Prevention Science
https://doi.org/10.1007/s11121-021-01241-8

Local Effects of Intervention: a Configural Analysis

Alexander von Eye[1] · Wolfgang Wiedermann[2] · Keith C. Herman[2] · Wendy Reinke[2]

Accepted: 1 April 2021

Abstract

In standard statistical data analysis, the effects of intervention or prevention efforts are evaluated in terms of variable relations. Results from application of regression-type methods suggest whether, overall, intervention is successful. In this article, we propose using configural frequency analysis (CFA) either in tandem with regression-type methods or by itself. CFA allows one to adopt a person-oriented perspective in which individuals are targeted that can be characterized by particular profiles. The questions asked in CFA concern these individuals instead of variables. In prevention research, one can ask whether, for particular profiles, the preventive measures are successful. In three real-world data examples, CFA is applied and compared to standard log-linear modeling. Examples consider non-randomized (observational) and randomized intervention settings. The results of these analyses suggest that person-oriented CFA and standard variable-oriented methods of analysis respond to different questions. We show that integrating person- and variable-oriented perspectives can help researchers obtain a fuller picture of intervention effectiveness. Extensions of the CFA approach are discussed.

Keywords Person-oriented research · Intervention effectiveness · Local effect · Configural frequency analysis · Log-linear model

Variable-centered analyses are the dominant methods used in social science and medical research (Bergman & Magnusson, 1997; Howard & Hoffman, 2018). Variable-centered approaches assume that an average set of parameters can best characterize data, whereas person-centered methods assume that subgroups exist within the data that have their own unique sets of parameters (Bergman, 2001; Bergman & Magnusson, 1997). In the most extreme case, every individual in the study has a unique set of parameters (a perspective usually taken in ideographic research; Molenaar, 2004). Although prevention scientists have long championed the importance of person-centered analyses, prevention outcome studies that report variable-centered or main effect analyses may unwittingly underestimate or mischaracterize an intervention's true benefit (Greenberg & Abenavoli, 2017). Instead, examining the natural heterogeneity of response to universal prevention strategies can help identify subgroups most responsive to intervention and may better capture the value of preventive interventions (Greenberg & Abenavoli, 2017).

It has repeatedly been shown (e.g., Mun et al., 2010; Thompson et al., 2019; von Eye & Bergman, 2003) that person- and variable-oriented approaches complement each other. For example, based on aggregate-level variable-oriented analysis, von Eye & Bergman (2003) showed that the beer consumption in self-diagnosed alcoholics follows a clear weekday-related autocorrelation pattern. However, there exist individuals who systematically deviate from this pattern. Without person-oriented, detailed analysis, this fact would escape the data analyst. Conversely, Bogat et al. (2016) showed that compared to men, women respond quite differently to intimate partner violence. Without variable-oriented analysis, this statement, while important, would miss that there are, nevertheless, widely visible general effects of such violence.

In the present article, we introduce a person-oriented method for categorical data to the audience of prevention scientists—*configural frequency analysis* (CFA; Lienert, 1968; von Eye & Gutiérrez Peña, 2004; von Eye et al., 2010). The remainder of this article is structured as follows. First, we provide an introduction to CFA as a statistical method that allows one to test hypotheses under an event-based

✉ Wolfgang Wiedermann
wiedermannw@missouri.edu

[1] Michigan State University, East Lansing, MI, USA

[2] Missouri Prevention Science Institute and Department of Educational, School, and Counseling Psychology, University of Missouri, 16 Hill Hall, Columbia, MO 65211, USA

perspective. This implies, in the analysis of categorical variables, that hypotheses are tested at the level of individual cells or groups of cells in a cross-classification. CFA is, thus, a method of analysis for person-oriented (Bergman & Magnusson, 1997; Wiedermann et al., 2016) and idiographic research (Molenaar, 2004). In contrast, standard variable-oriented data analysis specifies and tests hypotheses at the level of variable relations. The person-oriented approach starts from what is known as the *ecological fallacy* (see von Eye et al., 2015). This fallacy occurs when aggregation of data results in misleading conclusions about general trends. Data aggregation can make researchers overlook that groups of persons or individuals can systematically differ from the overall trend that can be found in data. If such individuals exist, they show development or treatment effects that are reliable but different from conclusions that are based on data aggregation. Therefore, theoretical concepts and methods of data analysis have been developed that are suitable for research from a person-oriented perspective. The present article can be seen in the context of this line of research. CFA is considered among the most important and useful methods of analysis in person-oriented research (von Eye et al., 2015).

Second, we present results from three re-analyses of data from a large-scale randomized controlled trial (RCT) to assess the impact of the Incredible Years Teacher Classroom Management (IY TCM) program on student social-emotional and academic outcomes (Reinke et al., 2018). In the first of these analyses, we perform an exploratory CFA in which we ask whether there exist patterns (*configurations*) that occur at unexpected rates. In the second, we use configural moderation analysis to ask whether frequencies of configurations of interest are moderated by demographic characteristics. In the third, we perform a confirmatory CFA in which we focus on particular cells, those that are a priori of interest in prevention research.

Configural Frequency Analysis

According to Goodman (1984), a cross-classification of categorical variables can be analyzed with the following three aims in mind: (1) examining the joint distribution of the variables that span a cross-classification, (2) examining the association structure of these variables, and (3) examining the dependency structure of these variables. von Eye and Mun (2016) add a fourth aim: (4) determining whether individual cells or groups of cells deviate from the expectancy that is specified using a probability model. CFA, originally proposed by Lienert (1968; cf. Lienert & Krauth, 1975; von Eye, 2002; von Eye & Gutiérrez Peña, 2004; von Eye et al., 2010), allows the researcher to test hypotheses that concern the frequency with which an individual cell or groups of cells in a cross-classification of categorical variables were observed. To test such hypotheses, expectancies are estimated in the context of a *base model* (von Eye, 2004). Statistically, these models often are log-linear models of the form $\log \hat{m} = X\lambda$, where $\hat{m}$ is the array of expected model frequencies, X is the design matrix that represents the base model, and λ is the parameter vector (more detail on base models follows in subsequent paragraphs). Parameters can be estimated within the generalized linear modeling (GLM) framework using a log link and a Poisson distribution.

Testing Hypotheses in CFA

Let $\hat{m}_{i\epsilon I}$ be the expected frequency of the ith cell (configuration) and I the total number of cells of the cross-classification under study. The corresponding observed cell frequency is $m_{i\epsilon I}$. Then, the CFA null hypothesis for Cell i is that the observed cell frequency equals the model cell frequency, or $H_0 : m_{i\epsilon I} = \hat{m}_{i\epsilon I}$. In exploratory CFA, this null hypothesis is tested for each of the I cells. In confirmatory CFA, this null hypothesis is tested for each of an a priori selected group of cells.

A large number of tests has been proposed for this null hypothesis (for an overview, see von Eye, 2002). Here, we present the binomial test as an example of an exact test and the z-test as an example of an approximative test. Other tests (not described here) include Pearson's χ^2-test and exact and approximative hypergeometric tests. The scores of the χ^2-test can be used as summands for a statistic that provides an overall goodness-of-fit test of the base model.

The binomial test can be used to estimate the probability B_i of the observed frequency m_i $(i = 1, \ldots, I)$ given the probability p obtained from a base model. For the observed cell frequency m_i, the exact tail probability of observing m_i and more extreme frequencies is $B_i(m_i) = \sum_{j=a}^{l} \binom{N}{j} p^j q^{N-j}$, where N is the sample size, $q = 1 - p$, $a = 0$ if $m_i < \hat{m}_i$ and $a = m_i$ if $m_i \geq \hat{m}_i$, $l = m_i$ if $a = 0$, and $l = N$ if $a = m_i$. When, as is the usual case, p is not known a priori, it is estimated from the sample. It is a priori known only in rare cases. The binomial test is exact. The probabilities of the observed and each more extreme frequency are calculated and added to the total. There is no need to assume that a theoretical distribution is reasonably approximated. The binomial test is recommended in particular when a sample is so small that the approximation properties of approximative tests are doubtful. The procedure is, however, known to be conservative (i.e., statistically less powerful than alternative significance procedures).

A sample approximative test is the z-test, $z = \left(m_i - Np\right)/\sqrt{Npq}$ where z follows the standard normal distribution under the null hypothesis. One rejects the null hypothesis $H_0 : m_{i\epsilon I} = \hat{m}_{i\epsilon I}$ against the alternative hypotheses (1) $H_A : m_{i\epsilon I} \neq \hat{m}_{i\epsilon I}$ when $|z| > N(0, 1)_{1-\alpha/2}$ (with $N(0,$

$1)_{1-\alpha/2}$ being the $1 - \alpha/2$ quantile of the standard normal distribution), (2) $H_A : m_{i\epsilon I} > \hat{m}_{i\epsilon I}$ when $z > N(0, 1)_{1-\alpha}$, and (3) $H_A : m_{i\epsilon I} < \hat{m}_{i\epsilon I}$ when $z < -N(0, 1)_{1-\alpha}$.

Both tests presented here can be used under any sampling scheme or base model. Application of these or other CFA tests will result in one of three possible outcomes. First, the observed cell frequency, m_i, corresponds to the expected cell frequency, $\hat{m}_i$. In this case, the CFA null hypothesis can be retained. Second, the observed cell frequency, m_i, exceeds the expected cell frequency, $\hat{m}_i$. In this case, the CFA null hypothesis can be rejected, and Cell i is said to constitute a *CFA type*. Third, the observed cell frequency, m_i, is smaller than the expected cell frequency, $\hat{m}_i$. In this case, the CFA null hypothesis can be rejected, and Cell i is said to constitute a *CFA antitype*. For a more formal description of these outcomes, see von Eye and Gutiérrez Peña (2004).

CFA tests differ, in addition to being differentially applicable under various sampling schemes, in the degree to which they suggest conservative decisions. The binomial test tends toward conservative decisions (i.e., the test tends to favor the null hypothesis). In contrast, the z-test tends to be less conservative (for an overview see von Eye & Wiedermann, 2021).

The CFA Base Model

In CFA, the specification of a *base model* is very important. This model is needed for proper interpretation of CFA types and antitypes. It serves as the frame of reference for the meaning of types and antitypes. Different base models can result in different interpretations of the same cell and its observed frequency. According to von Eye (2004), a suitable CFA base model possesses the following four characteristics (sample base models are presented in the following paragraphs): (1) *Uniqueness of interpretation* of types and antitypes. The base model must be specified such that there can be only one reason for the existence of types and antitypes. (2) *Completeness of design matrix*. The design matrix of the base model must contain only and all of those effects of the variables of the cross-classification that are *not* of interest to the researcher. These are effects that are expected due to the design of a study or effects that are not part of the theory to be tested. If the design matrix is complete in this sense, types and antitypes can reflect only those effects that the researcher is interested in. (3) *Parsimony*. A CFA base model must be as parsimonious as possible for clear-cut interpretation of types and antitypes. (4) *Consideration of sampling scheme*. Sampling schemes must be taken into account because they determine whether a CFA base model is admissible. In its simplest form, sampling is multinomial (i.e., cases are randomly assigned to all cells of the cross-classification table). In contrast, the marginals of variables that were observed under a product-multinomial sampling scheme are determined based on theory. They are not the result of sampling. These marginals must be reproduced in CFA. In other words, a CFA base model must contain the effects that allow one to reproduce these marginals. Base models that do not allow one to reproduce these marginal frequencies are not admissible.

We now present sample log-linear CFA base models. We begin with the original CFA model, that is, the one used by Lienert (1968). The question that was asked when this CFA model was developed is whether there exist *local associations* among categorical variables. Local associations are defined as associations between categories of variables instead of variables themselves. To derive estimated expected cell frequencies, the author started from the well-known Pearson χ^2-test statistic. For this statistic, the law of independent events is used. When two variables, say X and Y are jointly observed, this law implies that the probability that two independent variable categories (x_i and y_i) occur simultaneously equals the product of the probabilities of these two categories, i.e., the probability to observe x_i and y_i is given by $p(x_i \cap y_j) = p(x_i)p(y_j)$. This applies accordingly to more than two variables. The estimated expected cell frequency is, under the assumption of variable independence, then calculated as $\hat{m}_{ij} = Np(x_i \cap y_j)$. When the observed cell frequency, m_{ij}, is smaller than the expected, Cell ij constitutes a CFA antitype; when the observed cell frequency is greater than the expected, Cell ij constitutes a CFA type. The base model for these types and antitypes is one of variable independence. Just as a significant Pearson χ^2-test statistic suggests an association between two (or more) variables, a type or an antitype suggests that two categories from two variables are associated.

As von Eye (2002) suggested, this, and most other CFA base models can equivalently be expressed as a log-linear model. In the present example, the model is a main effect model. That is, the model proposes that the variables X and Y may exhibit main effects, but they are not associated. In log-linear model terms, this is the model $\log \hat{m} = \lambda + \lambda^X + \lambda^Y$, where the first term is the intercept and the single-superscripted terms indicate main effects. To illustrate, let X have two categories ($I = 2$), and let Y have three categories ($J = 3$). Then, the design matrix for this model is

$$X = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

The first column in this matrix (the vector of 1s) represents the model intercept. The second column represents the main effect of variable X. Only one parameter is estimated for the two categories, the second category constitutes the reference category. The last two columns represent the main effect terms of variable Y (here, two parameters are estimated, the third category constitutes the reference category). Throughout the article, we estimate log-linear models using effect coding because this coding method makes it easy to see which cells are contrasted (equivalently, dummy coding can be used).

This CFA base model is also called the *first-order base model*. When more than two variables are in a model, this base model still proposes that no interaction (association) exists. Types and antitypes can emerge when any interaction exists. That is, for three variables, types and antitypes can emerge when any or all of the two-way interactions exist, when the three-way interaction exists, or both. When the question is whether higher-interactions beyond first-order main effects and two-way interactions are the causes for types or antitypes, two-way interactions are included in the base model. The resulting base model is, then, a *second-order CFA base model*. For more than three variables, higher-order base models can be specified. In general, first- and higher-order CFA base models all propose that interactions higher than a particular level do not exist. Types and antitypes will then suggest that these interactions do exist, at least locally. The base models for this kind of hypothesis contain all possible effects, with the exception of those that represent the interactions of interest. If these base models are rejected, the hypothesized interactions exist, by necessity.

Prediction CFA: Another important class of base models is that of *Prediction CFA* (P-CFA). This CFA model focuses on relations among predictor and outcome variables. The base model for P-CFA must, therefore, include (1) the main effects of all variables, (2) all possible interactions on the predictor side of the model, and (3) all possible interactions on the outcome side of the model. The only effects that are not part of this base model are those that link predictor with outcome variables. If this model is rejected, predictor variables must be related to outcome variables, at least locally.

To illustrate, consider the three binary variables A, B, and C. Let A and B be predictors, and C is the outcome. The design matrix for the P-CFA base model for these three variables is

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 \end{bmatrix}$$

The first column in this matrix represents the model constant (the intercept). The following three columns represent the main effects of the three variables, A, B, and C. The last column represents the $A \times B$ interaction. The only terms that are missing for the model to be saturated are the interactions that link the predictor variables (A and B) and the outcome (C), that is, $A \times C$, $B \times C$, and $A \times B \times C$. Each of these links one or more variables on the predictor side with the variable on the outcome side. Therefore, when this base model fits the data poorly, predictor-outcome relations must exist.

The importance of the P-CFA model for prevention and intervention research is obvious. It resides in the classification of variables as predictors and outcomes. Types and antitypes indicate which predictor configurations exhibit effects on outcome configurations such that they are observed at rates that are discrepant from those that one would expect were the predictors are independent of the outcomes. In contrast to standard, variable-oriented research, these relations are not expressed at the level of variables but at the level of individual patterns, that is, configurations. These configurations are constituted by predictor and outcome categories. When, as usual, only a selection of predictor configurations is linked to a selection of outcome configurations, the remaining configurations do not deviate from expectancy. That is, these configurations represent sectors in the data space in which predictors and outcomes are independent.

Moderator Effects in CFA: P-CFA was developed to compare groups of variables. However, models of CFA have also been developed to compare groups of individuals. In this case, the grouping variable describes groups of people. It functions as a *moderator* variable. The procedure of such a CFA approach is the same as that of P-CFA as far as the specification of the base model is concerned. It differs in the steps that follow. Specifically, and when two groups are compared, instead of subjecting each individual cell to a CFA test, pairs of cells are contrasted for every configuration. To illustrate, consider two groups, A and B, that are compared in two variables, P_1 and P_2. Now, let configuration ij (i.e., $P_1 = i$ and $P_2 = j$) be the configuration of interest and $\neg\, ij$ (i.e., "not" ij) be all remaining configurations. Then, the grouping variable and the configuration indicator span a 2×2 table. In this table, cell frequencies a, b, c, and d can be compared with the exact Fisher-test, χ^2-tests, or even odds ratios (ORs). If the comparison groups differ statistically, the configuration ij is said to constitute a *discrimination type*. In contrast to standard CFA, there are no discrimination antitypes, because, usually, the grouping variable is nominal level.

An aspect that has not been discussed in the literature on CFA (see, e.g., von Eye et al., 2010) concerns the hypotheses researchers specify in prevention research. In prevention, interventions are often hypothesized to result in the *non-occurrence* of events. For example, teacher classroom management interventions are expected to result in positive student behavior, e.g., reduction of in-school or out-of-school suspensions. If the intervention is effective, the targeted events will occur at rates *below expectation*, that is, below the rates one would expect without the intervention. In termini of CFA, in prevention research, one often interprets antitypes as indicators of a successful intervention.

The Six Steps of CFA

This section introduces the six steps researchers perform in a typical CFA (for more detail, see von Eye, 2002). The steps are as follows:

1. Selection of frequentist or Bayesian CFA; thus far, in this article, frequentist CFA was introduced; Bayesian CFA can be used in particular when information concerning prior distributions can be made part of an analysis (Gutiérrez-Peña & von Eye, 2000);
2. Specification of a CFA base model and estimation of expected cell frequencies; as was explained above, the specification of a CFA base model is guided by (1) theoretical assumptions that concern the nature of the variables as either of equal status or grouped into predictors and criteria and (2) the sampling scheme under which the data were collected; both of these elements determine the effects that are included in the model;
3. Selection of a concept of deviation from independence; there exist multiple ways of deviating from independence; routinely, researchers use those ways that define residuals in log-linear models; Goodman (1991) showed, however, that other ways exist, and von Eye et al. (1995) have illustrated that these ways can result in quite different patterns of types and antitypes in CFA;
4. Selection of a significance test; in general, exact tests are preferred in CFA applications; however, when a sample is large enough, approximative tests can provide considerably more statistical power;
5. Identification of configurations that constitute CFA types or antitypes;
6. Interpretation of the resulting CFA types and antitypes.

In the following sections, we present three real-world-data examples of CFA in the context of intervention studies.

Data Examples

The data that are used in the following examples were collected in a large-scale RCT that evaluates the impact of the Incredible Years Teacher Classroom Management (IY TCM) program on student social-emotional and academic outcomes (Reinke et al., 2018). IY TCM was designed to promote effective classroom management practices (such as effective praise, proactive teaching strategies, and use of time out procedures) for preschool and early elementary teachers. In addition, the IY TCM focuses on ways to promote students' prosocial skills, increased parents' involvement, and positive teacher-parent relationships (Webster-Stratton et al., 2004). The study included 1817 students and 105 teachers in kindergarten to 3rd grade from nine schools in a school district in the Midwestern part of the USA. Using a cluster randomized design, teachers were randomly assigned to receive IY TCM or no treatment. The majority of students in the study identified as Black (76%; 22% White, 2% other) and received free/reduced lunch (61%). Fifty-two percent of the students were male, and about 9% received special education service. For the purpose of demonstrating the application of CFA, we focus on both, a non-randomized (observational) setting and a randomized intervention setting. In the non-randomized intervention setting, we use a baseline grouping variable that was not under experimental control (behavioral service provided at baseline). In the randomized intervention scenario, we focus on the impact of the IY TCM. For both scenarios, we excluded students with missing values in the baseline and post-treatment measures. The first data example included n = 1661 students (91%); the second data example used data from $n = 1671$ (92%) students. To test for potential differences between the discarded subsamples and the analysis samples, we used a multiple logistic regression model (with cluster-robust standard errors [SEs] clustered on teachers to account for the nested data structure) to predict missing values in baseline and posttreatment measures from available demographic information (students' age, race, gender), free/reduced lunch status, special education status, school membership, and treatment status. In both datasets, students in one school (School 6 in Table 1) were at higher risk for missing values (OR = 2.17, 95% CI = [1.15, 4.08] in the smaller data set and OR = 2.19, 95% CI = [1.16, 4.14] in the larger data set). No further significant differences were found between the analysis sample and the corresponding subsample with incomplete data. Table 1 shows the descriptive statistics for the larger analysis sample by treatment status. Analyses were performed using the R statistical programming environment (R Core Team, 2021). Configural frequency modeling was performed using the R package `confreq` (version 1.5.4–3; Stemmler & Heine, 2017). R scripts for all analyses are given in the online supplement of the article.

Example 1: Disruptive Behavior and Behavioral Services

For the first example, we use the variables *Behavioral Services or Support Provided at Baseline* (B), *Disruptive Behavior at Baseline* (D_1), and *Post-Treatment Disruptive Behavior* (D_2). Behavioral service/support was measured using a single item from the *Teacher Observation of Classroom Adaption-Checklist* (TOCA-C; Koth et al., 2009),

Table 1 Descriptive statistics for student emographics and study outcomes

		Control		Treatment		
		n = 835		n = 836		p-value*
Female	n (%)	414	(49.6)	401	(48.0)	0.541
Black	n (%)	622	(74.5)	637	(76.2)	0.452
Lunch	n (%)	510	(61.1)	499	(59.7)	0.596
Special ed	n (%)	81	(9.7)	70		0.389
Behavioral service						0.847
No		793	(95.0)	793	(94.9)	
Yes		41	(4.9)	41	(4.9)	
Don't know		1	(0.1)	2	(0.2)	
School						0.059
(1)	n (%)	76	(9.1)	88	(10.5)	
(2)	n (%)	100	(12.0)	76	(9.1)	
(3)	n (%)	87	(10.4)	114	(13.6)	
(4)	n (%)	111	(13.3)	101	(12.1)	
(5)	n (%)	97	(11.6)	83	(9.9)	
(6)	n (%)	56	(6.7)	71	(8.5)	
(7)	n (%)	110	(13.2)	118	(14.1)	
(8)	n (%)	96	(11.5)	73	(8.7)	
(9)	n (%)	102	(12.2)	112	(13.4)	
Age	M (SD)	7.06	(1.10)	7.15	(1.22)	0.125
Baseline emotion regulation	M (SD)	3.02	(1.18)	3.12	(1.21)	0.079
Post-treat. emotion regulation	M (SD)	3.48	(1.12)	3.63	(1.12)	0.005
Baseline disruptive behavior	M (SD)	1.78	(0.74)	1.76	(0.78)	0.516
Post-treat. disruptive behavior	M (SD)	1.92	(0.80)	1.85	(0.83)	0.104

n frequency, *M* mean, *SD* standard deviation

*Student t-tests were performed for continuous variables, and χ^2-tests were used for categorical variables

"Does this child currently receive additional classroom-based behavioral services or supports?" with responses 1 = *no* (94.7%), 2 = *yes* (5.1%), and 3 = *don't know* (0.2%). D_1 and D_2 are mean composite measures of the disruptive behavior subscale of the TOCA-C (e.g., "breaks rules" and "fights"; item responses ranged from *never* (1) to *almost always* (6). Higher scores indicate more disruptive behavior. Baseline and post-treatment Cronbach's alpha values for this subscale were 0.91 (95% CI = [0.90, 0.92]) and 0.92 (95% CI = [0.91, 0.93]). The general hypothesis under which we analyze the cross-classification of these three variables is that under consideration of D_1, B is related to D_2 in the sense that disruptive behavior decreases in frequency. Cases for whom it was unknown whether or not behavioral services or support had been provided were omitted from the analyses. Before performing CFA, we transformed variables so that a cross-classification can meaningfully be created. The transformation proceeded in two steps: (1) The scores of D_1 and D_2 were rounded to the next integer and (2) the resulting scores of D_1 and D_2 were set to 4 when they were 4 or greater. This was done for two reasons; first, the cross-classification of D_1, D_2, and B will contain fewer cells and will, thus, be easier to manage, and, second, scores above 4 were observed so rarely that the cross-classification that would be sparse had the larger scores not been condensed. These transformations resulted in a 2 × 4 × 4 cross-classification. The observed frequencies for this cross-classification appear in Table S1 of the electronic supplement.

We analyzed the data in the following steps. First, we estimated a log-linear main effects model $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2}$. As was discussed above, this is the base model of first-order CFA. If types and antitypes emerge from this model, they suggest that associations exist among the three variables that span the cross-classification. The expected cell frequencies for this model appear in Table S1, in the column "Model 1." Even a cursory comparison of the observed cell frequencies with those expected for Model 1 shows that there are dramatic discrepancies. We refrain, at this point, however, from interpreting the many types and antitypes. The main reason for this is that we did not ask whether there exist just any associations. We asked whether particular associations, specifically the associations among B and D_1 and D_2, result in types and antitypes. Therefore, we estimated a series of additional log-linear models and performed, finally, a CFA. In the second log-linear model, we included the interaction between D_1 and D_2. The reason for this is that in empirical longitudinal data, repeatedly observed behaviors tend to be strongly correlated. This correlation must be part of the base model because we hope to detect types and antitypes that are based on the relation of D_1 and D_2 with B, but not on the relation of D_1 with D_2. This model is $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2}$. The overall goodness-of-fit results for the first two and the three following models are given in Table 2. This table shows that the first two models are far from describing the frequency distribution in a satisfactory way. The same applies to the third model, Model 3. This model contains, in addition to $D_1 \times D_2$, the interaction $B \times D_2$. It is, thus, $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2} + \lambda^{B,D2}$. Table 2 shows that this model also fails to explain the frequency distribution. It shows, however, a significant improvement over Model 2. Model 4 adds to the terms in Model 3 the one that represents the interaction $B \times D_1$. This model is $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2} + \lambda^{B,D2} + \lambda^{B,D1}$ (adding the three-way interaction would render this model saturated). This is the only model that explains the data well (cf. Table 2). Removing the term for the interaction $B \times D_2$ results in Model 5, that is the model $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2} + \lambda^{B,D1}$. This model again results in relatively poor fit (cf. Table 2).

Table 2 contains, in its last column, the results of the model comparisons with the model in the line before. Each pair of these models is hierarchically related and can, therefore, directly be compared. From these comparisons, we conclude that up to Model 4, each model is significantly better than the more parsimonious model in the line before. Model 5 is more parsimonious than Model 4, but its fit is significantly worse. The conclusion from these results is as follows: model fit improves significantly when the relations between B and D_1 and D_2 are made part of the model. We now ask whether the effects of behavioral service on disruptive behavior are uniform or exceptionally strong or weak in particular sectors of the data space, that is, the cross-tabulation. This question can be answered by CFA.

Following the CFA steps listed above, we first decide to perform a frequentist CFA. The reasons for this are first, that this is by far the most popular version of CFA, and second, that the framework for a Bayesian CFA was not developed for the present example. Second, we develop a base model. In the preliminary analyses in which we estimated a series of log-linear models, we found that when all 2-way interactions that involve the variable Behavioral Service or Support (B) are included in the model, the data can be satisfactorily explained; that is, there remain no significant data-model discrepancies. There are no statements about the sectors in the data space in which the effects of the behavioral services can be seen. CFA focuses on the identification of these sectors. In other words, CFA identifies those configurations in which the deviations from the null hypothesis that psychological services have no effect are most pronounced. In the base model for CFA, we therefore exclude all interactions that involve B. These are the interactions $B \times D_1$, $P \times D_2$, and $B \times D_1 \times D_2$. The model is, thus, $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2}$. This is the same model as Model 2, above. Here, however, we are not interested in model fit and parameter interpretation. Instead, we are interested in the identification of types and antitypes, if they exist. According to the fourth step of a CFA, we select a proper significance test. Here, we opt for the z-test. The sample is large enough that we can trust the approximation characteristics of the test, and the z-statistic is routinely estimated in log-linear modeling as well, thus showing one of the relations of CFA and log-linear modeling. To protect the significance threshold, α, we opt for the Bonferroni procedure. This procedure results in a protected threshold $\alpha^* = \alpha/t$, where t is the total number of significance tests that are performed. In the present case, we perform one test for each cell of the cross-classification in Table S1, that is, $2 \times 4 \times 4 = 32$ tests. For a nominal $\alpha = 0.05$, we obtain a Bonferroni-protected α of $\alpha^* = 0.05/32 = 0.0016$. The corresponding z-score is $z^* = \pm 2.955$. z-scores smaller than

Table 2 Likelihood ratio (LR) goodness-of-fit tests and model comparisons for both empirical examples

No	Model specification	Model Fit	Model comparison
Model 1	[B][D1][D2]	LR-$\chi^2(24) = 951.53$, $p < 0.001$	-
Model 2	[B][D1 D2]	LR-χ^2 $(15) = 171.20$, $p < 0.001$	M_1 vs. M_2: $\Delta\chi^2$ $(9) = 780.34$, $p < 0.001$
Model 3	[D1 D2][B D2]	LR-χ^2 $(12) = 54.93$, $p < 0.001$	M_2 vs. M_3: $\Delta\chi^2$ $(3) = 116.27$, $p < 0.001$
Model 4	[D1 D2][B D1][B D2]	LR-χ^2 $(9) = 10.46$, $p = 0.314$	M_3 vs. M_4: $\Delta\chi^2$ $(3) = 44.47$, $p < 0.001$
Model 5	[D1 D2][B D1]	LR-χ^2 $(12) = 23.21$, $p = 0.026$	M_5 vs. M_4: $\Delta\chi^2$ $(3) = 12.75$, $p = 0.005$

B behavioral service at baseline, *D1* baseline disruptive behavior, *D2* post-treatment disruptive behavior; Model specification is summarized using bracket notation for log-linear models. Here, [*X*] refers to the main effect of *X*, [*X Y*] refers to the interaction effect of *X* and *Y* with main effects being implied. Reading example: [B][D1 D2] represents the model $\log \hat{m} = \lambda + \lambda^B + \lambda^{D1} + \lambda^{D2} + \lambda^{D1,D2}$

$-$ 2.955 indicate CFA antitypes, and z-scores > + 2.955 indicate CFA types. The results of this CFA are presented in Table S2 in the online appendix.

The results suggest that one CFA antitype and seven CFA types emerged. Before, however, we present detailed interpretations; we note one interesting result. All types and the antitype emerge for those students who are provided behavioral services. For students that do not receive behavioral services, there is no effect of B. This result is most plausible. What follows is the individual type-antitype interpretation.

Antitype 2 1 1. This antitype suggests that fewer students than expected exhibit low levels of disruptive behavior at both observation points when they are provided behavioral services. This is because students with higher risk for disruptive behavior were more likely to be offered behavioral services in the first place. Six students showed this pattern, but over 45 had been expected.

Type 2 2 3. This type describes students who, over time, increase their levels of disruptive behavior from relatively low range to mid-range. Nine students showed this pattern, but only about three had been expected to show it. These are the students who counter the intentions of the services that were provided.

Type 2 3 2. These students lower their level of disruptive behavior over time. This suggests that the behavioral services that were provided can be successful. Eleven students showed this pattern, but only about two had been expected to show it.

Type 2 3 3. These students do not change the level of disruptive behavior over the course of the observation period. Here again, one would conclude that for these students, the provided services are not successful. Fifteen students showed this pattern, but only about three had been expected.

Type 2 3 4. These students increase their level of disruptive behavior from moderate to high. This is the third type that could lead to the conclusion that for these students, the provided services are not successful. Five students showed this pattern, but only about one had been expected to show it.

Type 2 4 2. These students decrease their level of disruptive behavior from high to moderate. They, thus, also respond to the service as intended. Three students showed this pattern, but close to nobody had been expected to show it.

Type 2 4 3. These students decrease their level of disruptive behavior from high to a value that is lower, but still above average. Here again, we conclude that the intervention has met with success. Three students showed this pattern, but fewer than one had been expected to show it.

Type 2 4 4. The last type describes the four students (fewer than one was expected) who keep their level of disruptive behavior at an extremely high level. This is another configuration that indicates lack of success of the intervention.

In sum, we note that behavioral services or support met with mixed success. This result can be used by intervention researchers to contrast characteristics of students that respond as intended with characteristics of students that do not respond as intended. The service could possibly be optimized based on student characteristics.

Finally, we integrate person- and variable-oriented perspectives and ask whether CFA Type membership can be explained based on baseline demographics (gender, race, age, free/reduced lunch, receiving special education service, and school membership) in a variable-oriented follow-up analysis. Specifically, we used multiple logistic regression (with cluster-robust SEs clustered on teachers) to predict membership in one of the four CFA types describing suboptimal developmental trajectories, i.e., Types 2 2 3 ($n = 9$), 2 3 3 ($n = 15$), and 2 3 4 ($n = 5$), and 2 4 4 ($n = 4$). No significant effects were observed for students' age, race, and free/reduced lunch status. However, female students show significantly lower risks to be part of these CFA types (OR = 0.32, 95% CI = [0.15, 0.71]) compared to males, and students receiving special education services are at higher risk to belong to these CFA Types (OR = 2.52, 95% CI = [1.03, 6.18]).

Example 2: Configural Moderation Analysis

In the following example, we perform a configural moderator analysis. Based on the variable-oriented gender effects given above, we now ask whether female and male students who do versus do not receive behavioral services (B) differ in their development of disruptive behavior from a configural perspective. In the moderator model, we consider B, D_1, and D_2, the discrimination variables, and Gender (G) the grouping variable. The CFA base model that is used to evaluate overall goodness-of-fit considers the main effects of all variables, and all possible interactions among B, D_1, and D_2, that is $B \times D_1$, $B \times D_2$, $D_1 \times D_2$, and $B \times D_1 \times D_2$. However, this base model proposes independence between the discrimination variables on one side and the grouping variable on the other. Each configuration of the discrimination variables is subjected to a test as illustrated in Table S3 in the online appendix. α is protected not based on the number of cells, t, but on the number of 2×2 tables that is examined. This number is $t/2$. To perform moderation CFA in the present example, we substitute the estimation of expected cell frequencies by the configuration-wise group comparisons. Table S3 in the online supplement displays the results of these comparisons. In the present context, we opt for the χ^2-related concept of deviation from independence. The reason for this is that this concept is marginal-dependent (Goodman, 1991). In other words, the test of the 2×2 table uses the marginal frequencies as weights. Alternatively, a marginal-free measure could have been used, for example, the odds ratio (for a discussion and comparison of marginal-dependent and marginal-free measures, see Goodman, 1991). The results in Table S3 suggest the existence of two discrimination types. The first of these is constituted by Configuration 1 1 1. This profile describes students who did not receive behavioral services and displayed minimal disruptive behavior at both observation points. This profile is observed by 403 male and 515 female students. The second discrimination type is constituted by Configuration 1 1 2. These students did not receive behavioral services, exhibited minimal disruptive behavior at D_1 but increased disruptive behavior at D_2. A total of 128 male students showed this profile, but only 69 female students. Thus, the most common discrimination type 1 describes students who were accurately perceived as not needing behavior support and who exhibited no disruptive behaviors over time. The second type may be construed to be students, predominately males who were accurately identified as not needing support at baseline but developed problems that increased their disruptive behaviors. This type may benefit from subsequent screening to allow for intervention prior to the onset of the emergence of disruptive behaviors.

Example 3: IY TCM and Emotional Regulation

For the last example, we use variables from the same data set as in the previous examples. Specifically, we use the variables Treatment (T; 0 = no treatment; 1 = IY TCM treatment), emotional regulation at baseline (E_1), and post-treatment emotional regulation (E_2). The two emotional regulation variables are mean composites of the corresponding subscale of the *Revised Teacher Social Competence Scale* (T-COMP; Gifford-Smith, 2000; e.g., "show verbal and physical aggression," "handle disagreements"; responses range from *almost never* (0) to *almost always* (5). Baseline and post-treatment Cronbach's alpha estimates were 0.79 (95% CI = [0.78, 0.81]) and 0.90 (95% CI = [0.89, 0.91]). These two continuous variables were categorized for the same reasons and walking the same steps as in the first data example. The resulting sample size was $n = 1671$. Crossed, these three variables span a $2 \times 4 \times 4$ table. In a fashion analogous to the first example, we ask whether the randomized IY TCM intervention resulted in a reduction of emotional regulation. The preliminary log-linear analyses and the comparison of CFA results with those from log-linear and regression-type analyses are parallel to the ones executed in the first example. In difference to the first example, we employ confirmatory instead of exploratory CFA. In exploratory CFA, all cells in a cross-classification are subjected to CFA tests. In confirmatory CFA, a selection of individual cells or groups of cells is tested. The main advantages of this procedure are that (1) very detailed hypotheses can be targeted and (2) α protection results in less extreme thresholds as in exploratory CFA. In this example, we perform two CFAs. The first uses the log-linear base model of first-order CFA $\log \hat{m} = \lambda + \lambda^T + \lambda^{E1} + \lambda^{E2}$. Table S4 in the online appendix shows the results of this analysis. The goodness-of-fit χ^2 suggests that the base model is not tenable ($\chi^2 = 486.6$; $df = 24$; $p < 0.001$). Accordingly, seven types and eight antitypes emerged.

Instead of interpreting the types and antitypes, we ask in particular, whether Configurations 3 4 and 4 4 show treatment effects. That is, we ask whether, in comparison with the control group, IY TCM training (1) increases students' emotional regulation from level $E_1 = 3$ to level $E_2 = 4$ and (2) resulted in more stable cases that stayed in the high regulation group ($E_1 = E_2 = 4$). Statistically, the group comparisons are similar to tests of configural moderation analysis described above. That is, a dummy indicator representing the configuration of interest (e.g., 1 = configuration of interest, 0 = remaining configurations) and the treatment status (1 = IY TCM, 0 = no treatment) span a 2×2 table and cell frequencies are compared with a χ^2-test. Table S5 in the online appendix displays the results of a group comparison of the $E_1 \times E_2 \times T$ cross-classification. The table does not show the usual classification of pairs of configurations as discrimination type (or not) that would be based on significance thresholds that consider all 16 comparisons. Instead, we focus, in a confirmatory manner, on the two selected pairs 3 4 0 vs. 3 4 1 and 4 4 0 vs. 4 4 1. Based on Holm adjustment, the protected significance threshold for the first of these two comparisons is $\alpha^* = 0.05/2 = 0.025$. The protected threshold for the second comparison is $\alpha^* = \alpha$. Here, the first protected significance threshold is the same as the one under Bonferroni. For the second, we used the statistically more powerful Holm threshold.

The Pearson goodness-of-fit test suggests that the base model for this two-group CFA fails to describe the data well ($\chi^2 = 42.3$; $df = 15$; $p < 0.001$). Therefore, we anticipate that the confirmatory tests can unearth discrimination types. The first of these tests is performed for Configuration 3 4. This configuration suggests an improvement in emotion regulation by one step from Level 3 to 4. 40 respondents of the control group, but 76 respondents in the treatment group exhibited this improvement. This difference is significant, because $p = 0.0005$ is less than the protected $\alpha^* = 0.025$. The second confirmatory test is performed for Configuration 4 4. This pattern describes stable respondents. These students showed at both points in time high levels of emotion regulation. A total of 214 of these are in the IY TCM group, and 177 in the control group. This discrepancy is also significant because the protected α^* for this test equals the nominal $\alpha = 0.05$, and the probability of this group difference is $p = 0.034$. Both of these patterns can be interpreted as successes for the treatment. Results suggest that the IY TCM increases students' emotional regulation in particular for those who are comparatively well controlled to begin with. IY TCM supports teachers in providing social emotional coaching to students to increase their emotional regulation skills. These data suggest that these efforts are particularly impactful for students with adequate baseline levels of regulation; students with serious dysregulation at baseline may require more intensive supports to develop effective self-regulation capacities. This is an important finding consistent with a universal prevention approach with the goal of strengthening skills for students even with lower risks and thereby reducing the number of students who need more intensive support in the future.

In the last step, we again integrate person- and variable-oriented perspectives and use multiple logistic regression with cluster-robust SEs (clustered on teachers) to evaluate whether the selected configurations (3 4 and 4 4) can be further explained by student covariates. Again, students' gender, race, age, free/reduced lunch status, special education service status, and school membership were used as predictors. In addition, we accounted for treatment effects and tested all possible two-way interactions. For Configuration 3 4, no significant results other than the treatment effect (OR = 2.18, 95% CI = [1.33, 3.55]) were observed and all two-way interaction were non-significant. For Configuration 4 4, we obtain a non-significant treatment effect (OR = 1.33, 95% CI = [0.82, 2.17]). However, female students had a higher chance of keeping high levels of emotional regulation (OR = 2.52, 95% CI = [1.88, 3.38]) and students receiving special education service were at risk of lower emotional regulation over time (OR = 0.59, 95% CI = [0.39, 0.91]).

Discussion

In this article, we introduce configural frequency analysis (CFA) as a method for detailed, person-oriented analysis of prevention and intervention effects. Exploratory CFA allows one to search for sectors in a cross-classification in which the null hypothesis of no intervention effect is contradicted. Confirmatory CFA allows the researcher to test particular, cell-specific hypotheses about the effects of intervention. In prevention research, researchers often ask whether prevention measures were successful in the sense that specific behaviors do not exist or cease to be exhibited. CFA is particularly suited to test this kind of hypotheses because CFA antitypes suggest behaviors that are observed at rates below expectation. In the present context, expected values are estimated under the null hypothesis of no intervention effect. The option of searching for specific antitypes is new in the CFA literature.

When compared with methods from the family of generalized linear models, CFA produces results that target individual cells or groups of cells. By implication, CFA can tell the researcher, where, in a cross-classification, effects are particularly strong, and where they are missing, that is, where observed frequencies do not differ from expectancy. As was shown in the data examples, statements of variable relations are not necessarily true in all sectors of a cross-classification. CFA can show the researcher where they are true and where they are not true.

In the present study, CFA was first used to document patterns of response to a naturally occurring intervention offered to youth in schools—classroom behavior supports or services as reported by their teachers. CFA identified eight patterns of disruptive behaviors at baseline and end-of-year for students who received these services, three of them indicating successful response to the behavior service or support. The other patterns were indicative of non-response and thus identify youth who needed different or more intensive behavior services. It is important to note that the behavior services or supports reported by teachers were delivered at the classroom level, and were not intensive, individualized interventions. Instead, these supports are best construed as Tier 2, or selective/indicated types of supports that are typically economical and low intensity (Bruhn et al., 2014). Thus, it is not surprising that a subset of students would respond to these preventative supports and another comparably large subset would need more intensive services. CFA offers a method for identifying youth most likely to benefit from these less intensive supports and those that are likely to need something more. For instance, subsequent analyses revealed that males and students in special education were significantly more likely to have patterns indicating non-response to intervention. Thus, the data suggest that youth with these characteristics who are receiving classroom behavior supports should be closely monitored for worsening or not improving levels of disruptive behavior and provided different or more individualized services earlier in the school year.

A second set of analyses examined configural moderation effects. Results identified two additional pattern types. The most common pattern described a large subset of students who did not receive behavior supports and who displayed low levels of problem behaviors at both timepoints. These reflect a subset of youth who were accurately identified as not needing behavior supports. The second pattern involved youth, mostly males, who were accurately identified as not needing support at baseline but who escalated their disruptive behaviors by the end of the year. Given that males are at greater risk for life-course persistent aggressive and disruptive behaviors (Schaeffer et al., 2006), early identification of youth exhibiting this pattern of increasing disruptive behaviors in the absence of intervention is essential. Additional, screening later in the fall semester may be needed to identify youth who do not need supports at the start of the year but experience increasing problems as the year progresses.

The final set of analyses focuses on patterns of emotional regulation response to the IY TCM intervention. The findings are consistent with Greenberg and Abenavoli (2017) call for examining heterogeneous responses to universal preventive interventions. Rather than summarizing main effects across participants, CFA allows for the unpacking of nuanced intervention responses that would otherwise have remained undetected. For instance, the finding that students with adequate self-regulation skills at baseline experienced significant improvement in their skills in response to IY TCM provides important information about the value and particular impact of this universal prevention approach. It implies that IY TCM may have a preventative effect by making it less likely students who do not currently need supports will experience an even stronger protective buffer making it less likely they will need intensive supports in the future. A universal prevention program like IY TCM may help overcome the *prevention paradox* (Greenberg & Abenavoli, 2017), that is, the documented reality that the majority of cases of a disease come from the segment of the population at low or moderate risk for the disease. By bolstering the regulation skills of youth at low risk for dysregulation, IY TCM may reduce the population prevalence of emotional dysregulation and its corollaries (e.g., conduct and disruptive behavior problems). Additionally, the heterogeneous benefit also implies that youth at baseline without adequate self-regulation skills will need additional supports to recover. That is, the universal IY TCM approach may not be a potent enough intervention for those with more intensive emotional regulation deficits.

Extensions of the models that were discussed here can go in a number of directions. One direction is that of specification of configural mediator models (von Eye et al., 2010; Wiedermann & von Eye, 2021). Here, one explicitly distinguishes between a predictor, an outcome, and an intermediate variable (the mediator) that transmits the effect from the predictor to the outcome. Variable-oriented mediation analysis is commonly used to explain why an intervention works. In CFA, we ask which configurations carry the mediation mechanism at the level of individual cells of cross-classifications.

Another extension of CFA methodology concerns the analysis of cluster randomized designs. In these designs, groups of cases, instead of individuals, are randomized. In general, ignoring clustered data structures is known to bias SEs of statistical models—point estimates of regression slopes, however, tend to be less affected. This also holds for the Poisson regression model (Demidenko, 2007) used to estimate CFA base models. Thus, in CFA, biases in type/antitype decisions as a result of ignoring clustering can be expected to be low, because expected cell frequencies (the key information used to detect types/antitypes) only rely on estimated regression slopes of the base model. When one wishes to explore clustering effects from a configural perspective, cluster information must be made part of the base model. Moderation analysis is a common approach for examining subgroup effects but has limitations that interfere with their utility in prevention research (Lanza & Rhoades, 2013). For instance, moderation analyses of complex processes can require an exorbitant number of tests resulting in heavily inflated Type I error rates. In addition, analyses with multiple moderators may require higher-order interaction terms (3- and 4-way interactions) that can be complicated to interpret, and the power is limited to detect them. In the example used in this article, participants were students in nine schools. Assuming that prevention effects vary across these nine schools, this design feature could be considered. One option for this is to consider school membership as a moderator. This, however, would have made the analyses rather complex and unwieldy. The number of cells of the cross-classification to be analyzed would have increased from $4 \times 4 \times 2 = 32$ (cf. Table S1) to $4 \times 4 \times 2 \times 9 = 288$. Many cells would, then, have contained just a very small number of cases, or even none, and types and antitypes would be hard to identify. Alternatively, schools could have been classified, e.g., in those in well-to-do neighborhoods and those in disadvantaged neighborhoods. This would considerably reduce the size of the cross-classification to be analyzed. The thus created table could be analyzed with, for example, multiple group CFA. Results of this kind of analysis include explicit, that is, statistical comparisons of groups of schools, and that in addition to the here presented results. This is material for future work.

One question that needs to be answered concerns the relation of the hybrid approach proposed here and stand-alone CFA. The latter approach is part of the hybrid approach. The value-added by also performing logistic regression concerns the relation between variable- and person-oriented research strategies. The former provides broad-stroke overall results that can be of use when general trends are of interest. The latter can be of use when the assumption is tested that effects exist in specific sectors of the data space. Therefore, the two approaches complement each other, as was illustrated in the data examples in this article.

In sum, CFA is a flexible method that can be used instead of or in tandem with variable-oriented methods such as log-linear modeling or logistic regression.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11121-021-01241-8.

Funding This study was supported by a 5-year grant from the US Department of Education, Institute of Education Sciences (no. R305A100342; to the third and fourth authors) submitted to CFDA 84.324A.

Declarations

Ethics Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants in the study.

Conflict of Interest The authors declare no competing interests.

References

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

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

Bogat, G. A., von Eye, A., & Bergman, L. R. (2016). Person-oriented approaches. In D. Chicchetti (Ed.), *Handbook of developmental psychopathology.* (pp. 797–845). Wiley.

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2014). A review of tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders, 22*(3), 171–189.

Demidenko, E. (2007). Poisson regression for clustered data. *International Statistical Review, 75*(1), 96–113.

Gifford-Smith, M. (2000). *Teacher social competence scale, fast track project technical report*. Duke University.

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories*. Harvard University Press.

Goodman, L. A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. *Journal of the American Statistical Association, 86*, 1085–1111.

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness, 10*(1), 40–67.

Gutiérrez-Peña, E., & von Eye, A. (2000). A Bayesian approach to configural frequency analysis. *Journal of Mathematical Sociology, 24*, 151–174.

Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods, 21*, 846–876.

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.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*, 157–168.

Lienert, G. A. (1968). *Die "Konfigurationsfrequenzanalyse" als Klassifikationsmethode in der klinischen Psychologie*. Paper presented at the 26. Kongress der Deutschen Gesellschaft für Psychologie in Tübingen 1968.

Lienert, G. A., & Krauth, J. (1975). Configural frequency analysis as a statistical tool for defining types. *Educational and Psychological Measurement, 35*, 231–238.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific Psychology—This time forever. *Measurement, 2*, 201–218.

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

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

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.

Schaeffer, C. M., Petras, H., Ialongo, N., Masyn, K. E., Hubbard, S., Poduska, J., & Kellam, S. (2006). A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction to young adult antisocial outcomes. *Journal of Consulting and Clinical Psychology, 74*, 500.

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.

Thompson, A. M., Wiedermann, W., Herman, K. C., Reinke, W. M. (2019). Effect of daily teacher feedback on subsequent motivation and mental health outcomes in fifth grade students: A person-centered analysis. *Prevention Science*.

von Eye, A. (2002). *Configural Frequency Analysis: Methods, models, and applications*. Lawrence Erlbaum.

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

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.

von Eye, A., Bergman, L. R., & Hsieh, C. -A. (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). Wiley.

von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis—The search for extreme cells. *Journal of Applied Statistics, 31*, 981–997.

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

von Eye, A., & Mun, E. -Y. (2016). Configural frequency analysis for research on developmental processes. In D. Cicchetti (Ed.), *Handbook of developmental psychopathology.* (pp. 866–921). Wiley.

von Eye, A., Spiel, C., & Rovine, M. J. (1995). Concepts of nonindependence in configural frequency analysis. *Journal of Mathematical Sociology, 20*, 41–54.

von Eye, A., & Wiedermann, W. (2021). *CFA. Configural Frequency Analysis*. Springer (in preparation).

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

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

Wiedermann, W., & von Eye, A. (2021). A Simple Configural Approach for Testing Person-Oriented Mediation Hypotheses. Integrative Psychological and Behavioral Science. online first. https://doi.org/10.1007/s12124-020-09598-1