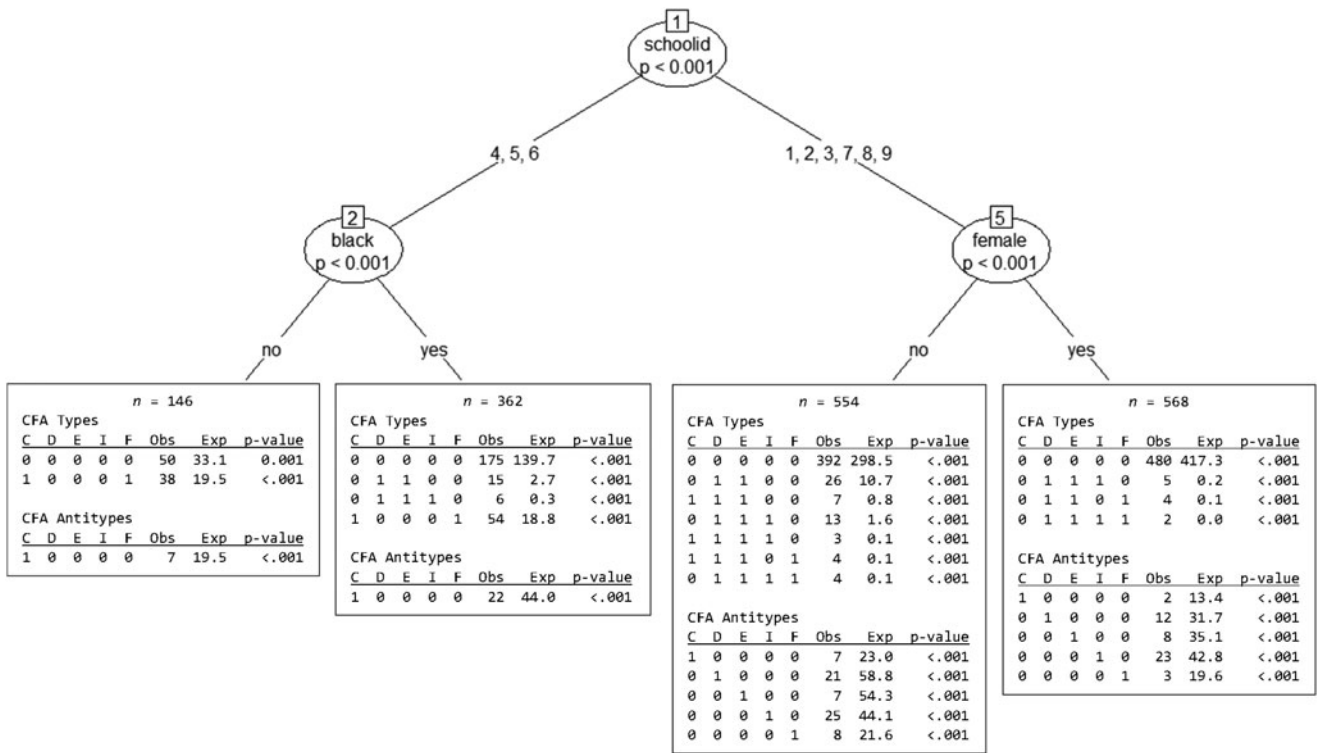
Figure 3. Recursively partitioned first-order configural frequency analysis of student problem behavior indicators (C = concentration problems, D = disruptive behavior, E = emotional dysregulation, I = internalization problems, F = family problems) based on $n = 1{,}630$ students.

questions are usually addressed using (multinomial) logistic regression models). Instead, P-CFA asks whether specific patterns of categorical predictors go hand-in-hand with specific outcome patterns (von Eye et al., 2005). To introduce P-CFA, let Y be a binary outcome variable and X_1 and X_2 are two binary predictors. Crossed, the three variables span a $2 \times 2 \times 2$ cross contingency table. The P-CFA base model proposed by Lienert and Krauth (1973) has the following characteristics: (a) the base model is saturated in the predictor variables (thus, types/antitypes cannot emerge due to the association among predictors), (b) the base model is saturated in the outcome variables (thus, types/antitypes cannot emerge due to the association among outcome variables), and (c) the base model assumes independence between predictor and outcome variables. In other words, types/antitypes can only emerge due to the presence of a predictor–outcome association. Specifically, if, for a specific predictor pattern, more cases are observed than expected under the P-CFA base model, the predictor pattern is said to predict the occurrence of the outcome pattern and, thus, constitutes a prediction type. In contrast, if fewer cases are observed, the predictor pattern is said to predict the nonoccurrence of outcome patterns and constitutes a prediction antitype. For the three variables (Y, X_1, and X_2), the corresponding base model is

$$\log(E[m]) = \lambda + \lambda^{X_1} + \lambda^{X_2} + \lambda^{Y} + \lambda^{X_1 X_2}$$

with λ being the intercept, λ^{X_1}, λ^{X_2}, and λ^{Y} being the main effects of the three variables, and $\lambda^{X_1 X_2}$ specifies the two-way interaction on the predictor side of the model. Terms that are missing in the base model are those that link the predictors (X_1 and X_2) to the outcome Y. The corresponding effect coded design matrix of

the model is

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 \end{pmatrix},$$

where the first column represents the intercept (λ), columns 2–4 indicate the three main effects for X_1, X_2, and Y, and the last column represents the interaction term $X_1 X_2$.

In contrast to standard P-CFA, the proposed P-CFA tree approach assumes that subgroups ($g = 1, \ldots, G$) exist for which potentially different prediction type/antitype patterns occur. P-CFA types/antitypes that exist for all G subgroups constitute *population prediction types/antitypes* and types/antitypes that are unique to specific subgroups describe *subpopulation prediction types/antitypes*. Again, the P-CFA tree is constructed using a priori specified splitting variables which can be categorical or continuous in nature. Formally, for three binary variables (X_1, X_2, and Y), the MOB P-CFA model for $g = 1, \ldots, G$ subgroups can be written as

$$\log(E[m]_g) = \lambda_g + \lambda_g^{X_1} + \lambda_g^{X_2} + \lambda_g^{Y} + \lambda_g^{X_1 X_2}$$

with λ_g representing the subgroup-specific intercept, $\lambda_g^{X_1}$, $\lambda_g^{X_2}$, λ_g^{Y} referring to the subgroup-specific main effects, and $\lambda_g^{X_1 X_2}$ being the subgroup-specific interaction effect. Because log-linear

model parameters are allowed to vary across subgroups, the algorithm is able to identify subgroup-specific prediction types and antitypes. To implement MOB P-CFA, one requires a case-wise design matrix. Consider two individuals, three binary variables (two predictors X_1 and X_2 and one outcome Y; 0 = event not observed, 1 = event observed), and two continuous splitting variables U and V. For Person 1, suppose we observe $X_1 = X_2 = Y = 1$, that is the configuration {1 1 1}, together with $U = u_1$ and $V = v_1$. For Person 2, we observe the configuration {0 1 0} and $U = u_2$ and $V = v_2$. Then, the corresponding case-wise design matrix for the two participants is (the intercept is implied in the model matrix)

Person	m	X_1	X_2	Y	$X_1 X_2$	U	V
1	0	0	0	0	0	u_1	v_1
1	0	1	0	0	0	u_1	v_1
1	0	0	1	0	0	u_1	v_1
1	0	1	1	0	1	u_1	v_1
1	0	0	0	1	0	u_1	v_1
1	0	1	0	1	0	u_1	v_1
1	0	0	1	1	0	u_1	v_1
1	1	1	1	1	1	u_1	v_1
2	0	0	0	0	0	u_2	v_2
2	0	1	0	0	0	u_2	v_2
2	1	0	1	0	0	u_2	v_2
2	0	1	1	0	1	u_2	v_2
2	0	0	0	1	0	u_2	v_2
2	0	1	0	1	0	u_2	v_2
2	0	0	1	1	0	u_2	v_2
2	0	1	1	1	1	u_2	v_2

with m being the observed frequencies for all possible patterns. For Person 1, one uses $m = 1$ for the observed pattern {1 1 1} and $m = 0$ for the remaining configurations. In a similar fashion, for Person 2, one uses $m = 1$ for the configuration {0 1 0} and $m = 0$ otherwise. Similar to the first-order MOB CFA described above, values of the splitting values U and V are each repeated eight times for each person. To estimate the MOB P-CFA, one uses a MOB Poisson regression model where m serves as the response, the four dummy variables for X_1, X_2, Y, and $X_1 X_2$ are the independent variables, and U and V serve as MOB splitting variables. Again, generalized M-fluctuation tests (Zeileis & Hornik, 2007) are used to detect instabilities of model parameters along the splitting variables. Parameter instabilities are captured by fitting separate Poisson regression models per subgroup. The resulting partitions represent terminal nodes of a P-CFA tree. Each terminal node contains a separate P-CFA model with partition-specific parameters and (potentially) prediction types and antitypes. To evaluate the performance of MOB P-CFA, we performed a second Monte Carlo simulation experiment.

Simulation Study 2

In the second Monte Carlo simulation experiment, we focus on a data scenario in which both, a variable- and a person-oriented effect exist simultaneously. Specifically, we focus on a pre–post comparison for two experimental groups, a treatment group and a business as usual group, and a binary outcome variable Y (0 = event not observed, 1 = event observed). The data-generating mechanism followed a logistic regression model with one outcome and two predictors, that is, $\log(p/[1 - p]) = \beta_0 + \beta_1 T + \beta_2 X$ with p being the probability of the event occurring ($Y = 1$), T being a treatment indicator (0 = business as usual, 1 = treatment), and X being a continuous covariate (representing continuous pre-treatment scores). T was randomly drawn from a binomial distribution with a probability of $T = 1$ of 50% and X was generated from a standard normal distribution. Mimicking properties of a successful random assignment of subjects to either the control or treatment group, T and X were generated independently with $\text{Cor}(T, X) = 0$. Following Huang (2019), the intercept β_0 was set to zero (representing the predicted probability in the reference group of 50%), the treatment effect was set to $\beta_1 = 0.5$, and the covariate effect was $\beta_2 = 1$. The outcome was sampled from a Bernoulli distribution with probability $p = \exp(\beta_0 + \beta_1 T + \beta_2 X)/[1 + \exp(\beta_0 + \beta_1 T + \beta_2 X)]$.

Data were generated using the same regression tree as in Simulation Study 1 (cf. Figure 1). That is, four subgroups were simulated using the three partitioning variables Z_1, Z_2, and Z_5 among a set of n_{cov} covariates. The overall number of covariates was $n_{cov} = 5$ or 10. Because P-CFA assumes that variables are categorical in nature, the continuous X variable was dichotomized at the mean. In the first subgroup (i.e., if $Z_1 \leq 17$ and $Z_2 \leq 30$), observed configuration of X, T, and Y were replaced with the type configuration {1 1 1} with probabilities $\delta = 0.20$, 0.25, 0.30, and 0.35. In the second and third subgroups (i.e., if $Z_1 > 17$ and $Z_2 \leq 30$ or if $Z_2 > 30$ and $Z_5 \leq 63$) no types occurred. In the fourth subgroup ($Z_2 > 30$ and $Z_5 > 63$), the type configuration was {0 1 1} occurred with probability δ. In other words, in Subgroup 1, the occurrence of $Y = 1$ is associated with $T = 1$ and $X = 1$ (i.e., X scores above the sample mean) which is also in line with the overall variable-oriented effect (i.e., large X scores and $T = 1$ increase the chance of $Y = 1$). In contrast, in Subgroup 4, $Y = 1$ is likely to occur together with $T = 1$ and $X = 0$ (i.e., X scores below the sample mean). Note that the latter person-oriented effect contradicts the global variable-oriented effect. Here, lower X scores and $T = 1$ is likely to be observed together with $Y = 1$. Overall sample sizes were set to $n = 500$, 1,000, and 1,500. The simulation design was fully crossed resulting in 2 (number of covariates) × 3 (sample size) × 4 (effect size δ) = 24 simulation conditions. One hundred samples were generated per condition.

For each sample, the case-specific design matrix of the P-CFA was generated to predict the observed configuration frequencies. Next, MOB was employed to identify P-CFA subgroups. MOB estimation was based on a GLM tree with a log link and a Poisson error distribution. The variables $Z_1, \ldots, Z_{ncov}$ were used as potential splitting variables. Bonferroni-corrected parameter stability tests were performed with a familywise nominal significance level of 5%. Similar to Simulation Study 1, regression trees were estimated with a maximal depth of three, limiting the number of terminal nodes to eight. Finally, for each identified subgroup, we estimated a separate P-CFA with Bonferroni-corrected binomial tests to test for the presence of prediction types and antitypes.

Results

Figure 4 summarizes the performance of the MOB P-CFA approach as a function of sample size, number of covariates,
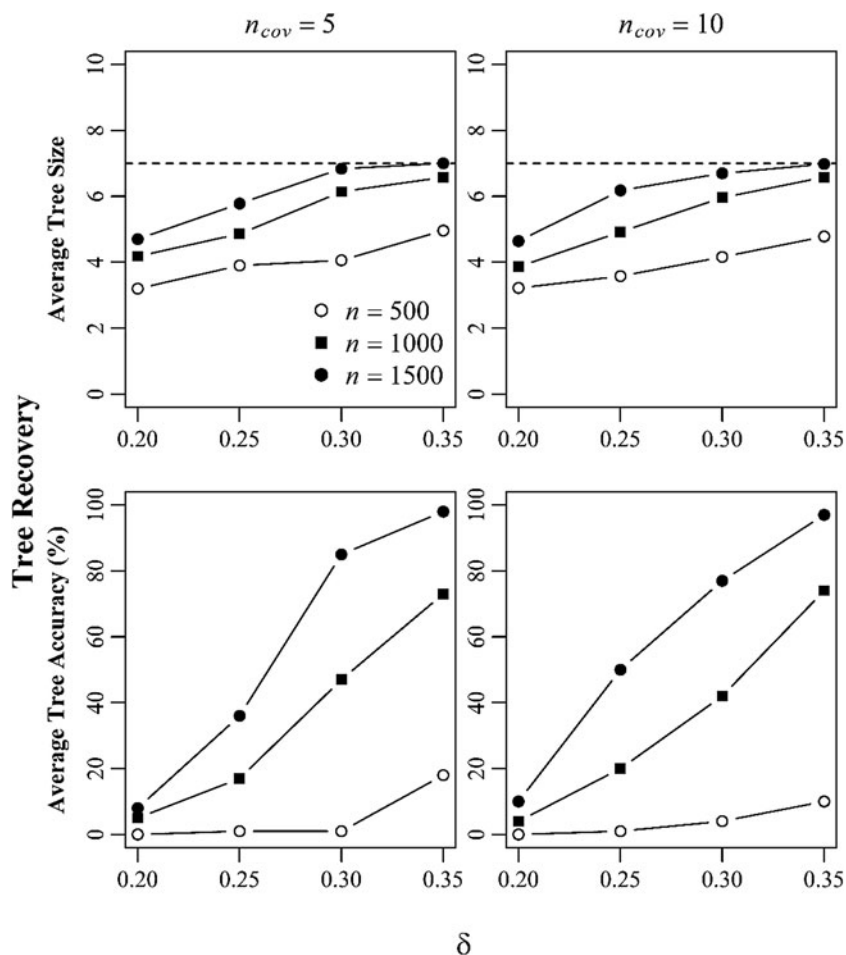
Figure 4. Average tree size and accuracy of recursively partitioned prediction configural frequency analysis (P-CFA) as a function of sample size (*n*), number of covariates (n_{cov}), and probability of CFA types (δ).

and probability of CFA types (δ). The upper panel of Figure 4 gives the average tree size per simulation condition, the lower panel summarizes tree accuracy (using the same definition as in Simulation Study 1). For smaller sample sizes and δ effects, the estimated trees tend to be smaller than the "true" tree structure. However, for large samples and δ ≥ 0.30, the average tree size is close to the "true" value. For *n* = 500, 1,000, and 1,500, we observe average tree sizes of 4.03 (*SD* = 1.35), 5.44 (*SD* = 1.53), and 6.08 (*SD* = 1.34) for n_{cov} = 5, and 3.94 (*SD* = 1.37), 5.33 (*SD* = 1.62), and 6.13 (*SD* = 1.35) for n_{cov} = 10. Similarly, large sample sizes (*n* = 1,500) and larger δ effects (δ ≥ 0.30) are required to obtain tree accuracy rates larger than 80% (see Figure 4 lower panel). Overall, we conclude that, similar to the first-order MOB CFA, MOB P-CFA is able to identify the underlying data-generating tree structure. However, again, larger samples are preferable when studying P-CFA tree structures.

Finally, we now focus on the percentage of correctly identified prediction types per subgroup. Again, we restrict our analysis to cases in which the MOB P-CFA algorithm correctly identified four subgroups. For Subgroup 1, we focus on the percentages to detect the prediction type {1 1 1}; for subgroup 4, we focus on the configuration {0 1 1}. Percentages to correctly identify {1 1 1} as a CFA type of Subgroup 1 ranged from 33.3% to 100% across all simulation conditions with an average of 89.5% (*SD* = 18.4%). Note that percentages larger than 80% were observed for 19 out of 24 simulation conditions. The remaining five simulation conditions describe data scenarios involving small sample sizes and

small δ effects. For example, the minimum of 33.3% occurred for *n* = 500, δ = 0.25, and n_{cov} = 5. Percentages of correctly identifying {0 1 1} as a CFA type in Subgroup 4 were lower than the corresponding percentages in Subgroup 1. This can be explained by the fact that the configuration {0 1 1} contradicts the underlying variable-oriented effect. Large percentages were only observed for *n* = 1,500 and δ = 0.35. Here, for n_{cov} = 5, {0 1 1} was correctly identified in 80.8% of the cases; for n_{cov} = 10, the type configuration was detected in 73.0% of the cases. Thus, overall, we can conclude that MOB P-CFA is able to detect subgroup-specific prediction types. Detection, however, is more accurate for type configurations that are in line with the global variable-oriented effect (such as {1 1 1} in the present context). When prediction type patterns contradict the global variable-oriented effect (such as {0 1 1}), large samples and large δ effects are required to detect these patterns with sufficient accuracy.

Empirical Example 2

In the following section, we illustrate the application of the MOB P-CFA approach again using the IY TCM data from Reinke et al. (2018). Now, we focus on the effectiveness of the IY TCM program to reduce teacher-rated internalization problems of students. Pre- and post-treatment internalization problems were measured using the TOCA-C (Koth et al., 2009). Scores were dichotomized at the 90th percentile. The binary internalization indicators were then analyzed using MOB P-CFA. Treatment status (*T*; 0 = business as usual;
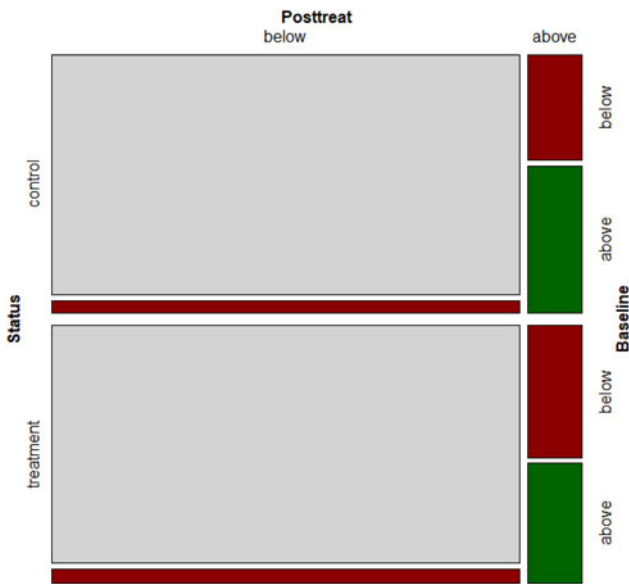
Figure 5. Mosaic plot of teacher-rated internalization problems for the total sample (*n* = 1,630 students); Status (left dimension) = control versus Incredible Years Teacher Classroom Management (IY TCM) treatment, posttreat (top dimension) = post-treatment internalization problems, baseline (right dimension) = pre-treatment internalization problems; red = configural frequency analysis (CFA) antitype, green = CFA type, gray = no deviation from expectation.

1 = IY TCM) and pre-treatment internalization status (I_1; 0 = <90th percentile; 1 = ≥90th percentile) were used as predictors and post-treatment internalization problems (I_2; 0 = <90th percentile; 1 = ≥90th percentile) was treated as outcome variable. The corresponding base model was saturated in the predictors and assumed independence between outcome and predictors. Formally, the model can be written as $\log(E[m]) = \lambda + \lambda^{I_1} + \lambda^T + \lambda^{I_2} + \lambda^{I_1 T}$. The covariates gender (0 = male, 1 = female), grade (K-3rd grade), race (1 = black, 0 = others), and school membership were used as splitting variables in the MOB algorithm.

First, we again focus on the P-CFA results of the total sample of 1,630 students given in Table 3 (the corresponding mosaic display is given in Figure 5). According to the LR goodness-of-fit test, the model does not fit the data well ($\chi^2(3) = 244.68$, $p < .001$) and we can conclude that prediction types/antitypes may exist. Again, we use the binomial test to evaluate whether cell frequencies significantly deviate from expectation. We used Bonferroni correction to protect the nominal significance level leading to an adjusted significance level of 0.05/8 = 0.00625. Overall, we observe two prediction types and four prediction antitypes. Interestingly, the two largest cells describing the absence of internalization problems in both treatment groups ({0 0 0} and {0 1 0}) do not deviate from expectation under the P-CFA model. Prediction types describe patterns of continuing internalization problems for both experimental groups. For the control group, that is for the cell {1 0 1}, we observed 49 (3.0%) students although only 8.6 (0.5%) had been expected. In the treatment condition, we observed 40 (2.5%) students with pre- and post-treatment internalization problems while only expecting 8.4 (0.5%) cases. The first two out of the four prediction antitypes, {1 0 0} and {1 1 0}, describe decreasing internalization problems in both groups. Here, only 34 cases (2.1%) were observed in the control group but 74.5 (4.6%) had been expected. Similarly, in the treatment group, 41 cases (2.5%) have been observed and

72.7 (4.5%) had been expected. Finally, the last two prediction antitypes describe an increase of internalization problems for both groups, {0 0 1} and {0 1 1}. Thirty-five control participants (2.2%) have been observed but 75.5 (4.6%) had been expected. In the treatment group, 44 students (2.7%) showed this pattern while 75.7 (4.6%) should have been observed according to the base P-CFA model.

Next, we ask whether the existing prediction types/antitypes hold on the population level or whether some prediction types/antitypes are specific to certain subpopulations. In addition, we evaluate whether the two patterns that did not significantly deviate from model expectation constitute prediction types/antitypes in a subpopulation. The MOB P-CFA algorithm can be used to address both questions. Figure 6 summarizes the MOB P-CFA tree for the splitting variables gender, grade, race, and school membership. Out of the four splitting variables two proved to be useful to partition the data, grade and school membership. Note that school membership was selected as a splitting variable for both grade partitions. Thus, the resulting regression tree consists of four terminal nodes. The first subgroup (*n* = 703, 43.1%) consists of K-, first- or second-grade students in Schools 1, 2, 4, 5, or 8. Students in the second subgroup (*n* = 605, 37.1%) are in K-, first, or second grade in Schools 3, 6, 7, or 9. The third subgroup is characterized as third-grade students in Schools 1, 4, or 6 (*n* = 95, 5.8%), and the last subgroup consists of third-grade students in Schools 2, 3, 5, 7, 8 or 9 (*n* = 227, 13.9%). For all four subgroups Bonferroni-corrected binomial tests suggest that types/antitypes exist.

In contrast to Empirical Example 1, we use mosaic displays to summarize end nodes (cf. Figure 6). In the mosaic plots, we use green areas to indicate P-CFA types, red areas to indicate P-CFA antitypes, and gray areas when no significant discrepancies exist. First, in comparison with the mosaic plot for the total sample (Figure 5), only Subgroup 2 shows similar dependency structures (interestingly, this is not the largest subgroup in the sample). Next, we observe different mosaic structures and, therefore, different dependencies between variables, across partitions, as expected. Overall, we observe the same type/antitype patterns as for the total sample. That is, using the order {baseline, status, post-treatment} to indicate baseline internalization problems (0 = below, 1 = above), status (0 = control, 1 = IY TCM treatment), and post-treatment internalization problems (0 = below, 1 = above), the configurations {1 0 1} and {1 1 1} constitute CFA types and {1 0 0}, {1 1 0}, {0 0 1}, and {0 1 1} are CFA antitypes. None of these patterns, however, hold on a population level. The type pattern {1 0 1} for example (i.e., high baseline scores – control group – high post-treatment scores), holds for the subpopulation of Subgroups 1, 2 and 4, and the CFA type {1 1 1} (high baseline scores – IY TCM treatment – high post-treatment scores) is only observed for Subgroups 2 and 3. Similarly, the antitype {1 0 0} (high baseline scores – control group – low post-treatment scores) holds in Subgroups 1 and 2, {1 1 0} is only observed in Subgroups 2 and 3, and {0 0 1} and {0 1 1} only exists in Subgroup 2.

Evaluating the Stability of CFA Trees

Tree structures found in the data can be unstable and highly sample dependent, that is, small changes in the input data can have tremendous effects on the resulting tree structure (Philipp, Rusch, Hornik, & Strobl, 2018). Thus, carefully evaluating the stability of a regression tree constitutes an important element of
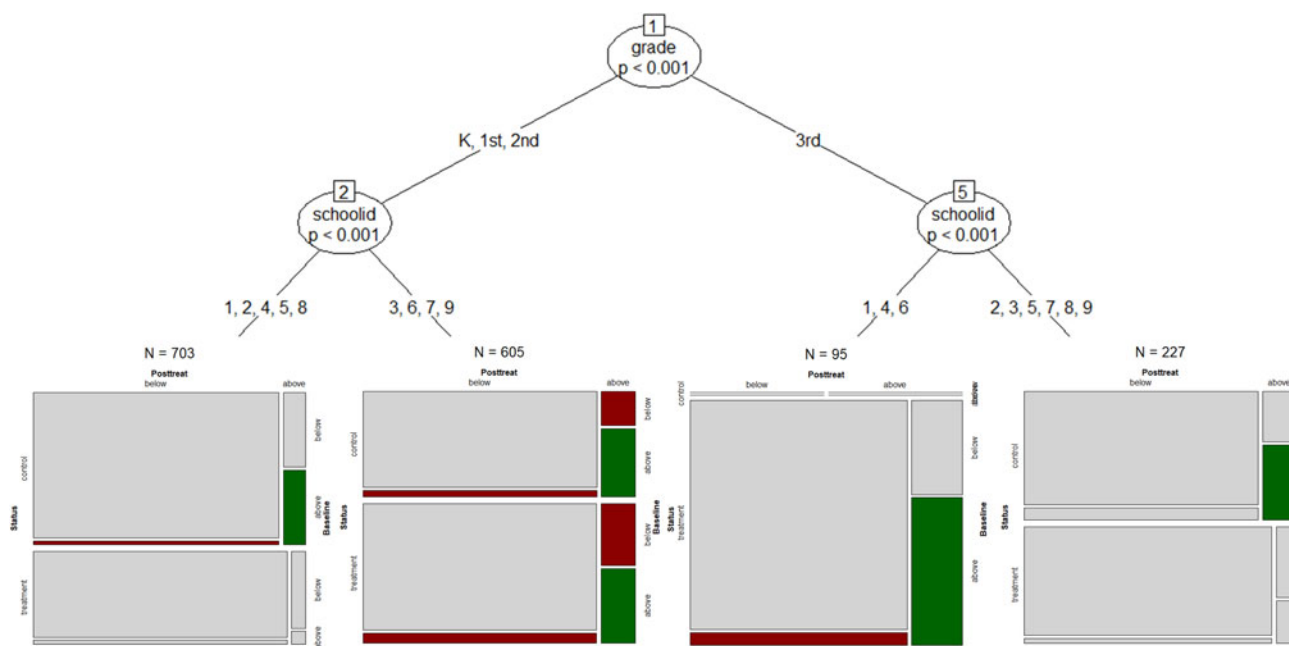
Figure 6. MOB P-CFA tree of teacher-rated internalization problems based on *n* = 1,630 students. Status (left dimension) = control versus Incredible Years Teacher Classroom Management (IY TCM) treatment, posttreat (top dimension) = post-treatment internalization problems, baseline (right dimension) = pre-treatment internalization problems; red = CFA antitype, green = CFA type, gray = no deviation from expectation.

every MOB application. Nonparametric bootstrapping has for example been suggested to quantify the robustness of the initial tree results (Philipp et al., 2018). Because different tree structures can lead to equivalent partitions and interpretation (Turney, 1995), instead of evaluating the structure of re-sampled trees, one can focus on two elements of a tree: (a) variable selection and (b) cut-point selection.

Let $b = 1, …, B$ denote the index of the B bootstrap samples (i.e., sampling with replacement from the original data) and $j = 1, …, J$ indicates the variables (Z_j) considered for partition. Further, let s_{bj} be an indicator that takes value 1 if Z_j is selected for partitioning and 0 otherwise. Then, the *variable selection percentage* is defined as $B^{-1} \sum_b s_{bj} \times 100$ and is expected to be close to 100% for those partitioning variables that had been selected in the original tree. To evaluate potential variability of cut-points,

two strategies are available depending on the measurement scale of the partitioning variable Z_j. For categorical splitting variables, *cut-point selection percentages* of the B re-samples can be used. In case of a continuous splitting variable, descriptive statistics such as means and standard deviations can be used to summarize the distribution of cut-points obtained from B re-samples (e.g., average cut-points of Z_j are expected to be close to the corresponding cut-point in the initial tree).

To illustrate MOB stability analysis in the context of configural frequency trees, we again use the data of the two empirical examples. For the first data example, we generated 100 re-samples of the five baseline student behavior indicators (concentration problem, disruptive behavior, emotional dysregulation, internalization problems, and family problems) and estimated a first-order MOB CFA for each re-sample. In line with the initial tree (cf. Figure 3), school membership was selected in 100% of the re-samples. Race was selecting in 57% and gender was selected in 38% of the re-samples. In addition, grade was identified as being a relevant splitting variable in 12% of the re-samples. In other words, with respect to the stability of the initial configural frequency tree, we observe mixed results. While school membership is identified as a stable component of the trees, race and gender show instabilities. Further, cut-point selection analysis of school membership suggests that the observed partition in Figure 3, [4, 5, 6] and [1, 2, 3, 7, 8, 9] was replicated in 68% of the re-samples. However, the most prevalent school membership partition (replicated in 82% of the re-samples) was one where School 6 is member of the other partition, that is, [4, 5][1, 2, 3, 6, 7, 8, 9]. This indicates that partitioning can be considered stable for Schools 4 and 5, but selection variability exists for School 6.

For the second data example, we generated 100 re-samples using pre- and post-treatment internalization and treatment status indicators. For each re-sample, we used MOB P-CFA to detect potential subgroups. The two splitting variables of the initial

Table 3. Prediction configural frequency analysis (P-CFA) results for the total sample (*n* = 1,630)

I_1	T	I_2	m	$E[m]$	*p* value	Decision
0	0	0	697	656.55	0.02208	–
1	0	0	34	74.45	0.00000	Antitype
0	1	0	690	658.35	0.05818	–
1	1	0	41	72.65	0.00003	Antitype
0	0	1	35	75.45	0.00000	Antitype
1	0	1	49	8.55	0.00000	Type
0	1	1	44	75.65	0.00004	Antitype
1	1	1	40	8.35	0.00000	Type

Note. I_1 = pre-treatment internalization problems, I_2 = post-treatment internalization problems, m = observed frequencies, $E[m]$ = expected frequencies. Bonferroni-corrected $\alpha = 0.00625$.

tree (school membership and grade) were selected in 100% of the re-samples indicating that, in contrast to the tree in Example 1, this configural frequency tree can be considered stable (note that, in addition, gender was selected as a splitting variable in 3% of the re-samples). Further, cut-point analysis suggests that in the majority of re-samples (68%) MOB P-CFA separated K-, first and second graders from third graders (which is in line with the initial tree), followed by partitions of the form [K, 3rd] [1st, 2nd] (18%) and [K, 1st, 3rd][2nd] (17%). Further, the two school membership partitions of the initial tree, that is, [1, 2, 4, 5, 8][3, 6, 7, 9] for K-, first, and second graders and [1, 4, 6][2, 3, 5, 7, 8, 9] for third graders was selected in the majority of re-samples (27% and 24% respectively), followed by the partitions [1, 2, 3, 4, 5, 7, 8][6, 9] (22%). The remaining partitions were less likely to occur (all ≤8%). Thus, overall, we can conclude that the initial configural frequency tree is sufficiently stable with respect to the selected cut-points.

Discussion

The present study introduced principles of configural frequency modeling and recursive partitioning and proposed a combination of the two methods to study moderation processes from a person-oriented (i.e., a pattern) perspective. CFA identifies the location of a categorical data space for which effects can be observed and MOB identifies subgroups for which model parameters significantly vary. Combining the two approaches leads to a powerful algorithm to evaluate moderation processes while preserving a person-oriented perspective. Configural frequency trees can be used to identify population-level and subpopulation-level types and antitypes. Application of the proposed configural frequency tree framework was illustrated focusing on first-order CFA and P-CFA trees. For both applications, Monte Carlo simulation results suggested that MOB CFA is able to recover true tree structures with adequate probability in particular for large sample sizes and large effect sizes. Further, two real-world empirical examples from school mental health research were presented for illustrative purposes. In both examples, MOB CFA identified four subgroups (partitions) that differ with respect to the likelihood of observing specific CFA type/antitype patterns.

The findings highlight the promise of CFA for identifying meaningful heterogeneity within populations and in response to interventions. The empirical example identified common patterns of youth behavior problems in relation to their expected occurrence. Youth with no behavior problems and those with co-occurring problems (concentration and family problems as well as emotional dysregulation and disruptive behaviors with and without internalization) were over-represented types. Those with family problems alone and internalization problems alone emerged as underrepresented antitypes. These findings are consistent with extensive research showing the common co-occurrence of youth behavior problems which arise rarely in isolation or as singular symptoms (Angold, Costello, & Erkanli, 1999; Herman, Ostrander, Walkup, Silva, & March, 2007). For instance, one study found that 80% of youth with major depressive disorder had clinically significant levels of other symptoms including concentration problems and disruptive behaviors (Herman et al., 2007). In addition, much prior research has highlighted family problems as a strong risk factor for youth mental health concerns (Dishion & Patterson, 2006; George, Herman, & Ostrander, 2006; Rothenberg, Hussong, & Chassin, 2016), so it is not surprising that an antitype of family problems without youth behavior

problems emerged from these analyses. CFA offers a parsimonious method for characterizing patterns of co-occurring symptoms, environmental contexts, and timing of risk conditions in line with a developmental psychopathology perspective (Cicchetti & Toth, 1995).

Subsequent analyses that included school, race, and gender as partitioning variables, suggested that particular schools were associated with higher risk for youth behavior problems, especially the concentration plus family problems type. School contexts influence student risk and how family risk (e.g., the percentage of families with economic challenges) intersects with school risk to influence negative youth outcomes (Biglan, Flay, Embry, & Sandler, 2012; Reinke & Herman, 2002). That the family-only antitype was most common in the other six schools is consistent with the importance of school context in the likelihood of youth expressing emotional and behavior symptoms. These schools may have provided a protective buffer for youth in them that reduced their risk of mental health concerns, at least as observed and reported by their teachers, in the presence of family problems. It was also noteworthy that the "no problems" pattern emerged as an overrepresented population type, again suggesting that the collective school environment in this district may have provided a protective buffer that mitigated expected levels of youth mental health concerns.

The CFA of treatment effects highlights the promise of the method in uncovering contextual factors associated with youth outcomes in response to intervention that does not have the downside or complexities of using multiple moderation analyses. In this example, both treatment conditions experienced comparable trends. MOB findings highlight the promise of CFA in helping understanding the generalizability and replication of RCT findings. Recent research has suggested that treatment effects may be contingent on unspecified and unexamined contextual factors that are inherent to any study. For instance, Kaplan et al. (2020) conducted a multisite study that involved 10 RCTs of the same design and found widely discrepant effect sizes based on the context of each study. These authors argued that causal mechanisms need to be contextualized and that future studies need to include careful elaboration and evaluation of potential contextual moderators of study effects. Recently, Herman, Dong, Reinke, and Bradshaw (2020) also found the findings of an educational RCT varied before and after a prominent historical event based on both participant race and treatment condition. Both Kaplan et al. (2020) and Herman et al. (2020) used moderation analyses to identify these contextual variations of study effects. MOB CFA offers a more sophisticated and powerful tool to detect and summarize multiple contextual processes simultaneously.

The presented MOB CFA framework can be extended in various ways. In essence, any CFA model that can be cast as a log-linear model can serve as a candidate for MOB CFA. Here, we want to focus on two potential extensions that seem particularly relevant for developmental psychopathology: k-sample CFA and longitudinal CFA. k-sample CFA has been suggested to test whether configurations of categorical variables (so-called discrimination variables) differ across the categories of a grouping variable. In this approach, the grouping variable can be considered a moderator. MOB CFA could be used to investigate the effects that additional continuous or categorical moderators have, and to determine whether population-based group differences hold when additional moderators are the basis of finer-grained classifications. Similarly, longitudinal CFA can be used to determine whether patterns of development are predictive of endpoints of

development. Here, MOB CFA can be used to find finer-grained endpoint patterns as well as finer-grained developmental trajectories. In other words, MOB CFA can be used to determine whether a population is homogeneous in type/antitype patterns of development.

Another extension concerns sensitivity analyses associated with MOB CFA applications. Previous studies have emphasized that single decision trees can provide low predictive performance and that ensemble techniques such as bagging (Breiman, 1996), boosting (Freund & Schapire, 1995), and random forests (Breiman, 2001) can overcome some of these issues (Fokkema & Strobl, 2020; Strobl et al., 2009). We expect that making use of ensemble methods in the context of MOB CFA can also lead to model improvements.

Although CFA is usually presented for "true" categorical variables, that is, variables that measure constructs in the form of (un)ordered distinct categories (taxa), configural frequency modeling can also be useful in the context of categorized continuous variables (e.g., a dichotomized composite measure) leading to a broad area of potential applications in developmental psychopathology research. In the presented real-world data examples, for instance, composite measures of student behavior were dichotomized at the 90th percentile to indicate problem behavior. It is important to note that data categorization is not free of limitations and has been hotly debated in quantitative methods research. Some authors argue that dichotomization should be avoided due to the tremendous loss of information affecting the validity of subsequent statistical models (Cohen, 1983; DeCoster, Iselin, & Gallucci, 2009; MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993; Rucker, McShane, & Preacher, 2015), others argue that the cost of dichotomization is comparatively low given potential benefits in clarity and communicability of results (Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015; von Eye & Mair, 2012; Westfall, 2011; Wiedermann & von Eye, 2020a). Furthermore, and most important for the present context, all previous studies have discussed benefits and costs of dichotomization from a purely variable-oriented perspective. Less is known about the effects of dichotomization when research hypotheses are tested in terms of patterns, that is, in the person-oriented research context. Although further studies are needed to evaluate the gains and losses of data categorization in person-oriented analyses, our results of Simulation Study 2 may provide some preliminary insights. In the second simulation experiment, we simulated a continuous covariate that was subsequently dichotomized at the mean. Simulation results indicate that MOB identifies underlying partitions with high accuracy even when a dichotomized variable is included in the model. Thus, in the present context, adverse effects of categorization can be expected to be low in configural frequency modeling.

MOB CFA does not come without some limitations. Both simulation studies suggested that large sample sizes are preferable to detect tree structures. The requirement of sufficiently large sample sizes will often be fulfilled in cross-sectional research that adopts a person-oriented/pattern perspective. In precision public health and precision prevention science (Khoury, Iademarco, & Riley, 2016; Supplee, Parekh, & Johnson, 2018), for example, it is well-recognized that large samples are needed to take individual variability into account and to make valid statements about subgroups or clients. That is, since collected information of people is more informative when diverse people of the underlying population are included and large enough subgroup-specific sample sizes are necessary to guarantee an accurate representation of the corresponding subpopulation, the overall sample sizes can be expected to be sufficiently large for applying MOB CFA. However, often, person-oriented studies focus on intensive longitudinal individual data (e.g., daily diary data) with moderate samples sizes. Configural frequency modeling of intensive longitudinal data has been discussed by von Eye et al. (2010). Here, CFA models are available to evaluate *runs* (asking whether adjacent scores are equal or different) and *lags* of a series (e.g., relating information from one time point to information from the next). Information that describes runs and lags can then be categorized and crossed with other categorial indicators (e.g., demographic variables). The benefit of ideographic data information usually comes at the cost of small to moderate sample sizes. Although, our simulation studies clearly indicated that, in a cross-sectional data setting, large samples are preferable, the ability of MOB CFA to detect heterogenous subgroups will also depend on the complexity of the tree and the magnitude of the effects. For example, extreme groups may also be detectable in the form of a two-group MOB solution even when sample sizes are moderate – apparently, ultimate statements about the performance of MOB CFA under such extreme conditions require further studies.

The requirement of larger sample sizes also puts constraints on the number of variables that can be incorporated to define configurations of interest. For example, with 1,500 subjects and 10 binary variables the case-wise data for MOB has $2^{10} \times 1,500 = 1,536,000$ rows. Thus, if the number of indicators is large, researchers should consider the application of a data reduction method prior to MOB CFA. For example, latent class analysis techniques (Collins & Lanza, 2010; Lazarsfeld & Henry, 1968) can be used to summarize potentially complex profiles of subgroups of indicators in the form of a small number of latent classes. Latent class membership information across subgroups of indicators can then be used for recursive partitioning (for a discussion of latent class analysis in the context of CFA modeling see Wiedermann & von Eye, 2016).

Overall, MOB CFA can be a powerful tool to study moderation processes from a pattern perspective provided that sample sizes are sufficiently large, and one takes effort in critically evaluating the robustness of the findings. To make MOB CFA accessible to researchers, we provide an implementation of the discussed models in R in the online supplement to this article.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley & Sons.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*, 821. doi:10.2307/2951764

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*, 57–87.

Bergman, L. R. (2001). A person approach in research on adolescence: Some methodological challenges. *Journal of Adolescent Research*, *16*, 28–53. doi:10.1177/0743558401161004

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, *9*, 291–319. doi:10.1017/S095457949700206X

Bergman, L. R., Magnusson, D., & El Khouri, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Biglan, A., Flay, B. R., Embry, D. D., & Sandler, I. N. (2012). The critical role of nurturing environments for promoting human well-being. *American Psychologist*, 67, 257–271. doi:10.1037/a0026796

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140. doi:10.1023/A:1018054314350

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. L. (1984). *Classification and regression trees.* Belmont: Wadsworth.

Cairns, R. B. (1979). *Social development: The origins and plasticity of interchanges.* San Francisco: Freeman.

Cairns, R. B., Elder, G. H., & Costello, E. J. (Eds.). (1996). *Developmental science* (1st ed.). Cambridge University Press. doi:10.1017/CBO9780511571114

Christensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). New York: Springer.

Cicchetti, D., & Schneider-Rosen, K. (1986). An organization approach to childhood depression. In M. Rutter, C. Izard, & P. Read (Eds.), *Depression in young people, clinical and developmental perspectives* (pp. 71–134). New York: Guilford Press.

Cicchetti, D., & Toth, S. L. (1995). Developmental psychopathology and disorders of affect: Risk, disorder, and adaptation. In D. Cicchetti, & D. J. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (Vol. 2, pp. 369–420). New York: Wiley & Sons.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–253. doi:10.1177/014662168300700301

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis for the social, behavioral, and health sciences.* New York: Wiley and Sons.

Daukantaitė, D., Lundh, L.-G., & Wångby-Lundh, M. (2019). Association of direct and indirect aggression and victimization with self-harm in young adolescents: A person-oriented approach. *Development and Psychopathology*, 31, 727–739. doi:10.1017/S0954579418000433

DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, 14, 349–366. doi:10.1037/a0016956

Dishion, T. J., & Patterson, G. R. (2006). The development and ecology of antisocial behavior in children and adolescents. In D. Cicchetti, & D. J. Cohen (Eds.), *Developmental psychopathology (vol. 3): Risk, disorder, and adaption* (2nd ed., pp. 503–541). Hoboken: Wiley & Sons.

Dunnett, C. W., & Tamhane, A. C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, 87, 162–170. doi:10.1080/01621459.1992.10475188

Dusseldorp, E., & Van Mechelen, I. (2014). Qualitative interaction trees: A tool to identify qualitative treatment-subgroup interactions. *Statistics in Medicine*, 33, 219–237. doi:10.1002/sim.5933

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50, 2016–2034. doi:10.3758/s13428-017-0971-x

Fokkema, M., & Strobl, C. (2020). Fitting prediction rule ensembles to psychological research data: An introduction and tutorial. *Psychological Methods*, doi:10.1037/met0000256

Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory.* Newbury Park: Sage.

Freund, Y., & Schapire, R. E. (1995). *A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting.* European Conference on Computational Learning Theory.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89, 190–200. doi:10.1080/01621459.1994.10476460

Friendly, M. (1995). Conceptual and visual models for categorical data. *The American Statistician*, 49, 153. doi:10.2307/2684630

George, C., Herman, K. C., & Ostrander, R. (2006). The family environment and developmental psychopathology: The unique and interactive effects of depression, attention, and conduct problems. *Child Psychiatry and Human Development*, 37, 163–177. doi:10.1007/s10578-006-0026-5

Greenberg, M. T., Speltz, M. L., Deklyen, M., & Jones, K. (2001). Correlates of clinic referral for early conduct problems: Variable- and person-oriented approaches. *Development and Psychopathology*, 13, 255–276. doi:10.1017/S0954579401002048

Hartigan, J. A., & Kleiner, B. (1984). A mosaic of television ratings. *The American Statistician*, 38, 32. doi:10.2307/2683556

Hatzinger, R., & Dittrich, R. (2012). Prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software*, 48. doi:10.18637/jss.v048.i10

Herman, K. C., Dong, N., Reinke, W. M., & Bradshaw, C. P. (2020). Accounting for traumatic historical events in social, behavioral, and educational randomized control trials. *(Under Review).*

Herman, K. C., Ostrander, R., Walkup, J. T., Silva, S. G., & March, J. S. (2007). Empirically derived subtypes of adolescent depression: Latent profile analysis of co-occurring symptoms in the Treatment for Adolescents with Depression Study (TADS). *Journal of Consulting and Clinical Psychology*, 75, 716–728. doi:10.1037/0022-006X.75.5.716

Hildebrand Karlén, M., Nilsson, T., Wallinius, M., Billstedt, E., & Hofvander, B. (2020). A bad start: The combined effects of early onset substance use and ADHD and CD on criminality patterns, substance abuse and psychiatric comorbidity among young violent offenders. *Journal for Person-Oriented Research*, 6, 39–55. doi:10.17505/jpor.2020.22045

Hjort, N. L., & Koning, A. (2002). Tests for constancy of model parameters over time. *Journal of Nonparametric Statistics*, 14, 113–132. doi:10.1080/10485250211394

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75, 800–802. doi:10.1093/biomet/75.4.800

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics*, 43, 417. doi:10.2307/2531823

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651–674. doi:10.1198/106186006X133933

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.

Huang, F. L. (2019). Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education*, 1–16. doi:10.1080/00220973.2019.1699769

Hunt, P. R., Friesen, M. C., Sama, S., Ryan, L., & Milton, D. (2015). Log-linear modeling of agreement among expert exposure assessors. *Annals of Occupational Hygiene*, 59, 764–774. doi:10.1093/annhyg/mev011

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25, 652–665. doi:10.1016/j.jcps.2014.12.002

Kagan, J. (1994). *Galen's prophecy: Temperament in human nature.* New York: Basic Books.

Kaplan, A., Cromley, J., Perez, T., Dai, T., Mara, K., & Balsai, M. (2020). The role of context in educational RCT findings: A call to redefine "evidence-based practice.". *Educational Researcher*, 49, 285–288. doi:10.3102/0013189X20921862

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119. doi:10.2307/2986296

Keller, T. E., Spieker, S. J., & Gilchrist, L. (2005). Patterns of risk and trajectories of preschool problem behaviors: A person-oriented analysis of attachment in context. *Development and Psychopathology*, 17. doi:10.1017/S0954579405050170

Khoury, M. J., Iademarco, M. F., & Riley, W. T. (2016). Precision public health for the era of precision medicine. *American Journal of Preventive Medicine*, 50, 398–401. doi:10.1016/j.amepre.2015.08.031

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation—checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*, 42, 15–30. doi:10.1177/0748175609333560

Lautsch, E., & Ninke, L. (2000). Kombinierter Einsatz von CHAID und KFA bei der soziodemographischen Beschreibung von Kriminalitätsfurcht. *Psychologische Beiträge*, 42, 347–361.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Lienert, G. A. (1968). *Die Konfigurationsfrequenzanalyse als Klassifikationsmethode in der Klinischen Psychologie. [Configural Frequency Analysis as Classification Method in Clinical Psychology.]. Paper Presented at the 26. Kongress der Deutschen Gesellschaft für Psychologie in Tübingen.*

Lienert, G. A., & Krauth, J. (1973). Die Konfigurationsfrequenzanalyse als Prädiktionsmodell in der angewandten Psychologie. In H. Eckensberger (Ed.), *Bericht über den 28. Kongreß der deutschen gesellschaft für psychologie in saarbrücken 1972* (pp. 219–228). Göttingen: Hogrefe.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40. doi:10.1037//1082-989X.7.1.19

Magnusson, D. (1985). Implications of an interactional paradigm for research on human development. *International Journal of Behavioral Development*, 8, 115–137. doi:10.1177/016502548500800201

Magnusson, D. (2001). The holistic-interactionistic paradigm: Some directions for empirical developmental research. *European Psychologist*, 6, 153–162. doi:10.1027//1016-9040.6.3.153

Magnusson, D., & Allen, V. L. (1983). Implications and applications of an interactional perspective for human development. In D. Magnusson, & V. L. Allen (Eds.), *Human development: An interactional perspective* (pp. 369–387). New York: Academic Press.

Magnusson, D., & Stattin, H. (1998). Person-context interaction theories. In W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 685–759). New York: Wiley & Sons.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113, 181–190. doi:10.1037/0033-2909.113.1.181

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Melcher, A. H., Lautsch, E., & Schmutz, S. (2012). Non-parametric methods – tree and P-CFA – for ecological evaluation and assessment of suitable aquatic habitats: A contribution to fish psychology. *Psychological Test and Assessment Modeling*, 54, 293–306.

Molenaar, P. C. M. (2010). Testing all six person-oriented principles in dynamic factor analysis. *Development and Psychopathology*, 22, 255–259. doi:10.1017/S0954579410000027

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18, 112–117. doi:10.1111/j.1467-8721.2009.01619.x

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58, 415–434. doi:10.1080/01621459.1963.10500855

Müller, M. J., Netter, P., & von Eye, A. (1997). Catecholamine response curves of male hypertensives identified by Lehmacher's two sample configural frequency analysis. *Biometrical Journal*, 39, 29–38. doi:10.1002/bimj.4710390104

Mun, E. Y., Bates, M. E., & Vaschillo, E. (2010). Closing the gap between person-oriented theory and methods. *Development and Psychopathology*, 22, 261–271. doi:10.1017/S0954579410000039

Mun, E. Y., von Eye, A., Fitzgerald, H. A., & Zucker, R. A. (2001). Using mosaic displays in configural frequency analysis. *Methods of Psychological Research-Online*, 6, 3.

Olejnik, S., Li, J., Supattathum, S., & Huberty, C. J. (1997). Multiple testing and statistical power with modified Bonferroni procedures. *Journal of Educational and Behavioral Statistics*, 22, 389. doi:10.2307/1165229

Philipp, M., Rusch, T., Hornik, K., & Strobl, C. (2018). Measuring the stability of results from supervised statistical learning. *Journal of Computational and Graphical Statistics*, 27, 685–700. doi:10.1080/10618600.2018.1473779

Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* San Mateo: Morgan Kaufmann.

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Reinke, W. M., & Herman, K. C. (2002). Creating school environments that deter antisocial behaviors in youth. *Psychology in the Schools*, 39, 549–559. doi:10.1002/pits.10048

Reinke, W. M., Herman, K. C., & Dong, N. (2018). The incredible years teacher classroom management program: Outcomes from a group randomized trial. *Prevention Science*, 19, 1043–1054. doi:10.1007/s11121-018-0932-3

Rose, T. (2015). *The end of average: How we succeed in a world that values sameness.* San Francisco: HarperOne/HarperCollins.

Rothenberg, W. A., Hussong, A. M., & Chassin, L. (2016). Intergenerational continuity in high-conflict family environments. *Development and Psychopathology*, 28, 293–308. doi:10.1017/S0954579415000450

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*, 25, 666–678. doi:10.1016/j.jcps.2015.04.004

Rusch, T., & Zeileis, A. (2013). Gaining insight with recursive partitioning of generalized linear models. *Journal of Statistical Computation and Simulation*, 83, 1301–1315. doi:10.1080/00949655.2012.658804

Schrepp, M. (2006). The use of configural frequency analysis for explorative data analysis. *British Journal of Mathematical and Statistical Psychology*, 59, 59–73. doi:10.1348/000711005X66761

Stattin, H., & Magnusson, D. (1996). Antisocial development: A holistic approach. *Development and Psychopathology*, 8, 617–645. doi:10.1017/S0954579400007331

Stemmler, M., & Heine, J.-H. (2017). Using configural frequency analysis as a person-centered analytic approach with categorical data. *International Journal of Behavioral Development*, 41, 632–646. doi:10.1177/0165025416647524

Stemmler, M., Heine, J. H., & Wallner, S. (2019). Analyzing tree structures with configural frequency analysis and the R-package confreq. *Psychological Test and Assessment Modeling*, 61, 419–433.

Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, 22, 239–254. doi:10.1017/S0954579410000015

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14, 323–348. doi:10.1037/a0016973

Supplee, L. H., Parekh, J., & Johnson, M. (2018). Principles of precision prevention science for improving recruitment and retention of participants. *Prevention Science*, 19, 689–694. doi:10.1007/s11121-018-0884-7

Turney, P. (1995). Technical note: Bias and the quantification of stability. *Machine Learning*, 20, 23–33. doi:10.1007/BF00993473

van Wie, M., Li, X., & Wiedermann, W. (2019). Identification of confounded subgroups using linear model based recursive partitioning. *Psychological Test and Assessment Modeling*, 61, 365–387.

von Eye, A. (1990). *Introduction to configural frequency analysis.* Cambridge: Cambridge University Press.

von Eye, A. (2002). *Configural frequency analysis: Methods, models, and applications.* Mahwah: Lawrence Erlbaum Associates.

von Eye, A. (2004). Base models for configural frequency analysis. *Psychology Science*, 46, 150–170.

von Eye, A. (2010). Developing the person-oriented approach: Theory and methods of analysis. *Development and Psychopathology*, 22, 277–285. doi:10.1017/S0954579410000052

von Eye, A., & Bergman, L. R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology*, 15, 553–580. doi:10.1017/S0954579403000294

von Eye, A., Bergman, L. R., & Hsieh, C. H. (2015). Person oriented methodological approaches. In W. F. Overton, & P. C. M. Molenaar (Eds.), *Handbook of child psychology and developmental science: Theory and methods* (pp. 789–841). Hoboken: Wiley & Sons.

von Eye, A., & Bogat, G. A. (2005). Logistic regression and prediction configural frequency analysis – a comparison. *Psychology Science*, 47, 326–341.

von Eye, A., & Gutiérrez Peña, E. (2004). Configural frequency analysis: The search for extreme cells. *Journal of Applied Statistics*, 31, 981–997. doi:10.1080/0266476042000270545

von Eye, A., & Mair, P. (2008). A functional approach to configural frequency analysis. *Austrian Journal of Statistics*, 37, 161–173.

von Eye, A., & Mair, P. (2012). On the effects of dichotomizing information. In A. Barragan, F. M. Martínez, L. E. N. Barajas, & C. C. Covarrubias (Eds.), *Memorias del XXIV Foro National de Estadistica* (pp. 11–19). Aguascalientes: Instituto Nacional de Estadistica y Geografia.

von Eye, A., Mair, P., & Bogat, G. A. (2005). Prediction models for configural frequency analysis. *Psychology Science*, 47, 342–355.

von Eye, A., Mair, P., & Mun, E. Y. (2010). *Advances in configural frequency analysis*. New York: Guilford Press.

von Eye, A., Mun, E. Y., & Bogat, G. A. (2008). Temporal patterns of variable relationships in person-oriented research: Longitudinal models of configural frequency analysis. *Developmental Psychology*, 44, 437–445. doi:10.1037/0012-1649.44.2.437

von Eye, A., Mun, E. Y., & Mair, P. (2009). What carries a mediation process? Configural analysis of mediation. *Integrative Psychological and Behavioral Science*, 43, 228–247. doi:10.1007/s12124-009-9088-9

von Eye, A., & Schuster, C. (1998). On the specification of models for configural frequency analysis—sampling schemes in prediction CFA. *Methods of Psychological Research Online*, 3.

von Eye, A., Schuster, C., & Rogers, W. M. (1998). Modelling synergy using manifest categorical variables. *International Journal of Behavioral Development*, 22, 537–557. doi:10.1080/016502598384261

von Eye, A., & Wiedermann, W. (2021). *CFA. Configural frequency analysis*. New York: Springer (in prep.).

von Eye, A., Wiedermann, W., & von Weber, S. (2019). Log-linear and configural analysis of tree structures. *Psychological Test and Assessment Modeling*, 61, 435–451.

von Weber, S., Lautsch, E., & von Eye, A. (2003). On the limits of configural frequency analysis: Analyzing small tables. *Psychology Science*, 45, 339–354.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child & Adolescent Psychology*, 33, 105–124. doi:10.1207/S15374424JCCP3301_11

Westfall, P. H. (2011). Improving power by dichotomizing (even under normality). *Statistics in Biopharmaceutical Research*, 3, 353–362. doi:10.1198/sbr.2010.09055

Wiedermann, W., Bergman, L. R., & von Eye, A. (2016). Developments in methods for person-oriented research. *Journal for Person-Oriented Research*, 2, 1–4. doi:10.17505/jpor.2016.01

Wiedermann, W., Reinke, W., & Herman, K. (2020). Prosocial skills causally mediate the relation between effective classroom management and academic competence: An application of direction dependence analysis. *Developmental Psychology*, 56, 1723–1735.

Wiedermann, W., & von Eye, A. (2016). Local associations in latent class analysis: Using configural frequency analysis for model evaluation. *Journal for Person-Oriented Research*, 2. doi:10.17505/jpor.2016.15

Wiedermann, W., & von Eye, A. (2020a). Log-linear models to evaluate direction of effect in binary variables. *Statistical Papers*, 61, 317–346. doi:10.1007/s00362-017-0936-2

Wiedermann, W., & von Eye, A. (2020b). Reciprocal relations in categorical variables. *Psychological Methods*, 25, 708–725. doi:10.1037/met0000257

Zeileis, A., & Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability. *Statistica Neerlandica*, 61, 488–508. doi:10.1111/j.1467-9574.2007.00371.x

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17, 492–514. doi:10.1198/106186008X319331

Zhang, H., & Singer, B. (2010). *Recursive partitioning and applications* (2nd ed.). New York: Springer.