

## Power and Sample Size Computation for Wald Tests in Latent Class Models

Dereje W. Gudicha

Tilburg University, The Netherlands

Fetene B. Tekle

Janssen Pharmaceutica, Belgium

Jeroen K. Vermunt

Tilburg University, The Netherlands

**Abstract:** Latent class (LC) analysis is used by social, behavioral, and medical science researchers among others as a tool for clustering (or unsupervised classification) with categorical response variables, for analyzing the agreement between multiple raters, for evaluating the sensitivity and specificity of diagnostic tests in the absence of a gold standard, and for modeling heterogeneity in developmental trajectories. Despite the increased popularity of LC analysis, little is known about statistical power and required sample size in LC modeling. This paper shows how to perform power and sample size computations in LC models using Wald tests for the parameters describing association between the categorical latent variable and the response variables. Moreover, the design factors affecting the statistical power of these Wald tests are studied. More specifically, we show how design factors which are specific for LC analysis, such as the number of classes, the class proportions, and the number of response variables, affect the information matrix. The proposed power computation approach is illustrated using realistic scenarios for the design factors. A simulation study conducted to assess the performance of the proposed power analysis procedure shows that it performs well in all situations one may encounter in practice.

**Keywords:** Latent class models; Sample size; Statistical power; Information matrix; Wald test; Design factor.

---

This work is part of research project 406-11-039 “Power analysis for simple and complex mixture models” financed by the Netherlands Organization for Scientific Research (NWO).

Corresponding Author’s Address: D.W. Gudicha, P.O.Box 90153, 5000LE Tilburg, Department of Methodology and Statistics, Tilburg University, The Netherlands, email: [D.W.Gudicha@uvt.nl](mailto:D.W.Gudicha@uvt.nl).

Published online: 2 May 2016

## 1. Introduction

Latent class (LC) analysis was initially introduced in the 1950s by Lazarsfeld (1950) as a tool for identifying subgroups of individuals giving similar responses to sets of dichotomous attitude questions. It took another two decades until LC analysis started attracting the attention of other statisticians. Since then, various important extensions of the original LC model have been proposed, such as models for polytomous responses, models with covariates, models with multiple latent variables, and models with parameter constraints (Goodman 1974; Dayton and Macready 1976; Formann 1982; McCutcheon 1987; Dayton and Macready 1988; Vermunt 1996; Magidson and Vermunt 2004). More recently, statistical software for LC analysis has become generally available—e.g., Latent GOLD (Vermunt and Magidson 2013b), Mplus (Muthén and Muthén 2012), LEM (Vermunt 1997), the SAS routine PROC LCA (Lanza, Collins, Lemmon, and Schafer 2007), and the R package poLCA (Linzer and Lewis 2011)—which has contributed to the increased popularity of this model among applied researchers. Applications of LC analysis include building typologies of respondents based on social survey data (McCutcheon 1987), identifying subgroups based on health risk behaviors (Collins and Lanza 2010), identifying phenotypes of stalking victimization (Hirtenlehner, Starzer, and Weber 2012), and finding symptom subtypes of clinically diagnosed disorders (Keel et al. 2004). Applications which are specific for medical research include the estimation of the sensitivity and specificity of diagnostic tests in the absence of a gold standard (Rindskopf and Rindskopf 1986; Yang and Becker 1997) and the analysis of the agreement between raters (Uebersax and Grove 1990).

Despite the increased popularity of LC analysis in a broad range of research areas, no specific attention has been paid to power analysis for LC models. However, as in the application of other statistical methods, users of LC models wish to confirm the validity of their research hypotheses. This requires that a study has sufficient statistical power; that is, that it is able to confirm a research hypothesis when it is true. Also, reviewers of journal publications and research grant proposals often request sample size and/or power computations (Nakagawa and Foster 2004). However, in the literature on LC analysis, methods for sample size and/or power computation as well as a thorough study on the design factors affecting the power of statistical tests used in LC analysis, are lacking.

In this paper, we present a method for assessing the power of tests related to the class-specific response probabilities, which in confirmatory LC analysis are the parameters of main interest. Relevant tests include tests for whether response probabilities are equal across latent classes, whether response probabilities are equal to specific values, whether response proba-

bilities are equal across response variables (indicators), and whether sensitivities or specificities are equal across indicators (Goodman 1974; Holt and Macready 1989; Vermunt 2010b). Since the class-specific response probabilities are typically parameterized using logit equations (Formann 1992; Vermunt 1997), as in logistic regression analysis, hypotheses about these LC model parameters can be tested using Wald tests (Agresti 2002). The proposed power analysis method is therefore referred to as a Wald based power analysis.

For logistic regression models, Demidenko (2007, 2008) and Whittemore (1981) described the large-sample approximation for the power of the Wald test. In this paper, we show how to use this procedure in the context of LC analysis. An important difference compared to standard logistic regression analysis is that in a LC analysis the predictor in the logistic models for the responses, the latent class variable, is unobserved. This implies that the uncertainty about the individuals' class memberships should be taken into account in the power and sample size computation. As will be shown, factors affecting this uncertainty include the number of classes, the class proportions, the strength of the association between classes and indicator variables, and the number of indicator variables (Collins and Lanza 2010; Vermunt 2010a).

The remainder of this paper is organized as follows. Section 2 presents the LC model for dichotomous responses and discusses the relevant hypotheses for the parameters of the LC model. Section 3 discusses power computation for Wald tests in LC analysis and, moreover, shows how the LC specific design factors affect the power via the information matrix. Section 4 presents a numerical study in which we assess the performance of the proposed method and illustrates power/sample size computation for different scenarios of the relevant design factors. In the last section, we provide a brief discussion of the main results of our study.

## 2. The LC Model

The LC model is a probabilistic clustering or unsupervised classification model for dichotomous or categorical response variables (Goodman 1974; McCutcheon 1987; Hagenaars 1988; Magidson and Vermunt 2004; Vermunt 2010b). Taking the dichotomous case as an example, let  $y_{ij}$  be the value of response pattern  $i$  for the binary variable  $Y_j$ , for  $j = 1, 2, 3, \dots, p$ , where  $y_{ij} = 1$  represents a positive response and 0 a negative response. We denote the full-response vector by  $\mathbf{y}_i$ . For example, for  $p = 3$ ,  $\mathbf{y}_i$  takes on one of the following eight triplets of 0 and/or 1's:

$$\{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

The three response variables could, for example, represent the answers to the following questions: “Do you support gay marriage?”, “Do you support a raise of minimum wages?”, and “Do you support the initiative for health care reform?”. In a sample of size  $n$  persons, a particular person could answer these questions with ‘no’, ‘yes’, and ‘yes’, respectively, in which case the response pattern for this subject becomes  $(0, 1, 1)$ . In such an application, the aim of the analysis would be to determine whether one can identify two latent classes with different response tendencies (say republicans and democrats), and subsequently to classify subjects into one of these classes based on their observed responses, or to compare the probability of positive responses to a given response variable between the republican and the democrat classes.

In general, for  $p$  dichotomous response variables, we have  $2^p$  tuples of 0 and/or 1’s. We denote the number of individuals with response pattern  $\mathbf{y}_i$  by  $n_i$ , where the total sample size  $n = \sum_{i=1}^{2^p} n_i$ . The LC model assumes that the response probabilities depend on a discrete latent variable, which we denote by  $X$  with categories  $t = 1, 2, 3, \dots, c$ . The probability of having response pattern  $\mathbf{y}_i$  is modeled as a mixture of  $c$  class-specific probability functions (Dayton and Macready 1976; Goodman 1974; McCutcheon 1987; McLachlan and Peel 2000; Vermunt 2010b). That is,

$$P(\mathbf{y}_i, \Psi) = \sum_{t=1}^c P(X = t)P(\mathbf{Y} = \mathbf{y}_i|X = t), \quad (1)$$

where  $P(X = t)$ , which we also denote by  $\pi_t$ , represents the relative size of class  $t$ , and  $P(\mathbf{Y} = \mathbf{y}_i|X = t)$  is the corresponding class-specific joint response probability. The class-specific probabilities for binary variable  $Y_j$  is usually modeled using a logistic parameterization; that is,  $\theta_{jt} = P(Y_j = 1|X = t) = \frac{\exp(\beta_{jt})}{1 + \exp(\beta_{jt})}$ , where  $\beta_{jt}$  is the log-odds of giving a positive response on item  $j$  in class  $t$ . Moreover, assuming that the response variables are independent within classes—which is referred to as the local independence assumption—the LC model represented by equation (1) can be rewritten as follows:

$$P(\mathbf{y}_i, \Psi) = \sum_{t=1}^c \pi_t \prod_{j=1}^p \theta_{jt}^{y_{ij}} (1 - \theta_{jt})^{1-y_{ij}}, \quad (2)$$

where  $\pi_t$  is such that  $0 < \pi_t < 1$  and  $\sum_{t=1}^c \pi_t = 1$ . The vector of parameters,  $\Psi$ , consists of the sub-vector  $\boldsymbol{\pi}$ , the class proportions, and the sub-vector  $\boldsymbol{\beta}$ , the class-specific logits for the indicator variables. For example, for  $c = 2$  and  $p = 3$ , the parameter vector will be:  $\Psi' = (\boldsymbol{\pi}', \boldsymbol{\beta}') = (\pi_1, \beta_{11}, \beta_{21}, \beta_{31}, \beta_{12}, \beta_{22}, \beta_{32})$ . In the application presented above, these parameters would correspond to the proportion of ‘republicans’, the log-

odds of a republican responds ‘yes’ instead of ‘no’ to questions  $Y_1$ ,  $Y_2$ , and  $Y_3$ , and the log-odds of a democrat responds ‘yes’ instead of ‘no’ to questions  $Y_1$ ,  $Y_2$ , and  $Y_3$ .

In general, for a LC model having  $c$  classes and  $p$  binary indicator variables, we have  $m = c - 1 + c \cdot p$  free model parameters. These parameters are usually estimated by maximum likelihood (ML) (Dayton and Macready 1976; Goodman 1974; McLachlan and Peel 2000; Vermunt 2010b), which involves seeking the values of  $\Psi$ , say  $\hat{\Psi}$  which maximize the log-likelihood function:

$$l(\Psi) = \sum_{i=1}^{2^p} n_i \log P(\mathbf{y}_i, \Psi). \quad (3)$$

Maximizing the log-likelihood function in (3) produces a unique estimate for  $\Psi$ , provided that the LC model in equation (1) is identifiable. As indicated by Goodman (1974), a necessary condition for an LC model to be identified is that the number of independent response patterns is at least as large as the number of free model parameters. That is,  $2^p - 1 \geq m = c - 1 + c \cdot p$ . A sufficient condition for local identification is that the Jacobian is full rank (McHugh 1956). Because the analytic evaluation of the rank of the Jacobian is very difficult, Forcina (2008) proposed checking identification of LC models by evaluating the rank of the Jacobian for a large number of random parameter values. For the scenarios considered in this paper we applied Forcina’s method, which showed that the models were identified.

Typically, researchers using LC models do not only wish to obtain point estimates for the  $\Psi$  parameters, but are also interested in tests concerning these parameters. For simplicity we will focus on a single type of test, which in most applications is the test of main interest. That is, the test to determine whether there is a significant association between the latent classes and a particular indicator variable. Inference regarding this association involves testing the null hypothesis that the response logit does not differ across latent classes for the indicator variable concerned. This null hypothesis can be formulated as  $H_0 : \beta_{j1} = \beta_{j2} = \dots = \beta_{jc}$ , for  $j = 1, 2, 3, \dots, p$ . An equivalent formulation of this hypothesis is

$$\begin{aligned} H_0 : \beta_{j1} - \beta_{j2} &= 0 \\ \beta_{j1} - \beta_{j3} &= 0 \\ &\cdot \\ &\cdot \\ &\cdot \\ \beta_{j1} - \beta_{jc} &= 0. \end{aligned}$$

Or, using matrix notation, as  $H_0 : \mathbf{H}\beta_j = 0$ , where  $\mathbf{H}$  is a  $c - 1$  by  $c$  design matrix with linear contrasts and  $\beta_j$  is a  $c$  by 1 column vector with the parameters for  $Y_j$ , i.e.,  $\beta_j' = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jc})$ . Under the null hypothesis of no association, the difference  $\beta_{j1} - \beta_{jt}$  occurs by chance alone, implying that the indicator does not contribute to the definition of classes in a statistically significant way.

As already indicated in the introduction section, various other types of hypotheses concerning the class-specific logit parameters may be of interest. Examples include tests for whether  $\beta_{jt}$  is equal to a particular value (e.g.,  $\beta_{11} = 1$ ), whether the  $\beta_{jt}$  parameters are equal across two or more items (e.g.,  $\beta_{1t} - \beta_{2t} = 0$ ), and whether the value is the opposite of the value for another class (e.g.,  $\beta_{11} + \beta_{12} = 0$ ) (Goodman 1974). In medical research, we may be interested in comparing the sensitivity and specificity of diagnostic tests (see, for example Yang and Becker 1997), yielding hypotheses such as  $\beta_{11} - \beta_{21} = 0$  and  $\beta_{12} - \beta_{22} = 0$ , respectively. Note that all these hypotheses can be expressed in the general form  $\mathbf{H}\beta = 0$ .

### 3. Wald Based Power Analysis for LC Models

#### 3.1 The Wald Statistic and Its Asymptotic Properties

One of the properties of the ML estimator is that, under certain regularity conditions (McHugh 1956; White 1982), the estimator  $\hat{\Psi}$  converges in probability to  $\Psi$  as the sample size tends to infinity. That is, for any sequence  $\hat{\Psi}_n$  we have  $\hat{\Psi}_n \xrightarrow{a.s.} \Psi$ . The other interesting property of the ML estimator is that it has a limiting normal distribution. More specifically, for large sample size  $n$ ,

$$\sqrt{n}(\hat{\Psi}_n - \Psi) \longrightarrow N(\mathbf{0}, \mathbf{V}), \quad (4)$$

where  $\longrightarrow$  denotes convergence in distribution,  $\mathbf{V} = \mathbf{I}^{-1}(\Psi)$  is the asymptotic co-variance of  $\sqrt{n}\hat{\Psi}_n$ , and  $\mathbf{I}(\Psi)$  is the  $m$  by  $m$  information matrix (McHugh 1956; Redner 1981; Rencher 2000; Wald 1943; Wolfe 1970). The latter has the following block structure:

$$\mathbf{I}(\Psi) = \begin{bmatrix} \mathbf{I}_1 = \{(\pi_t, \pi_s)\} & \mathbf{I}_2 = \{(\pi_t, \beta_{jl})\} \\ \mathbf{I}_3 = \{(\beta_{jq}, \pi_s)\} & \mathbf{I}_4 = \{(\beta_{jq}, \beta_{kl})\} \end{bmatrix},$$

for  $t, s = 1, 2, 3, \dots, c - 1$ ,  $l, q = 1, 2, 3, \dots, c$  and  $k, j = 1, 2, 3, \dots, p$ . The sub-matrices  $\mathbf{I}_1$ ,  $\mathbf{I}_2$ ,  $\mathbf{I}_3$ , and  $\mathbf{I}_4$  are of dimensions  $c - 1$  by  $c - 1$ ,  $c - 1$  by  $c \cdot p$ ,  $c \cdot p$  by  $c - 1$ , and  $c \cdot p$  by  $c \cdot p$ , respectively. The terms between braces indicate the parameters involved in the sub-matrix concerned.

Using the algebraic properties of block matrices, it follows that

$$\mathbf{V} = \mathbf{I}^{-1}(\Psi) = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{I}_1^{-1}\mathbf{I}_2\mathbf{B}^{-1} \\ -\mathbf{I}_4^{-1}\mathbf{I}_3\mathbf{A}^{-1} & \mathbf{B}^{-1} \end{bmatrix}, \quad (5)$$

where  $\mathbf{A} = \mathbf{I}_1 - \mathbf{I}_2\mathbf{I}_4^{-1}\mathbf{I}_3$  and  $\mathbf{B} = \mathbf{I}_4 - \mathbf{I}_3\mathbf{I}_1^{-1}\mathbf{I}_2$ . A necessary condition for  $\mathbf{A}$  to be invertible, which is a requirement to obtain the covariance matrix of  $\hat{\Psi}_n$ , is that both  $\mathbf{I}_1$  and  $\mathbf{I}_4$  are non-singular matrices (Rencher 2000). In the Appendix section, we provide details on the expressions for  $\mathbf{I}_1$ ,  $\mathbf{I}_2$ ,  $\mathbf{I}_3$ , and  $\mathbf{I}_4$ .

The consistency and multivariate normality discussed above apply to the estimators of the component parameters as well. That is, using the property of multivariate normal random variables which states that the sub-vectors of a multivariate normal are also normal, the limiting distributions of  $\hat{\pi}$  and  $\hat{\beta}$  become

$$\sqrt{n}(\hat{\pi}_n - \pi) \longrightarrow N(\mathbf{0}, \mathbf{A}^{-1}) \quad (6)$$

$$\sqrt{n}(\hat{\beta}_n - \beta) \longrightarrow N(\mathbf{0}, \mathbf{B}^{-1}). \quad (7)$$

Also sub-vector  $\hat{\beta}_j$  of  $\hat{\beta}$  is normally distributed, with mean  $\beta_j$  and with covariance  $\mathbf{V}_j$ , being a  $c$  by  $c$  sub-matrix of  $\mathbf{B}^{-1}$ . In the remaining part of the paper, we focus on this  $\beta_j$ .

Using the Continuous Mapping Theorem (Mann and Wald 1943), for a design matrix  $\mathbf{H}$  that defines the contrasts on the null hypothesis, one can show that  $\mathbf{H}\hat{\beta}_j \longrightarrow N(\mathbf{H}\beta_j, \mathbf{H}\mathbf{V}_j\mathbf{H}')$ . The quadratic form of the test for the hypothesis  $H_0 : \mathbf{H}\beta_j = \mathbf{0}$  yields the well-known Wald statistic:

$$W = n \left( (\mathbf{H}\hat{\beta}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\hat{\beta}_j) \right). \quad (8)$$

Under the null hypothesis, that is, if  $H_0 : \mathbf{H}\beta_j = \mathbf{0}$  holds, the Wald statistic  $W$  has an asymptotic (central) chi-square distribution with  $c - 1$  degrees of freedom (Rencher 2000; Wald 1943). That is,

$$W = n \left( (\mathbf{H}\hat{\beta}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\hat{\beta}_j) \right) \longrightarrow \chi_{(c-1)}^2. \quad (9)$$

Under the alternative hypothesis,  $W$  follows a non-central chi-square distribution with  $c - 1$  degrees of freedom and non-centrality parameter  $\lambda$ . That is,

$$W = n \left( (\mathbf{H}\hat{\beta}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\hat{\beta}_j) \right) \longrightarrow \chi_{(c-1, \lambda)}^2 \quad (10)$$

where  $\lambda = n(\mathbf{H}\beta_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\beta_j)$ .

### 3.2 Power and Sample Size Computation

With the establishment of the distribution of the test statistic under the null and alternative hypotheses and the availability of a closed form expression for the non-centrality parameter  $\lambda$ , it becomes possible to compute the power of the test for a given sample size or the sample size for a given power. As in any power analysis, we first have to define the population model. In our case, this involves defining the number of classes and the number of response variables, and, moreover, specifying the values for the class proportions  $\pi$  and the class-specific logits  $\beta$ . For the assumed population model, we can compute the inverse information matrix  $\mathbf{V}$  which appears in the formula of the non-centrality parameter.

Once the population parameters are set and  $\mathbf{V}$  is computed, power computation for a given sample size and required sample size computation for a given power proceeds along the steps described below.

#### 3.2.1 Steps for Power Computation

Power computation proceeds as follows:

1. Compute the non-centrality parameter  $\lambda$  for the specified sample size  $n$  (use the expression in equation 10).
2. For a given value of type I error  $\alpha$ , read the  $100(1-\alpha)$  percentile value from the (central) chi-square distribution. That is, find  $\chi_{(1-\alpha)}^2(c-1)$  such that under the null hypothesis,  $P\left(W > \chi_{(1-\alpha)}^2(c-1)\right) = \alpha$ . This value is referred to as the critical value of a test.
3. Compute the power as the probability that a random variable  $W$  from the non-central chi-square distribution (with non-centrality parameter  $\lambda$  given in step 1) will assume a value greater than the critical value obtained under step 2.

#### 3.2.2 Steps for Sample Size Computation

Sample size computation proceeds as follows:

1. For a given value of  $\alpha$ , read the  $100(1-\alpha)$  percentile value from the (central) chi-square distribution (see the second step for power computation).
2. For a given power and the critical value obtained in step 1, find the non-centrality parameter  $\lambda$  such that, under the alternative hypothesis, the condition that power is equal to  $P\left(W > \chi_{(1-\alpha)}^2(c-1)\right)$  is satisfied.



3. From the expression for  $\lambda$ , solve for the sample size as

$$n = \lambda \left( (\mathbf{H}\boldsymbol{\beta}_j)' (\mathbf{H}\mathbf{V}_j\mathbf{H}')^{-1} (\mathbf{H}\boldsymbol{\beta}_j) \right)^{-1}.$$

### 3.2.3 Software Implementation

The above procedure for power computation can be applied using existing software for LC analysis that allows defining starting values or fixed values for the logit parameters and that provides the (inverse) information matrix as output, for example, using LEM (Vermunt 1997), Mplus (Muthén and Muthén 2012), or Latent GOLD (Vermunt and Magidson 2013b). More specifically, with a LC analysis software package, the inverse information matrix  $\mathbf{V}$  can be obtained. This will typically require the following two steps:

- A. Create a data set containing all possible data patterns and with the expected frequencies according to the LC model of interest as weights. This can be achieved by running the LC software with the population parameters specified as fixed values and with the estimated frequencies as requested output. The created output is, in fact, a data set which is exactly in agreement with the population model. Such a data set is sometimes referred as an 'exemplary' data set (ÓBrien 1986).
- B. Analyze the (exemplary) data set created in step A with the LC model of interest and request the variance-covariance matrix of the parameters (the inverse information matrix) as output. Note that when analyzing a data set which is exactly in agreement with the model, the observed information matrix is identical to the expected information matrix. The same applies to the approximate observed information matrix based on the outer-product of the gradient contributions of the data patterns (see Appendix).

The above two steps provide us with the inverse information matrix  $\mathbf{V}$ . The actual power or sample size computations using the steps described above can subsequently be performed using software that allows performing matrix computations and that has functions for obtaining the critical value from the chi-squared distribution and the non-centrality value from the non-central chi-squared distribution. For this purpose, one can use R. An R script is available from the first author.

The procedure described above is fully automated in version 5.0 of the Latent GOLD program (Vermunt and Magidson 2013b). Users define the population model and specify either the sample size or the required power. The program computes the power or the required sample size for the Wald tests it reports by default, as well as for other Wald tests defined by the user. In the Appendix, we give an example of the Latent GOLD syntax for power computation.

### 3.3 Design Factors Affecting the Power of a Wald Test in LC Models

Now let us look in more detail at the factors affecting the power of the Wald test in LC models. It should be noted that the power is determined by the value of the type I error and the value of the noncentrality parameter  $\lambda$ . The larger the type I error and the larger  $\lambda$ , the larger the power. As can be observed from equation (10),  $\lambda$  is a function of the sample size  $n$ , the precision of the estimator ( $\mathbf{V}_j$ ), and the effect size  $\mathbf{H}\beta_j$ . Note that in our case the effect size is the difference between the class-specific  $\beta$  parameters or, equivalently, the strength of the association between the classes and the response variable concerned.

Specific for LC models is that the precision of the estimator is affected by the fact that class membership is unobserved; that is, that we are uncertain about a person's class membership. Recall from equation (5) that the block of  $\mathbf{V}$  concerning the  $\beta$  parameters is obtained as the inverse of  $\mathbf{B} = \mathbf{I}_4 - \mathbf{I}_3\mathbf{I}_1^{-1}\mathbf{I}_2$ . This means that  $\mathbf{B}$  becomes larger when  $\mathbf{I}_4$  and  $\mathbf{I}_1$  become larger and when  $\mathbf{I}_2$  and  $\mathbf{I}_3$  become smaller. To show how the uncertainty about the class membership affects  $\mathbf{B}$ , let us have a closer look at  $\mathbf{I}_4$ , which is the most important term in  $\mathbf{B}$ . Its elements are obtained as follows:

$$\mathbf{I}_4(\beta_{jq}, \beta_{kl}) = \sum_{i=1}^{2^p} P(X = q|\mathbf{y}_i)P(X = l|\mathbf{y}_i)(y_{ij} - \theta_{jq})(y_{ik} - \theta_{kl})P(\mathbf{y}_i), \quad (11)$$

where  $\theta_{jq} = \exp(\beta_{jq})/(1 + \exp(\beta_{jq}))$ . As can be seen, specific for a LC analysis is that the elements of the information matrix are not only a function of the model parameters, but also of the posterior class membership probabilities  $P(X = q|\mathbf{y}_i)$ . For example, the contribution of response pattern  $i$  to the information on parameter  $\beta_{jq}$  equals  $P(X = q|\mathbf{y}_i)^2(y_{ij} - \theta_{jq})^2P(\mathbf{y}_i)$ . In other words, response pattern  $i$  contributes with "weight"  $P(X = q|\mathbf{y}_i)^2$  to the information on a parameter of class  $q$ . The contribution to total of the parameters of all  $c$  classes equals  $\sum_{t=1}^c P(X = t|\mathbf{y}_i)^2$ . This shows that the information is maximal when  $P(X = q|\mathbf{y}_i)$  equals 1 for one class and 0 for the other classes, in which case the total contribution equals 1. This occurs when the classes are perfectly separated or when the class membership is observed rather than latent.

Also the entries of  $\mathbf{I}_1$  become larger when the posterior class membership probabilities get closer to either 0 or 1. The matrices  $\mathbf{I}_2$  and  $\mathbf{I}_3$  capture the overlap in information between the class proportions and the  $\beta$  parameters. The elements of this matrix are 0 when separation is perfect and become larger with lower class separation.

The implication of the above is that the power can be increased by increasing the separation between the classes; that is, by influencing the

factors affecting the posterior class membership probabilities. The posterior class membership probabilities depend on the number of classes, the class proportions, the class-specific conditional response probabilities, and the number of response variables (Collins and Lanza 2010; Vermunt 2010a). More specifically, class separation is better with less latent classes, a more uniform class distribution, response variables which are more strongly related to the classes, and a larger number of response variables.

Note that the conditional response probabilities have a dual role. The more the conditional response probabilities  $\theta_{jq}$  or the logit parameters  $\beta_{jq}$  differ across latent classes, the larger the effect size and thus also the higher the power of the test for the parameters of indicator variable  $Y_j$ . However, a larger difference between classes in the response on  $Y_j$  also increases the class separation, and thus the power of all tests, also the ones for the other response variables.

#### 4. Numerical Study

In this section, we present a numerical study that illustrates the Wald based power analysis for different configurations of design factors. As was shown in Section 3, in addition to the usual factors (i.e., sample size, level of significance, and effect size), power computation in LC models involves the specification of design factors such as the number of classes, the number of observed response variables, the class sizes, and the class-specific probabilities (or logits) for the response variables, which we refer to as LC-specific design factors.

As already indicated in Section 3.3, LC-specific design configurations yielding better separated classes, or posterior class membership probabilities which are closer to either 0 or 1, yield more precise estimators, and as a result larger power of the Wald tests. Therefore, in order to be able to compare different design configurations, it is important to have a measure for class separation. For this purpose, we use the entropy based R-square. The entropy of the posterior class membership probabilities for data pattern  $i$ , denoted by  $E_i$ , equals  $\sum_{t=1}^c -P(X = t|\mathbf{y}_i) \log P(X = t|\mathbf{y}_i)$ . Note that  $E_i$  gets closer to 0 when the posteriors are closer to 0 and 1. The average entropy across data patterns, denoted by  $E$ , equals  $\sum_{i=1}^{2^p} E_i P(\mathbf{y}_i)$ . The entropy based R-square can now be obtained as follows:  $R_{\text{entropy}}^2 = 1 - E/E(0)$ . Here,  $E(0)$  is the maximum entropy given the class proportions; that is,  $E(0) = \sum_{t=1}^c -P(X = t) \log P(X = t)$ . The entropy based R-square takes on values between 0 and 1, where larger  $R_{\text{entropy}}^2$  indicate larger separation between classes. Values lower than .5, between .5 and .75, and larger than .75 correspond to LC models with small, medium, and large class separation, respectively. Closer inspection of the expression  $R_{\text{entropy}}^2 = 1 - E/E(0)$  shows that the largest entropy based R-square is

obtained when  $E$  equals 0. This occurs when  $P(X = t|\mathbf{y}_i)$  is either 0 or 1 for each response pattern  $\mathbf{y}_i$ ; that is, when class separation is perfect.

#### 4.1 Manipulation of the Design Factors

The LC-specific design factors that were varied are the number of classes, the number of indicator variables, the class-specific conditional probabilities, and the class proportions. The number of classes varied from 2 to 4 (i.e.,  $c = 2, 3, 4$ ). The number of indicator variables was set to  $p = 6$  and  $p = 10$ . The class-specific conditional probabilities  $\theta_{jt}$  were 0.7, 0.8, and 0.9 (or, depending on the class, 1-0.7, 1-0.8, and 1-0.9), corresponding to a weak, medium, and strong association between classes and indicator variables. The  $\theta_{jt}$  were high for class 1, say 0.8, and low for class  $c$ , say 1-0.8; with  $c = 3$ , class 2 had high  $\theta_{jt}$  values for the first half of the items and low values for the other items; with  $c = 4$ , class 2 had low  $\theta_{jt}$  values for the first half of the items and high values for the other items, and class 3 had high  $\theta_{jt}$  values for the first half of the items and low values for the other items. The class sizes were equal or unequal, where for the unequal conditions we used class proportions of (0.75, 0.25), (0.5, 0.3, 0.2), (0.6, 0.3, 0.1), and (0.4, 0.3, 0.2, 0.1), respectively.

In addition to the four LC-specific design factors, we varied the sample size, power, level of significance, and effect size (Cohen, 1988). For power computation, the sample size was set to 75, 100, 200, 300, 500, 700, 1000, and 1500, whereas for sample size determination, the power was set to .8, .9, and .95. The type I error was fixed to 0.05. The effect size is already specified via the response probabilities  $\theta_{jt}$ , where it should be noted that the logit coefficients  $\beta_{jt}$  for which the Wald tests are performed equal  $\beta_{jt} = \log \theta_{jt}/(1 - \theta_{jt})$ .

#### 4.2 Effects of Design Factors on Power and Sample Size

Table 1 presents the entropy based R-square for several combinations of the LC-specific design factors. It shows how the value of this R-square measure is affected by the number of classes, the class proportions, the number of indicators, and the strength of the class-indicator associations, given specific values of the other design factors. As can be seen, the smaller the number of the classes, the larger the number of indicator variables, and/or the stronger the class-indicator associations, the larger the value of the entropy based R-square. Moreover, the more equal the class sizes, the larger the entropy. It can also be seen that the entropy based R-square may become very low when all conditions are less favorable.

To investigate the effect of class separation on the power of the Wald test for the significance of a class-indicator association, the power is com-

Table 1. Entropy based R-square values for different combinations of LC-specific design factors

		Class size		
		Equal	Unequal	More unequal
Number of classes (for $p = 6$ and $\theta_{j1}=0.8$ )	$c = 2$	.818	.811	
	$c = 3$	.627	.624	
	$c = 4$	.594	.589	
Number of indicators (for $c = 3$ and $\theta_{j1}=0.8$ )	$p = 6$	.627	.624	
	$p = 10$	.790	.788	
Class-indicator associations (for $c = 3$ , and $p = 6$ )	$\theta_{j1}=0.7$	.332	.330	.314
	$\theta_{j1}=0.8$	.627	.624	.607
	$\theta_{j1}=0.9$	.880	.879	.871

Note: the 'unequal' and 'more unequal' class size conditions refer to the level of deviation from uniform class distribution. For example, for  $c = 3$ , we used (0.5, 0.3, 0.2) and (0.6, 0.3, 0.1) to represent a smaller and larger deviation from a uniform class distribution, respectively.

puted for five of the design configurations that were presented in Table 1 under different sample sizes. The results are presented in Table 2. From this table, we can see that the power of a Wald test for class-indicator association strongly depends on the class separations. When classes are well separated, a sample size of 100 can be large enough to achieve a power of .8 or more. With a class separation of .330, .607, and .624, a sample size of 900, 370, and 140, respectively, is required to achieve such a power. With very badly separated classes as in the worst condition, even a sample size of 1500 is not large enough to achieve a power of .8.

Table 3 reports the required sample size for a specified power for various combinations of LC-specific design factors. We use the condition with  $c = 3$ ,  $p = 6$ , equal size classes, and medium class-indicator associations as the baseline. This condition requires sample sizes of 82, 108, and 131, respectively, to achieve the three reported power levels. The other conditions are obtained by varying one design factor at the time.

The results in Table 3 show that, as expected, the required sample size depends on the number of classes, the number of indicators, the strength of the class-indicator associations, and the class sizes. More specifically, keeping the other LC-specific design factors constant, the larger the number of classes and the fewer the number of indicators, the larger the required sample size to achieve the specified power level. The strength of the class-indicator associations turns out to be one of the key factors affecting the power; for example, to obtain a power of .80, we need at least 419 observations when these associations are weak, but only 34 observations when these are strong. Moreover, many more observations are required when the class sizes are un-

Table 2. Estimated power (%) for different class separation levels and different sample sizes

Sample size	Entropy based R-square				
	.314	.330	.607	.624	.790
75	7	12	22	51	94
100	8	14	28	64	98
200	10	24	52	92	100
300	13	34	71	99	100
500	19	53	91	100	100
700	25	69	98	100	100
1000	34	84	100	100	100
1500	49	96	100	100	100

$$H_0 : \beta_{j1} = \beta_{j2} = \dots = \beta_{jc} \text{ for which } j = 1 \text{ and } c = 3.$$

Table 3. Required sample size for different configurations of LC-specific design factors and different power levels

Power	Number of classes			Number of indicators		
	$c = 2$	$c = 3$	$c = 4$	$p = 6$	$p = 10$	
.8	33	82	83	82	49	
.9	45	108	108	108	64	
.95	55	131	130	131	78	
Power	Class-indicator associations			Class sizes		
	Low	Medium	High	Equal	Unequal	More unequal
.8	419	82	34	82	141	371
.9	550	108	45	108	185	487
.95	671	131	55	131	226	594

Note: The baseline model is the model with  $c = 3$ ,  $p = 6$ , equal size classes, and medium association between classes and indicators. One design factor is varied to get the other conditions reported in the table.

equal than when they are equal; for example, to achieve a power of .95, we need approximately 130, 225, and 600 observations for the (0.334, 0.333, 0.333), (0.5, 0.3, 0.2), and (0.6, 0.3, 0.1) condition, respectively.

In summary, these results show that the strength of the class-indicator associations and the class distribution have a much stronger impact on the power than the number of classes and the number of indicator variables. The fact the strength of the class-indicator association is so important can be explained by the fact it affects both the class separation and the effect size. For example, for  $P = 6$ ,  $C = 3$ , and equal class sizes, when the

$\theta_{jt}$  value changes from .9 to .7, the class separation drops from .880 to .332 and the difference between classes in their conditional response probabilities drops from .8 to .4. Thus, a  $\theta_{jt}$  value of .9 yields not only a much larger R-square value but also a much larger effect size than a  $\theta_{jt}$  value of .7. The class sizes are important because the power of a test regarding differences between groups depends strongly on the size of the smallest group.

### 4.3 Performance of the Power Computation Procedure

An important question is whether the theoretical power computed using the formulae presented in this paper agrees with the actual power when using the Wald with empirical data. To answer this question, we conducted a simulation study in which the theoretical power is compared with the actual power in data sets generated from the assumed population model. Note that the actual power equals the proportion of simulated data sets in which the null hypothesis is rejected.

The population model is a three-class LC model with six indicators and equal class sizes. We varied the strength of the class-indicator associations (same three levels as above) and the sample size (75, 100, 200, 300, 500, 700, and 1000). The actual power was computed using 500 samples from the population under the alternative hypothesis. For each of these samples, the LC model is estimated and it is checked whether the Wald value for the test of interest exceeds the critical value.

Table 4 presents the theoretical and actual power of the Wald test under the investigated simulation conditions. As can be seen, both measures show the same overall trend, namely that the power increases with increasing sample size and increasing effect size (and class separation). However, the actual power of the Wald test is always slightly lower than its theoretical value, where the differences are larger for the smaller sample size and the weaker class separation conditions. An explanation for these differences is that the estimated asymptotic variance-covariance matrix used in the simulated power computations overestimates the variability of the  $\beta_j$  parameters. On the other hand, substantive conclusions are the same for the simulated and theoretical power levels reported in Table 4. With the small effect size and the corresponding weak class separation condition, a sample size of 500 is needed to achieve a power of .8; with the medium class separation, a sample size of 100 suffices; and with the strong class separation, less than 75 observations are needed.

## 5. Discussion and Conclusion

In LC analysis, the association between class membership and the response variables is usually modeled using a logistic parametrization. This

Table 4. Theoretical and simulated power of the Wald test

Class-indicator		Sample size						
associations	Method	75	100	200	300	500	700	1000
Weak	Theoretical	.200	.254	.470	.649	.869	.958	.994
	Simulated	.145	.234	.444	.628	.838	.920	.960
Medium	Theoretical	.762	.877	.995	1.000	1.000	1.000	1.000
	Simulated	.714	.848	.944	.992	1.000	1.000	1.000
Strong	Theoretical	.989	.999	1.000	1.000	1.000	1.000	1.000
	Simulated	.986	1.000	1.000	1.000	1.000	1.000	1.000

The power presented here is for the null hypothesis  $H_0 : \beta_{j1} = \beta_{j2} = \dots = \beta_{jc}$  for which  $j = 1$ . Moreover,  $c = 3$ ,  $p = 6$ , and class sizes are equal.

paper dealt with power analysis for Wald tests for these logit coefficients, for example, for the hypothesis of no association between class membership and the response provided on one of the indicators. We showed that, in addition to the usual design factors—that is, effect size, sample size, and level of significance—the power of Wald tests in LC models depends on factors affecting the amount uncertainty about the subjects’ class memberships. More specifically, factors affecting the class separation also affect the power. The most important of these LC-specific design factors are the number of classes, the class proportions, the strength of the class-indicator associations, and the number of indicator variables.

A numerical study was conducted to illustrate the proposed power and sample size computation procedures. More precisely, it was shown how class separation—quantified using the entropy-based R-square—is affected by the number of classes, the class proportions, the strength of the class-indicator associations, and the number of indicator variables, and, moreover, how class separation affects the power. It turned out that under the most favorable conditions a sample size of 100 suffices to achieve a power of .8 or .9. For the situation where the entropy-based R-square is small, a considerably larger sample size is required. It was shown that under the least favorable conditions, even a sample size of 2000 did not suffice to achieve an acceptable power level. This demonstrates the importance of performing a power analysis prior to conducting a study that will make use of the LC analysis.

If power turns out to be too low given the planned sample size, instead of increasing the sample size, one may try to increase the class separation, for example, by using a larger number of indicators or, if possible, also by using indicators of a better quality. Note that improving the quality of indicators has a dual effect on the power of the Wald test for class-indicator associations: It increases both the effect size and the class separation. This dual effect could be seen in our numerical study where we saw a dramatic



reduction of the required sample size when the  $\theta_{jt}$  value increased from .7 to .9. In practice, improving the quality of the indicators will not be easy, even in the type of more confirmatory LC analyses we were dealing with.

A simulation study was conducted to evaluate whether the theoretical power corresponds with the actual power of the Wald test. It turns out that the estimated power obtained with the formulae provided in this paper is slightly larger than the actual power, where we see a larger overestimation for smaller sample sizes and lower power levels. This implies that to be on the safe side, to achieve the specified power, a slightly larger sample size may be used than the estimated sample size.

In this paper, we restricted ourselves to power computations for Wald tests. However, likelihood-ratio tests are often used in LC models as well, either for testing the same kinds of hypotheses as discussed here or for comparing models with different number of latent classes. Future research will focus on power computation for likelihood-ratio tests in LC models.

Another limitation of the current work is that we restricted ourselves to simple LC models. In future work, we will investigate whether the methods discussed in this paper can be extended to more complex LC models, such as LC models with covariates, latent Markov models, mixture growth models, and mixture regression models.

Most of the simulation studies on LC and mixture modelling show that larger sample sizes may be needed than those found with the power computation method described in the current paper (see for example, Yang 2006; Nylund, Asparouhov, and Muth n 2007; Tofghi and Enders 2008). Those studies are, however, about deciding on the number of classes, whereas here we focus on the class-indicator association for a single response variable assuming that the number of classes is known. Note also that these studies typically do not look at significance testing, but at the performance of measures like BIC, which may have less power because of their penalty for model complexity. Further research is needed on the power of statistical tests for deciding about the number of classes, for example, of the bootstrapped likelihood-ratio test.

## Appendix

### A.1 Elements of the Information Matrix in a LC Model for Binary Responses

The elements of the information matrix  $\mathbf{I}(\Psi)$ , with  $\Psi' = (\pi', \beta')$ , equal to minus the expected value of the second-order partial derivatives of the log-likelihood function defined in (3) with respect to the free parameters divided by the sample size.

In a LC model, these have the following rather simple form:

$$\begin{aligned} \mathbf{I}(\psi_l, \psi_q) &= -E \left( \frac{\partial^2 l(\Psi)}{\partial \psi_l \partial \psi_q} \right) / n \\ &= \sum \frac{\partial \log P(\mathbf{y}_i, \Psi)}{\partial \psi_l} \frac{\partial \log P(\mathbf{y}_i, \Psi)}{\partial \psi_q} P(\mathbf{y}_i, \Psi). \end{aligned}$$

This shows that the computation of the information matrix requires solving the first-order partial derivatives  $\frac{\partial \log p(\mathbf{y}_i)}{\partial \psi_l}$ . For a class-proportion  $\pi_t$  and a class-specific response logit  $\beta_{jt}$ , these take on the following form:

$$\begin{aligned} \frac{\partial \log P(\mathbf{y}_i, \Psi)}{\partial \pi_t} &= \frac{P(X = t|\mathbf{y}_i)}{\pi_t} - \frac{P(X = c|\mathbf{y}_i)}{\pi_c}, \\ \frac{\partial \log P(\mathbf{y}_i, \Psi)}{\partial \beta_{jt}} &= P(X = t|\mathbf{y}_i)(y_{ij} - \theta_{jt}). \end{aligned}$$

This yields the following forms for the entries of the sub-matrix  $\mathbf{I}_1$ ,  $\mathbf{I}_2$ ,  $\mathbf{I}_3$ , and  $\mathbf{I}_4$ :

$$\begin{aligned} \mathbf{I}_1(\pi_t, \pi_s) &= \sum_{i=1}^{2^p} \left( \frac{P(X = t|\mathbf{y}_i)}{\pi_t} - \frac{P(X = c|\mathbf{y}_i)}{\pi_c} \right) \\ &\quad \left( \frac{P(X = s|\mathbf{y}_i)}{\pi_s} - \frac{P(X = c|\mathbf{y}_i)}{\pi_c} \right) P(\mathbf{y}_i, \Psi), \\ \mathbf{I}_2(\pi_t, \beta_{jl}) &= \sum_{i=1}^{2^p} \left( \frac{P(X = t|\mathbf{y}_i)}{\pi_t} - \frac{P(X = c|\mathbf{y}_i)}{\pi_c} \right) \\ &\quad P(X = l|\mathbf{y}_i)(y_{ij} - \theta_{jl})P(\mathbf{y}_i, \Psi), \\ \mathbf{I}_3(\beta_{jq}, \pi_s) &= \sum_{i=1}^{2^p} P(X = q|\mathbf{y}_i)(y_{ij} - \theta_{jq}) \\ &\quad \left( \frac{P(X = s|\mathbf{y}_i)}{\pi_s} - \frac{P(X = c|\mathbf{y}_i)}{\pi_c} \right) P(\mathbf{y}_i, \Psi), \\ \mathbf{I}_4(\beta_{jq}, \beta_{kl}) &= \sum_{i=1}^{2^p} P(X = q|\mathbf{y}_i)(y_{ij} - \theta_{jq})P(X = l|\mathbf{y}_i) \\ &\quad (y_{ik} - \theta_{kl})P(\mathbf{y}_i, \Psi). \end{aligned}$$

Note that  $P(\mathbf{y}_i, \Psi) = \sum_{t=1}^c \pi_t \prod_{j=1}^p \theta_{jt}^{y_{ij}} (1 - \theta_{jt})^{1-y_{ij}}$  is the probability for response pattern  $\mathbf{y}_i$ . Moreover  $P(X = t|\mathbf{y}_i) = \frac{\pi_t P(\mathbf{Y}=\mathbf{y}_i)|X=t}{P(\mathbf{y}_i, \Psi)}$  is the posterior class membership probability, where  $P(\mathbf{Y} = \mathbf{y}_i | X = t) = \prod_{j=1}^p \theta_{jt}^{y_{ij}} (1 - \theta_{jt})^{1-y_{ij}}$  is the joint class-specific probability.

## A.2 An Example of the Latent GOLD Setup for Wald Based Power Computation

The Latent GOLD 5.0 (Vermunt and Magidson 2013a) Syntax system implements the power computation procedure described in this paper. In order to perform such a Wald power computation, one should first create a small “example” data set; that is, a data set with the structure of the data one is interested in. With six binary response variables ( $y_1$  through  $y_6$ ), this file could be of the form:

```
y1  y2  y3  y4  y5  y6
0   0   0   0   0   0
```

which is basically a data set with a single observation with a response of 0 on all six variables.

For this small data set, one defines the model of interest and requests the power or the required sample size using the output options. This is done as follows using the Latent GOLD “options”, “variables”, and “equations” sections:

```
options
  output parameters standarderrors
    WaldPower=<number> WaldTest='fileName';
variables
  dependent y1 2, y2 2, y3 2, y4 2, y5 2, y6 2;
  latent x nominal 2;
equations
  x <- 1;
  y1 - y6 <- 1 | x;
  {0.0000000000
  1.386294361 -1.386294361
  1.386294361 -1.386294361
  1.386294361 -1.386294361
  1.386294361 -1.386294361
  1.386294361 -1.386294361
  1.386294361 -1.386294361}
```

In the “variables” section, we define the variables which are in the model and also their number of categories. These are the six response variables and the latent variable “x”. The “equations” section specifies the logit equations defining the model of interest, as well as the values of the population parameters. Note that the value 1.386294361 for a logit coefficients corresponds to a conditional response probability of .80.

The “output” line in the “options” section lists the output requested. With `WaldPower=<number>`, one requests a power or sample size compu-

tation. When using a “number” between 0 and 1, the program reports the required sample size for that power, and when using a values larger than 1, the program reports the power obtained with that sample size. The optional statement `WaldTest='filename'` can be used to define user-specific Wald test in addition to the test which are provided by default. The linear contrasts for the user-defined hypotheses of interest are defined in a text file.

## References

- AGRESTI, A. (2002), *Categorical Data Analysis*, Hoboken NJ: John Wiley & Sons, Inc.
- COHEN, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale NJ: Lawrence Erlbaum Associates.
- COLLINS, L.M., and LANZA, S.T. (2010), *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*, Hoboken NF: John Wiley & Sons, Inc.
- DAYTON, C.M., and MACREADY, G.B. (1976), “A Probabilistic Model for Validation of Behavioral Hierarchies”, *Psychometrika*, 41(2), 189–204.
- DAYTON, C.M., and MACREADY, G.B. (1988), “Concomitant-Variable Latent-Class Models”, *Journal of the American Statistical Association*, 83(401), 173–178.
- DEMIDENKO, E. (2007), “Sample Size Determination for Logistic Regression Revisited”, *Statistics in Medicine*, 26(18), 3385–3397.
- DEMIDENKO, E. (2008), “Sample Size and Optimal Design for Logistic Regression with Binary Interaction”, *Statistics in Medicine*, 27(1), 36–46.
- FORCINA, A. (2008), “Identifiability of Extended Latent Class Models with Individual Covariates”, *Computational Statistics & Data Analysis*, 52(12), 5263–5268.
- FORMANN, A.K. (1982), “Linear Logistic Latent Class Analysis”, *Biometrical Journal*, 24(2), 171–190.
- GOODMAN, L.A. (1974), “Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models”, *Biometrika*, 61(2), 215–231.
- HAGENAARS, J.A. (1988), “Latent Structure Models with Direct Effects Between Indicators Local Dependence Models”, *Sociological Methods & Research*, 16(3), 379–405.
- HIRTENLEHNER, H., STARZER, B., and WEBER, C. (2012), “A Differential Phenomenology of Stalking Using Latent Class Analysis to Identify Different Types of Stalking Victimization”, *International Review of Cictimology*, 18(3), 207–227.
- HOLT, J.A., and MACREADY, G.B. (1989), “A Simulation Study of the Difference Chi-Square Statistic for Comparing Latent Class Models Under Violation of Regularity Conditions”, *Applied Psychological Measurement*, 13(3), 221–231.
- KEEL, P.K., FICHTER, M., QUADIEG, N., BULIK, C.M., BAXTER, M.G., THORNTON, L., HALMI, K.A., KAPLAN, A.S., STROBER, M., WOODSIDE, D.B., et al. (2004), “Application of a Latent Class Analysis to Empirically Define Eating Disorder Phenotypes”, *Archives of General Psychiatry*, 61(2), 192.
- LANZA, S.T., COLLINS, L.M., LEMMON, D.R., and SCHAFER, J.L. (2007), “Proc LCA: A SAS Procedure for Latent Class Analysis”, *Structural Equation Modeling*, 14(4), 671–694.
- LAZARSELD, P. (1950), “The Logical and Mathematical Foundation of Latent Structure Analysis and the Interpretation and Mathematical Foundation of Latent Structure Analysis”, in *Measurement and Predictions*, eds. S.A. Stouffer et al., pp. 362–472.

- LINZAR, D.A., and LEWIS, J.B. (2011), “poLCA: An r Package for Polytomous Variable Latent Class Analysis”, *Journal of Statistical Software*, 42(10), 1–29.
- MAGIDSON, J., and VERMUNT, J.K. (2004), “Latent Class Models”, in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan, Thousand Oaks CA: Sage, pp. 175–198.
- MANN, H.B., and WALD, A. (1943), “On Stochastic Limit and Order Relationships”, *The Annals of Mathematical Statistics*, 14(3), 217–226.
- MCCUTCHEON, A.L. (1987), *Latent Class Analysis*, Newbury Park CA: SAGE Publications.
- MCHUGH, R.B. (1956), “Efficient Estimation and Local Identification in Latent Class Analysis”, *Psychometrika*, 21(4), 331–347.
- MCLACHLAN, G., and PEEL, D. (2000), *Finite Mixture Models*, New York: John Wiley.
- MUTHÉN, L.K., and MUTHÉN, B.O. (2012), *Mplus. The Comprehensive Modelling Program for Applied Researchers: Users Guide 5*, Los Angeles CA: Muthén & Muthén.
- NAKAGAWA, S., and FOSTER, T.M. (2004), “The Case Against Retrospective Statistical Power Analyses with an Introduction to Power Analysis”, *Actathologica*, 7(2), 103–108.
- NYLUND, K.L., ASPAROUHOV, T., and MUTHÉN, B. O. (2007), “Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study”, *Structural Equation Modeling*, 14(4), 535–569.
- Ó BRIEN, R.G. (1986), “Using the SAS System to Perform Power Analyses for Log-Linear Models”, in *Proceedings of the 11th Annual SAS Users Group Conference*, pp. 778–784.
- RENDER, R. (1981), “Note on the Consistency of the Maximum Likelihood Estimate for Non-Identifiable Distributions”, *The Annals of Statistics*, 9(1), 225–228.
- RENCHEA, A.C. (2000), *Linear Models in Statistics*, New York: John Wiley.
- RINDSKOPF, D., and RINDSKOPF, W. (1986), “The Value of Latent Class Analysis in Medical Diagnosis”, *Statistics in Medicine*, 5(1), 21–27.
- TOFIGHI, D., and ENDERS, C.K. (2008), “Identifying the Correct Number of Classes in Growth Mixture Models”, *Advances in Latent Variable Mixture Models*, 317–341.
- UEBERSAX, J.S., and GROVE, W.M. (1990), “Latent Class Analysis of Diagnostic Agreement”, *Statistics in Medicine*, 9(5), 559–572.
- VERMUNT, J.K. (1996), *Log-Linear Event History Analysis: A General Approach with Missing Data, Latent Variables, and Unobserved Heterogeneity*, Tilburg: Tilburg University Press.
- VERMUNT, J.K. (1997), *LEM: A General Program for the Analysis of Categorical Data*, Tilburg: Tilburg University.
- VERMUNT, J.K. (2010a), “Latent Class Modeling with Covariates: Two Improved Three-Step Approaches”, *Political Analysis*, 18(4), 450–469.
- VERMUNT, J.K. (2010b), “Latent Class Models”, in *International Encyclopedia of Education*, 7, eds. P. Peterson, E. Baker, and B. McGaw, pp. 238–244.
- VERMUNT, J.K., and Magidson, J. (2013a), *LG-Syntax User's Guide: Manual for Latent GOLD 5.0 Syntax Module*, Belmont MA: Statistical Innovations Inc.
- VERMUNT, J.K., and Magidson, J. (2013b), *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*, Belmont MA: Statistical Innovations Inc.
- WALD, A. (1943), “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large”, *Transactions of the American Mathematical Society*, 54(3), 426–482.

- WHITE, H. (1982), "Maximum Likelihood Estimation of Misspecified Models ", *Econometrica: Journal of the Econometric Society*, 50(1), 1–25.
- WHITTEMORE, A.S. (1981), "Sample Size for Logistic Regression with Small Response Probability ", *Journal of the American Statistical Association*, 76(373), 27–32.
- WOLFE, J.H. (1970), "Pattern Clustering by Multivariate Mixture Analysis ", *Multivariate Behavioral Research*, 5(3), 329–350.
- YANG, C.C. (2006), "Evaluating Latent Class Analysis Models in Qualitative Phenotype Identification ", *Computational Statistics and Data Analysis*, 50(4), 1090–1104.
- YANG, I., and BECKER, M.P. (1997), "Latent Variable Modeling of Diagnostic Accuracy", *Biometrics*, 53(3) 948–958.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.