

On the Specification of Models for Configural Frequency Analysis – Sampling Schemes in Prediction CFA

Alexander von Eye*
Michigan State University

Christof Schuster
University of Michigan

Zusammenfassung

Zur Spezifikation von Modellen für die Konfigurationsfrequenzanalyse – Stichprobeneigenschaften und die Prädiktions-KFA

In diesem Artikel wird die Spezifikation von Modellen für die Konfigurationsfrequenzanalyse diskutiert. Üblicherweise beachten weder Methodiker noch Benutzer bei der Spezifikation von Modellen für die KFA Art und Eigenschaften der Stichprobenziehung. In diesem Artikel regen wir an, bei der Modellspezifizierung zu berücksichtigen, daß Variablen entweder multinomial oder produkt-multinomial sein können. Ein Ergebnis der Berücksichtigung dieser Eigenschaften ist, daß eines der ungelösten Probleme der KFA, und zwar die Nicht-Unterscheidbarkeit der Modelle für die Interaktionsstrukturanalyse (ISA) und der Prädiktions-KFA (PKFA), gelöst werden kann, weil sie sich in ihren zugrunde liegenden Modellen unterscheiden können. Sowohl die ISA als auch die PKFA unterteilen Variablen in zwei Gruppen. Das KFA-Modell, das bisher für beide Ansätze verwendet wurde, postuliert, daß (1) die Beziehungen zwischen den Variablen innerhalb der Gruppen saturiert sind und (2) zwischen den Variablengruppen Unabhängigkeit herrscht. Die Anwendbarkeit dieses Modells auf die PKFA ist debattierbar. Beispiele von Modellen, die für die PKFA geeignet sind, werden vorgestellt, und es werden Regeln formuliert, anhand derer geeignete Basismodelle für die KFA ausgewählt werden können. Es wird gezeigt, wie Typen und Antitypen interpretiert werden können, die auf der Basis unterschiedlicher KFA- Modelle mit unterschiedlichen Stichprobeneigenschaften gefunden wurden. Datenbeispiele illustrieren erneut, daß Muster von Typen und Antitypen, die aufgrund unterschiedlicher KFA- Modelle gefunden wurden, sich dramatisch voneinander unterscheiden können.

Schlüsselwörter: Konfigurationsfrequenzanalyse, KFA-Modelle, Stichprobenziehung, Interaktionsstrukturanalyse, Prädiktions-KFA

Abstract

On the Specification of Models for Configural Frequency Analysis - Sampling Schemes in Prediction CFA

*Address correspondence concerning this article to Alexander von Eye, Michigan State University, Department of Psychology, 119 Snyder Hall, East Lansing, Michigan 48824-1117, USA; e-mail: VONEYE@PILOT.MSU.EDU. Alexander von Eye's work on this article was supported in part by NIAAA grant # 2RO1 AA7065. The authors are indebted to one of the two anonymous reviewers and to Jürgen Rost for helpful and stimulating discussions of earlier versions of this article.

This article is concerned with the specification of base models for Configural Frequency Analysis. Typically, neither methodologists nor users of CFA consider the type of sampling when specifying CFA base models. In this article we propose considering the sampling scheme of the variables used in CFA as either multinomial or product multinomial. As a result of this consideration, one of the problems of CFA, that is, the indistinguishability of the base models for Interaction Structure Analysis (ISA) and Prediction CFA (PCFA) can be solved, because different models can be specified. Both ISA and PCFA divide variables in two groups. The base model that is commonly employed is saturated in these groups and assumes independence between groups. Application of this model in PCFA is debatable. Examples of base models are provided, and rules are presented that can be used for selection of suitable CFA base models. It is shown how to interpret types and antitypes based on different base models from different sampling schemes. Data examples illustrate again that the patterns of types and antitypes that emerge from different base models can be dramatically different from each other.

Key Words: Configural Frequency Analysis, CFA base models, sampling schemes, Interaction Structure Analysis, Prediction CFA

On the Specification of Models for Configural Frequency Analysis – Sampling Schemes in Prediction CFA

Configural Frequency Analysis (CFA; Lienert, 1969; von Eye, 1990) employs base models custom tailored to the assumptions made when searching for types and antitypes. These base models are termed global when all variables have the same status. The models are termed regional when variables differ in status. Sample models of CFA where variables differ in status include Interaction Structure Analysis (ISA; Krauth & Lienert, 1974) where variables form typically two groups. Interactions between these groups manifest in ISA types and antitypes. The groups are not classified in terms of independent and dependent variables, or predictors and criteria. Another regional model is Prediction CFA (PCFA; Lienert & Krauth, 1973; cf. von Eye, 1985). PCFA also specifies two groups of variables, but these are interpreted as the predictors and the criteria. Thus, the difference between ISA and PCFA is that whereas in ISA there is the assumption of a non-directed relationship between the two groups of variables, in PCFA the relationship is directed. This applies accordingly to k-sample or discrimination CFA (Lienert, 1971). This article is concerned with criteria for proper specification of base models. Specifically, it is shown that the base models for ISA and PCFA can differ under the appropriate sampling schemes.

In the following sections, a brief overview of CFA and its base models is given (for more detail see von Eye, 1990; von Eye, Spiel, & Wood, 1996). In the subsequent sections, the base models for ISA and PCFA are described in some detail, with special emphasis on sampling schemes.

1 Configural Frequency Analysis

Consider the saturated log-frequency model for the two groups of variables, A and B,

$$\log M = X_A \lambda_A + X_B \lambda_B + X_{AB} \lambda_{AB}, \quad (1)$$

where M is an array of expected frequencies, X_A is the matrix with the indicator variables for the variables in A such that the model is saturated in A , X_B is the matrix with the indicator variables for the variables in B such that the model is saturated in B , and X_{AB} is the matrix with the indicator variables for all first- and higher-order interactions between the variables in A and B (for more detail

on log-linear modeling see, e.g., Agresti, 1996; Bishop, Fienberg, & Holland, 1975; Christensen, 1997). The λ vectors contain the parameters that correspond to the columns of the indicator matrices.

The typical base model for CFA applications with two groups of variables omits the last term in (1). One thus obtains

$$\log M = X_A \lambda_A + X_B \lambda_B, \quad (2)$$

that is, a model that is saturated in the variables in A , saturated in the variables in B , but proposes that there be no interactions among the variables from the two different groups.

The model given in (2) is the CFA base model for both ISA and PCFA. Thus, at the level of base models there seems to be no difference between ISA and PCFA. Indeed, Krauth (1996, p. 138) states for PCFA that the difference between the two groups of variables exists only at the level of substantive interpretation¹. Thus, the base models thus far used for ISA and PCFA are statistically the same. In the present article we propose base models that allow one to discriminate between ISA and PCFA.

In applications of CFA one interprets members of a cell as members of a type if $E(M_i) > E(\hat{m}_i^{(B)})$, where i numbers the cells in the cross-classification of all variables, $E(m_i)$ are the expectancies of the observed cell frequencies, m , and $E(\hat{m}_i^{(B)})$ are the expected cell frequencies, estimated under the base model, B . If $E(M_i) < E(\hat{m}_i^{(B)})$ the cell members belong to an antitype. Using one of the many tests proposed for CFA (for an overview see von Eye, 1990; von Eye & Rovine, 1988) under the appropriate measure for protection of the experiment-wise α , one can come to a statistical decision about the presence of types and antitypes.

2 The Choice of Base Models in CFA and Sampling Schemes

In a recent discussion, Mellenbergh (1996) pointed to the importance of proper selection of a CFA base model. Mellenbergh showed in a data example that the patterns of types and antitypes identified using different base models can dramatically differ, and concluded that “CFA outcomes depend heavily on the null model and researchers must have sound theoretical or empirical arguments ...” (1996, p. 330). In the present section we therefore discuss criteria for the selection of CFA base models. We propose here that, in addition to the criterion proposed by von Eye (1990; von Eye et al., 1996), characteristics of the sampling scheme be considered. Special emphasis is placed on CFA models with two groups of variables. We consider the following criteria:

1. The first criterion is that there be only one way to deviate from the base model for clear-cut interpretation of types and antitypes. For example, first order CFA uses the log-linear main effect model for a base model. This model implies no interactions among variables. Deviations from this model are only possible if there are interactions. Another example is the standard base model for PCFA and ISA (see Equation 2). This model is saturated in each of the two groups of variables. Thus, the only possible deviation from this model is given by interactions (or associations) between predictors and criteria. Later in this article we posit that a more general base model is more appropriate for PCFA. Specifically, we posit that a base model for PCFA leaves room

¹The exact statement is “Der Unterschied zwischen den beiden Teilmengen von Merkmalen besteht nur auf der inhaltlichen oder interpretativen Ebene” (Krauth, 1996, p. 138).

for effects that involve the dependent variables. Substantive considerations determine the type of effects taken into account.

2. The second criterion, introduced in the present article, is that the sampling scheme be considered. Specifically, the multinomial and the product multinomial sampling schemes are of importance in the present context (see Christensen, 1997).
3. *Parsimony* is a third criterion for the selection of base models. More specifically, a base model must take into account interactions of the lowest possible order. We argue later in this article, that the complex standard base model of PCFA which is saturated in the dependent variables, is necessary only in rare cases.

To illustrate the interplay of these criteria consider the following example. If there are two variables with fixed margins and one or more variables with random margins, certain CFA base models that are in accordance with the first criterion are no longer admissible because of the second criterion. In this example, zero order and first order CFA are no longer admissible, because the bivariate margins of the variables with the fixed margins must be part of the base model. Thus, Criterion 2 serves to modify the selection of models based on Criterion 1, with Criterion 2 superseding Criterion 1. Criterion 3 prevents researchers from specifying a base model that is unnecessarily complex.

Section 4 discusses this issue in more detail and illustrates consequences for CFA base model selection. In the following paragraphs we provide a definition of the two sampling schemes considered in this paper (see Christensen, 1997; Jobson, 1992; see also von Eye & Schuster, in preparation).

When researchers perform multinomial sampling they draw a random sample from an infinite population. When creating the cross-classification of variables, individuals are randomly assigned to the cells of the resulting table. Suppose a two-dimensional table is created with R rows and C columns, and $i = 1, \dots, R$ and $j = 1, \dots, C$. Then the joint density of the sample cell frequencies is

$$f(m_{11}, m_{12}, \dots, m_{RC}) = \frac{n!}{\prod_{i=1}^R \prod_{j=1}^C m_{ij}!} \prod_{i=1}^R \prod_{j=1}^C \pi_{ij}^{m_{ij}}, \quad (3)$$

where π_{ij} indicates the probability for Cell ij , $\sum_{i=1}^R \sum_{j=1}^C \pi_{ij} = 1$, and $\sum_{i=1}^R \sum_{j=1}^C m_{ij} = n$. The expectancies of the m_{ij} are $E[m_{ij}] = n\pi_{ij}$. The variances of the m_{ij} are $V[m_{ij}] = n\pi_{ij}(1 - \pi_{ij})$ for $i = 1, \dots, R$ and $j = 1, \dots, C$. The covariances are $Cov[m_{ij}, m_{kl}] = -n\pi_{ij}\pi_{kl}$, for $i \neq k; j \neq l; i, k = 1, \dots, R$; and $j, l = 1, \dots, C$. Because the assignment of cases is to the cells in the entire table, there is no constraint on the marginal sums other than $\sum_i m_{i.} = \sum_j m_{.j} = n$.

The *product multinomial* distribution describes the joint distribution of two or more independent multinomial distributions. Consider an $R \times C$ cross-classification with *fixed row marginals* $m_{i.}$ for $i = 1, \dots, R$. Row marginals are fixed when the number of cases in the rows is determined a priori. In this case individuals in each row are considered members of sub-populations, for instance females and males, or smokers and non-smokers. The joint density of the R rows results from multiplying the row-specific multinomials. In an $R \times C$ table this product is

$$f(m_{11}, m_{12}, \dots, m_{RC}) = \prod_{i=1}^R \left[\frac{m_{i.}!}{\prod_{j=1}^C m_{ij}!} \prod_{j=1}^C \left[\frac{\pi_{ij}}{\pi_{i.}} \right]^{m_{ij}} \right]. \quad (4)$$

Table 1: (2 x 2) x (2 x 2 x 2) Cross-Classification

Variable Group 1	Cell Indices	Variable Group 2								Row Sums
		111	112	121	122	211	212	221	222	
	11	m_{ijklm}								$m_{11\dots}$
	12									
	21									
	22									
Column Sums		$m_{\cdot 111}$	$m_{\cdot 112}$	$m_{\cdot 121}$	$m_{\cdot 122}$	$m_{\cdot 211}$	$m_{\cdot 212}$	$m_{\cdot 221}$	$m_{\cdot 222}$	n

In words, the probability of observing the contingency table with cell frequencies $m_{11}, m_{12}, \dots, m_{RC}$ is given as the product of probabilities for observing each of the R rows or, in other words, as the product of conditional multinomial distributions, given the row totals. This applies accordingly if column marginals are fixed, or if the marginals are fixed for more than one variable.

While the estimation of parameters is the same for these two sampling schemes, type and number of models that can be considered, differ. Consider a research design with two independent classification variables, Drinking (D; yes - no) and Gender (G; female - male), and one dependent variable Liver Cancer (C; shows signs of liver cancer - does not show signs of liver cancer). Together, these three variables form a 2 x 2 x 2 cross-classification. Drinking and Gender are the independent variables, and Liver Cancer is the dependent variable. The margins of the two independent variables are fixed by design². Therefore, any model of these three variables must include a provision to reproduce the Gender - Drinking marginals, $m_{ij\cdot}$. All models that include the (hierarchical) terms for [DG] fulfill this condition. These are the models [DGC]; [DG][DC][GC]; [DG][DC]; [DG][GC]; and [DG][C]. All models without the [DG] term are not admissible. These models include, for instance, the main effect model [D][G][C], and the model [DC][GC].

When there are variables with fixed margins in a design researchers have to make sure that the fixed marginal totals are always reproduced. Margins can be fixed by design or because a variable has the status of an independent variable. Beyond that, the property of marginal frequencies as fixed, does not need to be considered. Estimation is as for multinomial sampling. However, it will be shown in the following sections that when investigating prediction hypotheses using PCFA, the base model can be different than the base model for ISA.

3 Base Models for PCFA

Consider two groups of variables. The first group contains the variables P_1 and P_2 , and the second group contains the variables C_1, C_2 , and C_3 . When analyzing these five variables using PCFA, the following log-linear base model is typically used: $[P_1 P_2][C_1 C_2 C_3]$. This model implies the main effects of all five variables and all group-specific interactions. In addition, this model posits independence among the variables from different groups. To perform a PCFA one typically cross-tabulates all possible patterns (= *configurations*) in one group with all possible configurations in the second group. Table 1 presents the arrangement for the data analysis for a situation where all five variables are dichotomous.

The PCFA base model for designs such as exemplified in Table 1 focuses on interactions in the sense that types and antitypes can emerge only if variables from Group 1 interact with variables from Group 2. Main effects of variables cannot lead

²The variable Drinking is fixed because an a priori determined number of alcohol consumers and resisters was included in the sample.

to types or antitypes because they are taken into account in the base model. In this respect, the ISA model is similar to the base model for first order CFA which also can lead to types and antitypes only if variables interact. The base model for ISA is symmetrical because a change in the order of the variable groups has no effect on the results of analysis.

We now focus on the characteristics of the variables. The variables in Group 1 are predictors, and the variables in Group 2 are criteria. Then it is of importance to specify a base model that reflects the assumptions that guide the researchers' search for types and antitypes. There is a number of assumptions that can meaningfully be entertained. For instance, researchers may proceed under a causal assumption and hypothesize that (1) the first order margins of the criteria, (2) the interactions among the criteria, and (3) the interactions between predictors and criteria are determined by the predictors. This assumption implies a CFA base model that has the following two characteristics:

1. The marginals of the predictors are fixed. Therefore, a base model must include the term $[P_1P_2]$. This term guarantees that cell-wise deviations from the base model are not due to main effects and interactions between the predictors. It is important to emphasize that there are several reasons why marginals of predictors can be fixed. The most important two reasons are that marginals are fixed (1) by design and (2) because of the particular role they play in a study. Variables are fixed by design if researchers determine a priori the number of respondents in a given configuration. Variables are fixed because of the role they play when they are predictors.
2. The criteria typically are free in the sense that neither their main effects nor their interactions are set by design. This specification is of importance when researchers assume that the causal variables, P_1 and P_2 , affect cell frequencies such that not only interactions but also main effects on the criterion side can reflect the causal effects. In other words, the criterion variables may have main effects which show in their first order margins. However, these main effects are not the results of design or sampling but reflect the causal effects.

If researchers consider hypotheses of this type, the PCFA base model cannot include the main effects nor the interactions among the criterion variables, C_1, C_2 , and C_3 . Neither can it include predictor-criterion interactions. Rather, the base model must be based on a mixed sampling scheme where the marginals of the predictors are fixed and the marginals of the criteria are random. The PCFA base model is thus reduced from (2) to

$$\log M = X_A \lambda_A, \quad (5)$$

that is, a model that is saturated in the predictors and contains no effects on the criterion side. If types and antitypes emerge, they reflect the effects that cause the criteria to display (1) main effects, that is, unequal first-order marginal totals, (2) interactions among themselves, and (3) interactions with the predictors. In the following sections we call PCFA models such as given in (5) *mixed-sampling PCFA base models*. Mixed-sampling PCFA base models do allow one to statistically distinguish between predictors and criteria because only predictor main effects and interactions are considered, and it is assumed that types and antitypes reflect the causal effects of the predictors. This applies also to main effects on the criterion side.

It should be noted that the model presented in Equation 5 also implies independence. Specifically, it implies independence (1) between predictors and criteria and (2) among the criteria. In addition, this model implies a uniform distribution for

the criterion variables. It should also be noted, that Equation 5 is but one of many PCFA base models. The selection of a base model for a specific research question must be guided by substantive considerations and meet the above three criteria for CFA base models.

To explain what we mean with the expression *predictors have effects on the main effects of the criteria* we now present an artificial data example. Consider the three variables, Alcohol Diagnosis (A; alcoholic versus not alcoholic), Driving Under the Influence of Alcohol (D; yes vs. no), and was Involved in a Car Accident in the last five Years (C; yes vs. no). A is the predictor, D and C are the criteria. The cross-classification of these three variables and an hypothetical frequency distribution appears in Table 1a, below.

Table 1a: Cross-classification of A, D, and C with hypothetical frequency distribution

Predictor	Criteria		Cell Frequencies
	Driving Drunk	Accident	
Alcohol Diagnosis			
alcoholic	yes	yes	20
alcoholic	yes	no	0
alcoholic	no	yes	0
alcoholic	no	no	10
not alcoholic	yes	yes	10
not alcoholic	yes	no	0
not alcoholic	no	yes	0
not alcoholic	no	no	20

The bivariate interactions $A \times C$ and $D \times C$ are zero. Only the $A \times D$ interactions exists. The univariate marginal totals of all three variables are 30 and 30. However, the marginal totals of neither the Driving Drunk nor the Accident variables are uniform any longer, when Alcohol Diagnosis is considered. More specifically, for the subsample of alcoholics the Driving Drunk marginal totals are 20 and 10, and the Accident marginal totals are also 20 and 10. For the subsample of not alcoholics the marginal totals are both 10 and 20. We thus conclude that the ratio of marginal totals varies across the strata of Alcohol Diagnosis. Application of PCFA yields types for the two cells with 20 observations, and antitypes for the four cells with no observations.

In the context of Prediction CFA, researchers may interpret a relationship of this kind such that the predictors allow one to predict the marginal totals of the criteria. In a context of causal analysis one may consider the main effect in the subtables *caused by* the independent variable. The following sections illustrate and classify the many base models that are possible for PCFA.

3.1 Fixed-Effect Base Models for PCFA

Base models that consider in addition to the effects on the predictor side any of the possible effects of the criterion variables, beginning with the main effects of the criterion variables, will be termed *fixed-effect PCFA base models*. A *first order fixed-effect PCFA base model* is a model that only considers the main effects of the criteria. Consider again the example with the two predictors, P_1 and P_2 , and the three criteria, C_1 , C_2 , and C_3 . The mixed-sampling PCFA base model for these variables is $[P_1 P_2]$. The first order fixed-effect PCFA base model for these variables is $[P_1 P_2][C_1][C_2][C_3]$. Notice that this is still not the standard PCFA model, which is, as was indicated above, $[P_1 P_2][C_1 C_2 C_3]$. Using the present terminology, this model can be termed a third order fixed-effect PCFA model. Between the last two models lies the second order fixed-effect PCFA model $[P_1 P_2][C_1 C_2][C_1 C_3][C_2 C_3]$.

Each of these models has its unique interpretation and may be useful in specific applications. The first order fixed-effect PCFA base model, $[P_1 P_2][C_1][C_2][C_3]$, is of interest when the number of cases per category of the criterion variables is not fixed, but known to deviate from a uniform distribution. Under this condition causal effects may not extend to the main effects of the criterion variables. Only first and higher order interactions among the criterion variables and the interactions among criterion and predictor variables can reflect the causal effects. Accordingly, application of the second order fixed-effect PCFA model, $[P_1 P_2][C_1 C_2][C_1 C_3][C_2 C_3]$, implies that both main effects and first order interactions in the criteria are known to exist, based on prior knowledge that indicates interactions among the criteria that are not caused by the predictors. Thus, the predictor effects can manifest only in the third order interaction, $[C_1 C_2 C_3]$, among the criterion variables or interactions of any order between predictor and criterion variables.

It thus becomes obvious that the standard model of PCFA, in the example $[P_1 P_2][C_1 C_2 C_3]$, has the characteristic that it restricts the manifestation of the causal effects of the predictors to the predictor-criterion interactions. In the following sections we present two data examples.

3.2 A Data Example

The PCFA model given in Equation 5 and the example in Section 3 are extreme in the sense that not a single effect of the criterion variables is part of the model. This may seem implausible because researchers typically assume that main effects that are caused by independent variables manifest only in variable interactions. The following hypothetical example illustrates that this is not necessarily the case. Consider an evaluation experiment in which researchers compare two training methods. A sample of $N = 22$ participants was randomly drawn from a pool of athletes who still had to take the qualification test for the next intergalactical games. 11 randomly selected athletes went to the camp where training method A was used, the remaining 11 went to the camp where training method B was used. After completion of the training program, all athletes took the qualification test. Only one athlete from each camp failed the test, all others passed.

We now analyze these frequencies using two PCFA base models. The first base model is that of standard PCFA which, in this case, is identical to the base model for first order CFA. This model takes the main effects of both the predictor and the criterion into account. Notice that this is also the model for the Pearson X^2 -test of variable independence. The second base model employed for analysis of these data is that of mixed-sampling PCFA which considers only the predictor effects, thus hypothesizing a joint frequency distribution that is fully determined by the predictor in this cross-classification, that is, Training Program (P with categories A and B). The Result of Test variable (T with categories pass and fail) is supposed to have no effect whatsoever. This model can yield types and antitypes if (1) the criterion has main effects, and/or (2) there exist predictor-criterion interactions. Results of these two analyses are summarized in Table 2. We used the Anscombe z -test and Bonferroni-adjustment of α which led to $\alpha^* = 0.0125$.

The mixed-sampling PCFA yields two antitypes, whereas standard fixed-effect PCFA indicates neither types nor antitypes. In fact, the goodness-of-fit Pearson $X^2 = 0.0$ indicates for this artificial data example that there is no association between Training Program and Success in the qualification test. In other words, the model of variable independence applies. In contrast, the mixed-sampling PCFA suggests that it is less likely than expected from the model that assumes no main effects for the Result of Test variable that participants in either training program fail the test. The goodness-of-fit for this base model is poor, as the Pearson $X^2 = 14.72$ ($df = 2$; $p(X^2) = 0.006$) indicates. This discrepancy between observed and estimated

Table 2: Analysis of Training Program Data Using Two Base Models of PCFA

Cells	Observed frequencies	Fixed-effect PCFA			Mixed-sampling PCFA		
		Exp. freq.	z	p(z)	Exp. Freq.	z	p(z)
Ap	10	10	0	.5	5.5	1.79	.036
Af	1	1	0	.5	5.5	2.32	.010 A
Bp	10	10	0	.5	5.5	1.79	.036
Bf	1	1	0	.5	5.5	2.32	.010 A

expected cell frequencies is solely due to the main effect of the criterion variable. This main effect is considered as caused by the independent variable, as is explained in the following paragraph.

To correctly interpret and apply this last statement one has to consider the assumptions made when applying the base model of mixed-sampling CFA. The model posits that the entire variability in a table can be accounted for by only considering the main effects and interactions of a subset of variables, that is, the predictors. Otherwise, and because there is no additional information that is being used, there is the assumption of a uniform frequency distribution. For the data example in Table 2 this assumption translates into a uniform distribution for the entire table because the experiment used the same number of athletes for each training program. This is equivalent to the assumption that, without the training, the selected athletes have a 50% chance of qualifying for the intergalactic games. Researchers who are unable or not willing to make this assumption may possess information that can be made part of the CFA base model (see Gutiérrez-Peña & von Eye, in preparation). If, for example, the chances of qualifying differ from 50%, one must take this information into account. Failure to take into account this information can render the types and antitypes detected by CFA invalid. In addition, a control group of athletes who do not participate in the training programs, may be considered for valid conclusions.

Thus, this example shows that the extreme base model of mixed-sampling PCFA in which no effects of the criterion are taken into account, can lead to meaningful results if the hidden assumption that the marginal frequencies on the criterion side are uniformly distributed can be defended. If this assumption cannot be defended, we recommend using the fixed-effect PCFA base model that is saturated in the predictors and considers the main effects of the criteria. Alternatively, Bayesian approaches to CFA (Wood, Sher, & von Eye, 1994; Gutiérrez-Peña & von Eye, in preparation) and parametric CFA (Spiel & von Eye, 1993) may be considered. More complex models may be needed depending on substantive assumptions.

4 Selection of Models

In this section we present an overview of possible CFA base models for cross-classifications with up to five variables. Two types of base models are distinguished. The first comprises global CFA base models, that is, models where all variables have the same status. The second type comprises two-group models. Among the two-group models we distinguish further between models of the PCFA type, that is models with predictors and criteria.

When discussing PCFA and selecting a base model for PCFA there are three aspects that need to be considered. The first aspect concerns the model for the predictors, the second aspect concerns the model for the criteria, and the third aspect concerns the model for the dependence of the criteria on the predictors. This dependence materializes typically in predictor-criterion interactions.

In the past, PCFA involved fitting a saturated model to the predictor variables even if their margins were random. There are two main reasons for this strategy. The first is that PCFA, in analogy to usual regression analysis, models the conditional distribution of the criteria, given the predictors. By fitting a saturated model to the predictors we treat them in the statistical analysis as fixed even if they are random. The second reason is that if the base model for the predictors is not saturated, types and antitypes can result that reflect predictor interactions. This would be counter to the concept of PCFA. Therefore, we retain the strategy of fitting a saturated base model to the predictors.

In contrast, we do not retain the strategy of also always fitting a saturated base model for the criterion variables. Although relationships between the criterion variables may be of secondary interest we do not see a theoretical argument for always assuming a saturated base model on the criterion side. A saturated model is trivially “true.” Therefore, there is no way to avoid a mis-specification of the criterion base model. In addition, one cannot always exclude the possibility that criterion interactions are caused by the predictors. These interactions will go unnoticed when the base model for the predictors is saturated.

Another important issue must be considered. Suppose a model that is more parsimonious than the saturated model fits the criteria. Then, the types and antitypes identified by PCFA can be different than the types and antitypes identified when the criterion base model is saturated. Any discrepancies between the fitting model and the saturated model may reflect predictor effects.

It is important to note that if the predictor and criterion variables are not connected in every base model, the joint distribution of all variables is collapsible onto the criterion variables (for a discussion of collapsibility see Asmussen and Edwards, 1983). Hence, we can consider a reasonable model for the criterion variables as a preliminary step to PCFA. After a predictor variable model is selected (typically the saturated model; see above), one can perform PCFA as usual. However, the fitting model for the criteria is substituted for the standard saturated model.

Table 3 presents an overview of base models for CFA and PCFA. Possible models are marked with a “x” and models that are excluded because of the variables sampling characteristics or status are marked with a “-”. The first column in Table 3 lists the number of variables, out of 5, with fixed margins. The middle panel of Table 5 displays the order of global CFA base models. This order is determined by the level of effects taken into account when estimating expected cell frequencies (von Eye, 1990). For instance, zero order CFA, also called Configural Cluster Analysis (Lienert & von Eye, 1985) assumes in its base model a uniform distribution, and First Order CFA takes all main effects into account.

The right hand panel displays the order of CFA base models for the criterion variables in PCFA. For this panel it is assumed that all predictors have random margins. The order of base models for the predictors does not need to be displayed, because it is always the same as the number of variables if all predictors have fixed marginals, which is typically the case. The last line of the right hand panel displays n/a for each cell, because there can be no CFA of two groups of variables when all variables are predictors with fixed margins. The table does not discriminate between models that posit different relationships between predictors and criteria. We do believe it is worth discussing models where certain predictor-criterion relationships are part of some base model. However this discussion is beyond the scope of the present article and will be taken up at another occasion.

Table 3 suggests that global base models must be of increasingly higher order when the number of variables with fixed margins increases. The last row in the table shows that CFA cannot be performed when all variables have fixed margins because, as was explained before, base models must be saturated in variables with fixed margins.

Table 3: Possible CFA Base Models for up to Five Variables in the Cross-Classification

# of Variables with fixed margins	Order of Global CFA Base Model					Order of Base Models for Criterion Variables in PCFA-type Models			
	0	1	2	3	4	0	1	2	3
0	x	x	x	x	x	x	x	x	x
1	-	x	x	x	x	-	x	x	x
2	-	-	x	x	x	-	-	x	x
3	-	-	-	x	x	-	-	-	x
4	-	-	-	-	x	-	-	-	-
5	-	-	-	-	-	n/a ^a	n/a	n/a	n/a

^an/a indicates that a model is not conceivable because there are no criterion variables left.

Table 4: Sample base models for Global CFA with varying numbers of fixed variables

Global CFA Base Model	Model Specification in Bracket Notation
Zero order Global Model	no effects, no fixed margins
First Order Global Model	[P1][P2][C1][C2][C3]
Second Order Global Model	[P1, P2][C1, C2][P1, C1][P1, C2][P2, C1][P2, C2]
Third Order Global Model	[P1, P2][C1, C2, C3]
Fourth Order Global Model	[P1][C1, C2, C3, C4]
Fifth Order Global Model	[C1, C2, C3, C4, C5] (Saturated)

To illustrate the use of Table 3 we now list sample CFA base models. The listed models are not necessarily the only possible base models. However, they can be used to derive other possible models. Table 4 contains sample base models for the left panel of Table 3. The variables used in the examples are the Predictors, P1 and P2, and the Criteria, C1, C2, C3, C4 and C5. Table 5 displays base models for PCFA.

5 Data Example

In this paragraph we re-analyze a data set that has been repeatedly used for illustration of PCFA and its characteristics (Lienert, 1978; Mellenbergh, 1996; von Eye et al. 1996). The data describe suicide attempts in a sample of $n = 482$ individuals. The following three categorical variables are analyzed: Gender (G; 1 = male, 2 = female); Motive for Suicide (M; 1 = illness; 2 = psychiatric disorder; 3 =

Table 5: Sample base models for PCFA with varying numbers of fixed criteria

Base Model for PCFA Criterion Variables	Model Specification in Bracket Notation
Zero Order Fixed-Effect PCFA Base Model	[P1][P2]
First Order Fixed-Effect PCFA Base Model	[P1][P2][C1][C2][C3]
Second Order Fixed-Effect PCFA Base Model	[P1, P2][C1, C2][C1, C3][C2, C3]
Third Order Fixed-Effect PCFA Base Model	[P1][P2][C1, C2, C3]
Fourth Order Fixed-Effect PCFA Base Model	[P1][C1, C2, C3, C4]
Fifth Order Fixed-Effect PCFA Base Model	n/a = not applicable because there are no criterion variables left

Table 6: Observed Cell Frequencies of Suicide Attempt Data

Cell Indices	Observed Frequencies for Outcome of Suicide Attempt	
GM	Survived	Died
11	64	18
12	76	47
13	7	8
21	86	16
22	61	25
23	47	27

alcoholism); and Outcome of Suicide Attempt (O; 1 = survived; 2 = dead). Table 6 displays the observed cell frequencies for the 2 x 3 x 2 cross-classification of these three variables.

For the following analyses we assume that G and M are the predictors and O is the criterion variable. To illustrate various assumptions that can be investigated in regard to a possibly emerging pattern of types and antitypes, we apply the following CFA base models:

1. **First order CFA [G][M][O]**. This model is included here only to provide readers with results from classical first order CFA. A comparison with results from PCFA is not possible because the model [G][M][O] does not discriminate between predictors and criteria. This model of variable independence is of interest when researchers are interested in types and antitypes that can be traced back to local interactions of variables but not to main effects. Only one variable can be fixed by design or sampling characteristics. Types and antitypes can result from any interaction in these three variables, but not from main effects.
2. **Mixed-sampling PCFA [GM]**. This model considers the margins of the predictor variables, Gender and Motive as fixed. The margins of the outcome variable, Outcome of Suicide Attempt, are considered random. The predictors may interact because it seems plausible that males and females differ in motive for suicide (Chipuer & von Eye, 1989). This model is, on the criterion side, a zero order CFA. Types and antitypes can result from (1) the main effect of the criterion, and (2) interactions among the predictors and the criterion.
3. **First order fixed-effect PCFA [GM][O]**. In the present example, this is the standard PCFA (and ISA) base model. It considers both the margins of the predictor variables and the margins of the criterion as fixed. Types and antitypes can only result from interactions between the predictors and the criterion.

Table 7 displays the results from these three CFA base models. For each model we adjusted $\alpha = 0.05$ using the Bonferroni method which resulted in an adjusted $\alpha^* = 0.0042$. The table presents the deviance residuals, z_{ijk} , and their one-sided tail probabilities.

Table 7 suggests that first order CFA identifies two types and two antitypes. The first type is constituted by Configuration 122. These are men that suffer from some psychiatric disorder and succeed in committing suicide. The second type is constituted by Configuration 232. These are alcoholic women that succeed in committing suicide. Each of these two patterns occurs more often than expected from the base model of variable independence. The first antitype, constituted by Configuration

Table 7: z-values and their Tail Probabilities for Three CFA Models for Suicide Attempt Data

Cell Indices GMO	m_{ijk}	[G][M][O]		[GM]		[GM][O]	
		z_{ijk}	$p(z_{ijk})^a$	z_{ijk}	$p(z_{ijk})^a$	z_{ijk}	$p(z_{ijk})^a$
111	64	.587	.279	3.150	.0008 T	.773	.220
112	18	-1.392	.082	-4.182	< α^* A	-1.280	.100
121	76	1.015	.155	3.769	< α^* T	-1.208	.114
122	47	3.289	.0005 T	-.101	.460	1.753	.040
131	7	-4.869	< α^* A	-3.422	.0003 A	-1.183	.118
132	8	-1.198	.116	-3.117	.0009 A	1.544	.061
211	86	1.752	.040	4.612	< α^* T	1.581	.057
212	16	-2.864	.002 A	-5.617	< α^* A	-2.781	.002 A
221	61	-2.258	.012	.550	.291	.020	.492
222	25	-1.494	.068	-4.751	< α^* A	-.031	.488
231	47	2.065	.020	4.101	< α^* T	-.753	.226
232	27	3.031	.001 T	.561	.287	1.107	.134

^aTail probabilities with four or more leading zeros are displayed as “< α^* ”; types are indicated by **T**; antitypes are indicated by **A**.

131, describes alcoholic men that fail to succeed in committing suicide. The second antitype, constituted by Configuration 212, describes women that suffer from some illness and succeed in committing suicide. Each of these two patterns occurs less often than expected from the base model.

When describing and interpreting these types and antitypes it is important to stay at a purely descriptive level. The CFA base model does not allow one to discriminate between predictors and criteria, nor does it support notions of causality (see von Eye & Brandtstädter, 1998). The sampling scheme that underlies this analysis is either multinomial or product multinomial, with no effect on the magnitude of the estimated expected cell frequencies or the model parameters (which are typically not of interest in CFA applications). In contrast, the second base model, [GM] does support the distinction between predictors and criteria. This model assumes that Gender and Motive for suicide attempt allow one to predict the outcome of a suicide attempt. In addition, the model [GM] is based on a mixed sampling scheme where the uni- and bivariate marginals of G and M are fixed and the marginals of O are random. As was indicated in Section 3, not considering the main effect of the criterion implies the assumption of a uniform marginal distribution for the criterion. If this assumption is of no substantive interest, criterion marginals must be made part of the base model (see below).

The model [GM] identifies four types and five antitypes. These types and antitypes overlap with the ones identified by first order standard CFA only in part. Specifically, the two types found by first order CFA are not found here.

When interpreting the results from the model [GM] one has to take into account the nature of the base model which treats G and M as predictors and O as the outcome variable. For instance, the pattern “male, suffering from some illness” allows one to predict that a suicide attempt is survived (Configuration 111). This pattern constitutes a type, indicating that it occurs more often than expected from the assumption that G and M are not predictors of (1) main effects of O and (2) interactions between G and M on the one hand side and O on the other. Whether (1) or (2) is the case cannot be determined from the information provided by CFA. Ac-

cordingly, the first antitype suggests that it occurs less often than expected from the base model that men that suffer from some illness succeed in committing suicide³.

Rather than interpreting the three types and four antitypes in detail we now inspect the results from the third model, [GM][O]. This model can be interpreted in two ways. It is either a standard PCFA, or it is a first order fixed-effect PCFA where researchers posit that the effects of the predictors manifest only in interactions between the two predictors and the criterion, and the marginals of the criterion are fixed by design or sampling characteristics. In either case, this model takes more information into account than the other two models when estimating the expected cell frequencies. Therefore, there are fewer ways to deviate from the expected cell frequencies and it does not come as a surprise that the number of types and antitypes is smaller than for the other two models. (It should be noted, however that this is not necessarily always the case.) The model identifies no type and only one antitype. It is Configuration 212 which had emerged as an antitype in the other two analyses also.

In sum, this example illustrates again that Mellenbergh's (1996) statement is correct. Different models of CFA can lead to dramatically different patterns of types and antitypes.

6 Discussion

It was the goal of this article to show that in CFA applications the selection of an appropriate base model is crucial for proper interpretation of results. It was shown that (1) different base models can lead to the detection of different patterns of types and antitypes, and (2) the sampling scheme that underlies a research design determines, in part, what base model can be considered. More specifically, the distinction of the multinomial and product multinomial sampling schemes was used to show that

1. Whenever a product multinomial sampling scheme is used, the uni- and multivariate marginals that are fixed must be considered. Therefore, zero-order CFA which compares the observed cell frequencies to an expected uniform distribution, can be employed in a product multinomial sampling design only if the design is balanced in the product multinomial variables. This applies accordingly to first and higher order CFA.
2. When a multinomial sampling scheme is used, the effects considered in the base model depend only on the substantive hypotheses. There are no constraints that the sampling scheme poses on effects.
3. When a mixed sampling scheme is used, the above arguments apply to the variables depending on what sampling was used.

We showed that PCFA can be applied in particular to data that were collected under sampling schemes where the predictors are fixed. Under the assumption that the predictor effects can manifest in (1) criteria main effects, (2) interactions among the criteria, and (3) predictor-criterion interactions, a problem was solved that was typical of the often unreflected standard application of CFA models (Krauth, 1996), that is, the problem that the base models of ISA and PCFA are the same although interpretation is very different at the substantive level. The present approach leads

³One might suspect that when the outcome variable is dichotomous types and antitypes always go hand in hand. This is obviously not the case. For instance, Pattern 121 constitutes a type, but Pattern 122 is far from constituting an antitype. Another example is the pattern pair 131 and 132 which both constitute antitypes.

to base models that are different for ISA and PCFA. As was shown using the suicide data example, there is only one instance where ISA and PCFA base models can be the same.

More specifically, only if a base model is saturated in both the predictors and the criteria, the distinction between PCFA and ISA disappears. Base models that (1) meet the three criteria put forth in Section 2 and (2) are more parsimonious on the criterion side are PCFA models and thus differ from ISA models. The distinction between base models for ISA and PCFA has important consequences. Two consequences will be discussed here. First, researchers now can statistically test assumptions concerning the causal or predictive relationships between variable groups. The types and antitypes identified by PCFA can dramatically differ depending on whether the standard ISA model or some custom PCFA model was specified. This was illustrated in Section 3 of this article.

It may be important to realize that, from a log-linear modeling perspective, it is trivial to note that residuals are confounded with models. In the context of CFA, however, the selection of base models reflects two important considerations, both of which are crucial for interpretability and meaningfulness of identified types and antitypes. The first consideration is the aim of the study. Depending on aim, researchers select a CFA model, for example, PCFA instead of ISA. The second consideration concerns the status of variables as fixed or random, as predictors or criteria etc. The present article presented rules that can be used to specify a particular CFA base model. Thus, emerging types and antitypes are not just confounded with a CFA model, but with a model that reflects some scientific theory and thinking. From this perspective, the present article removes some of the arbitrariness from the process of selecting a CFA base model.

A second consequence is that researchers have to match PCFA base model and substantive hypotheses. It is not appropriate to simply use the base model that is saturated on both the predictor and the criterion sides. One consequence of selecting an unnecessarily complex model on the criterion side is that types and antitypes can remain undetected and thus, justice is not done to the predictive or causal characteristics of the variables on the predictor side.

The flexibility of base models for PCFA discussed here applies also to other CFA models that distinguish groups of variables. One such model is 2- or more-sample CFA. This method compares two or more pre-existing groups of cases using a number of discriminant variables. Differences manifest in form of discriminant types (and antitypes, in three or more group comparisons). In many applications, the pre-existing groups are fixed in number, that is, by design. Therefore, discriminant CFA can also fruitfully be viewed as an approach for which mixed-sampling base models are most appropriate. That is, the uni- and multivariate marginals of the grouping variables must be reproduced by design. In contrast, the uni- and multivariate marginals of the discriminant variables must be only reproduced if required by some substantive hypothesis.

It is important to realize that the first criterion of selecting base models is fulfilled in mixed-sampling PCFA also. The first criterion states that CFA base models must leave only one option for deviations open. In mixed-sampling PCFA this option involves effects of the dependent variables. Consider the case where only predictor effects are part of the base model (cf. von Eye, 1985). This base model can be contradicted and thus lead to types and antitypes only if there are effects on the side of the dependent variables. Such effects include main effects, interactions among the dependent variables, and interactions among the dependent and the independent variables. This applies accordingly to fixed-effect PCFA.

The base models that can be considered for PCFA range from a no effect base model, also called zero-order CFA base model on the criterion side to a saturated model on the criterion side (see Table 3). These two base models can be viewed

as the poles of a continuum of complexity. The base model that is saturated in the criteria is the least parsimonious and requires special justification. The same applies to the most parsimonious model that is the model of no effects or zero order CFA on the criterion side (see Equation 5). A justification must (1) be grounded in substantive considerations and (2) conform to the three criteria listed in Section 2 of this article.

Future research needs to explore the consequences of considering sampling schemes in both more (1) depth and (2) breadth. First, one can ask whether other sampling schemes than the ones considered here also have consequences for the specification of CFA base models. Examples of such sampling schemes include but are not limited to *hypergeometric sampling*, *stratification sampling*, *cluster sampling*, and *complex sampling* which includes both stratification and cluster sampling. Second, one can ask whether the existing classification of CFA base models into global and regional models can be fruitfully expanded to also accommodate the sampling schemes and their consequences. In addition, one needs to determine the consequences of the present discussion for the recently introduced Bayesian CFA models (Wood et al., 1994; Gutiérrez-Peña & von Eye, in preparation).

From a more general research strategy perspective, this new criterion of considering sampling schemes (cf. von Eye & Schuster, in preparation) sheds new light on the characteristics of CFA as an exploratory approach, and on exploratory research in general. CFA, while in principle useable in explanatory research, is typically applied in exploratory contexts. Most researchers consider CFA a method that is largely free of assumptions that need to be made for proper application. However, even exploratory methods require specific conditions to be met. The present paper suggests that exploratory application of CFA cannot regress to blind application of base models that are unrelated to the substantive conceptions of the status of variables as predictors and criteria. The sampling and design characteristics of the variables must be taken into account. The data example in Section 4 illustrates how far results can be from meaningfully interpretable if these data characteristics are ignored.

In the development of CFA as a statistical method, the “borrowing from log-linear modeling” was routine. This practice went so far that defensively formulated articles were deemed in order that defended CFA as a method that allows on to answer questions that cannot be answered using log-linear modeling (see, for example, Lehmacher, 1984, or the debate between Langeheine, 1980, and Krauth, 1980). Recently, the picture has changed. CFA has become more and more independent of log-linear modeling and variants of CFA have been proposed that are quite different than standard log-linear residual analysis (Gutiérrez-Peña & von Eye, in preparation) or use fewer and fewer elements of log-linear modeling (von Eye, Spiel, & Rovine, 1995).

The present article presents a third element in this development. This article introduces concepts that are of relevance for the selection of base models in CFA. These concepts are well-known in the log-linear literature (Christensen, 1997). However, these concepts have, thus far, been chiefly discussed in regard to their implications for the estimation of parameters and odds ratios. After stating that the estimation of neither parameters nor odds ratios and their significance tests are not affected, most authors move on and talk about different topics. We believe, however, that the results presented in this article have implications for model selection in log-linear modeling as well. Specifically, when models are selected in which predictor-criteria relationships are modeled and/or some of the margins are fixed, the selection of the proper log-linear model is of colossal importance for the interpretation of results. While this is a topic for another discussion, it can be stated at this point that in log-linear modeling as well as in CFA, the selection of models is restricted depending on the nature of variables.

The present article discussed the selection of base models from the perspective of status of variables in PCFA. The main criteria for model specification were the status of variables as fixed versus random and as predictors versus criteria. The result of the application of the rules that were formulated was a base model that then was subjected to standard CFA routines. Yet, there are more criteria that can be used to select a base model. For instance, one can ask whether for the random variables on the criterion side, more parsimonious models than the saturated one can be considered, and how to identify these models. Table 3 indicates already that more parsimonious models may be possible. This applies accordingly when the predictors are random and the criteria are fixed, as is the case in Discriminant CFA (DCFA; see von Eye, Schuster, & Gutiérrez-Peña, in preparation). One can try to fit log-linear models to either only the predictors (PCFA) or only the criteria (DCFA) and thus arrive at a more parsimonious base model. The gain of this approach will be twofold. First, there will be more statistical power. Thus, it is more likely that types and antitypes will be detected. Second, and for the interpretation of types and antitypes equally important, these parsimonious models will show what main effects and interactions do versus do not exist in the criteria when the table is collapsed with respect to the predictors (criteria) thus creating a natural basis for interpretation. Future work will have to develop this argument in more detail (Schuster & von Eye, in preparation).

The present article focused on the model of Prediction CFA. PCFA is only one of a number of regional CFA models. These are models where variables can have a different status such as, for instance, predictors and criteria. Other examples of regional CFA models include Discrimination CFA where variables can be fixed on the predictor side. The consequences that different sampling schemes have for Discrimination CFA are not exactly the same as the consequences for PCFA. Therefore, a parallel paper is being prepared where sampling schemes are discussed for Discrimination CFA (von Eye, Schuster, & Gutiérrez-Peña, in preparation). Later works will have to address the issue of sampling schemes and their consequences for all of CFA.

References

- [1] Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
- [2] Asmussen, S., & Edwards, D.G. (1983). Collapsibility and response variables in contingency tables. *Biometrika*, *70*, 567 - 578.
- [3] Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis. Theory and practice*. Cambridge: MIT Press.
- [4] Chipuer, H. M., & von Eye, A. (1989). Suicide trends in Canada and Germany: An application of Configural Frequency Analysis. *Suicide and Life-Threatening Behavior*, *19*, 264- 276.
- [5] Christensen, R. (1997). *Log-linear models and logistic regression*, 2nd edition. New York: Springer.
- [6] Gutiérrez Peña, E., & von Eye, A. (in preparation). *A Bayesian approach to Configural Frequency Analysis*.
- [7] Jobson, J.D. (1992). *Applied multivariate data analysis. Volume II: Categorical and multivariate methods*. New York: Springer.
- [8] Krauth, J. (1980). Ein Vergleich der Konfigurationsfrequenzanalyse mit der Methode der log-linearen Modelle [A comparison of Configural Frequency Analysis with the method of log- linear models]. *Zeitschrift für Sozialpsychologie*, *11*, 233 - 247.
- [9] Krauth, J. (1996). *Einführung in die Konfigurationsfrequenzanalyse (KFA) [Introduction to Configural Frequency Analysis (CFA)]*. Weinheim: Beltz.

- [10] Krauth, J., & Lienert, G.A. (1974). Zum Nachweis syndromgenerierender Symptominteraktionen in mehrdimensionalen Kontingenztafeln (Interaktionsstrukturanalyse) [The detection of syndrome-generating interactions in multi-dimensional contingency tables]. *Biometrische Zeitschrift*, 16, 203 - 211.
- [11] Langeheine, R. (1980). Multivariate Hypothesentestung bei qualitativen Daten. *Zeitschrift für Sozialpsychologie*, 11, 140 - 151.
- [12] Lehmacher, W. (1980). Die Beziehung der Konfigurationsfrequenzanalyse zu anderen Methoden der Analyse qualitativer Daten [The relationship of CFA with other methods of analysis of categorical data]. In H. Enke, & J. Haerting (Eds.), *Biometrie and Biostatistik in der Medizin und verwandten Gebieten [Biometrics and Bio-statistics in medicine and related disciplines]* (pp. 83 - 91) Halle: Martin Luther Universität, Wissenschaftliche Beiträge.
- [13] Lienert, G.A. (1969). Die Konfigurationsfrequenzanalyse als Klassifikationsmethode in der Klinischen Psychologie [Configural Frequency Analysis as a method for classification in Clinical Psychology]. In M. Irle (Ed.), *Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie* (pp. 244 - 253) [Proceedings of the 26th congress of the German Psychological Association]. Göttingen: Hogrefe.
- [14] Lienert, G.A. (1971). Die Konfigurationsfrequenzanalyse III. Zwei- und Mehrstichproben-KFA in Diagnostik und Differentialdiagnostik [Configural Frequency Analysis III. Two- and more-sample CFA in diagnostics and differential diagnostics.] *Zeitschrift für Klinische Psychologie und Psychotherapie*, 19, 291 - 300.
- [15] Lienert, G.A. (1978). *Verteilungsfreie Methoden in der Biostatistik, Band II [Distribution-free methods, Vol. II]*. Meisenheim am Glan: Hain.
- [16] Lienert, G.A., & Krauth, J. (1973). Die Konfigurationsfrequenzanalyse als Prädiktionsmodell in der angewandten Psychologie [Configural Frequency Analysis as a prediction model in applied psychology.] In H. Eckensberger (ed.), *Bericht über den 28. Kongreß der Deutschen Gesellschaft für Psychologie in Saarbrücken 1972 [Proceedings of the 28th congress of the German psychological association in Saarbrücken 1972]* (pp. 219 - 228). Göttingen: Hogrefe.
- [17] Lienert, G. A., & von Eye, A. (1985). Die Konfigurationsclusteranalyse (KCA) und ihre Anwendung in der klinischen Psychologie [Configural Cluster Analysis (CCA) and its application in clinical psychology.]. In D. Albert (Ed.), *Bericht über den 34. Kongreß der Deutschen Gesellschaft für Psychologie 1984 in Wien [Proceedings of the 34th congress of the German psychological association in Vienna, 1984]* (pp. 167-169). Göttingen: Hogrefe.
- [18] Mellenbergh, G.J. (1996). Other null model, other (anti)type. *Applied Psychology: An International Review*, 45, 329 - 330.
- [19] Schuster, C., & von Eye, A. (1999). *Modeling turn-over tables using the GSK approach.*⁴
- [20] Spiel, C., & von Eye, A. (1993). Configural frequency analysis as a parametric method for the search for types and antitypes. *Biometrical Journal*, 35, 151-164.
- [21] von Eye, A. (1985). Konfigurationsfrequenzanalyse bei gerichteten Variablenbeziehungen (GKFA) [Configural Frequency Analysis for directed variable relationships]. *EDV in Medizin und Biologie*, 16, 37 - 51.
- [22] von Eye, A. (1990). *Introduction to Configural Frequency Analysis – The search for types and antitypes in cross-classifications*. Cambridge, UK: Cambridge University Press.
- [23] von Eye, A., & Brandtstädter, J. (1998). The Wedge, the Fork, and the Chain - Modeling dependency concepts using manifest categorical variables. *Psychological Methods*, 3, 169 - 185.
- [24] von Eye, A., & Rovine, M. J. (1988). A comparison of significance tests for Configural Frequency Analysis. *EDP in Medicine and Biology*, 19, 6-13.

⁴<http://www.pabst-publishers.de/mpr/issue5/art4/>

-
- [25] von Eye, A., & Schuster, C. (in preparation). *Testing causality assumptions under multinomial and product multinomial sampling schemes.*
- [26] von Eye, A., Schuster, C., & Gutiérrez Peña, E. (in preparation). *Configural Frequency Analysis under retrospective and prospective sampling schemes – frequentist and Bayesian approaches.*
- [27] von Eye, A., Spiel, C., & Rovine, M. J. (1995). Concepts of nonindependence in Configural Frequency Analysis. *Journal of Mathematical Sociology*, 20, 41 - 54.
- [28] von Eye, A., Spiel, C., & Wood, P.K. (1996). Configural Frequency Analysis in Applied Psychological Research. *Applied Psychology: An International Review*, 45, 301 - 327.
- [29] Wood, P. K., Sher, K., & von Eye, A. (1994). Conjugate methods in Configural Frequency Analysis. *Biometrical Journal*, 36, 387 - 410.