Australian & New Zealand Journal of Statistics

Check for updates

Aust. N. Z. J. Stat. **65**(3), 2023, 213–233 doi: 10.1111/anzs.12396

Embedding latent class regression and latent class distal outcome models into cluster-weighted latent class analysis: a detailed simulation experiment

Roberto Di Mari[1], Antonio Punzo[1] and Zsuzsa Bakk[2]

University of Catania and Leiden University

Summary

Usually in latent class (LC) analysis, external predictors are taken to be cluster conditional probability predictors (LC models with external predictors), and/or score conditional probability predictors (LC regression models). In such cases, their distribution is not of interest. Class-specific distribution is of interest in the distal outcome model, when the distribution of the external variables is assumed to depend on LC membership. In this paper, we consider a more general formulation, that embeds both the LC regression and the distal outcome models, as is typically done in cluster-weighted modelling. This allows us to investigate (1) whether the distribution of the external variables differs across classes, (2) whether there are significant direct effects of the external variables on the indicators, by modelling jointly the relationship between the external and the latent variables. We show the advantages of the proposed modelling approach through a set of artificial examples, an extensive simulation study and an empirical application about psychological contracts among employees and employers in Belgium and the Netherlands.

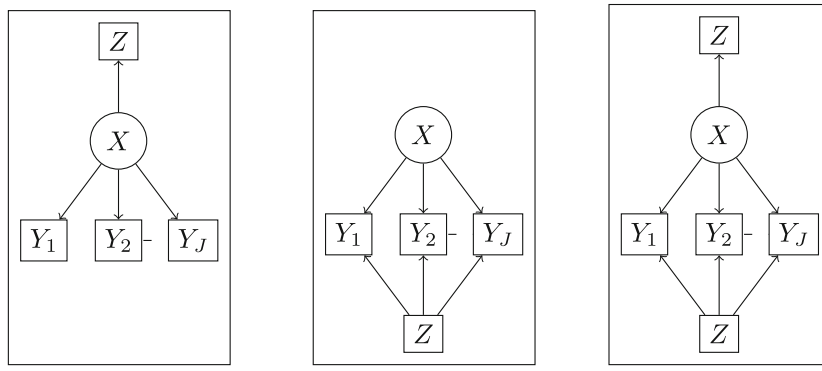
Key words: cluster-weighted models; continuous distal outcomes; direct effects; latent class analysis; latent class regression models; psychological contracts.

1. Introduction

Latent class (LC) analysis (McCutcheon 1985) is a model-based clustering technique that is very popular in the social and behavioural sciences, economics and health sciences alike. The approach is used to cluster a set of observed categorical variables (known as indicators) based on an underlying latent variable. Some example applications include patterns of mobile internet usage for travelling (Okazaki *et al.* 2015), or, from health sciences, types of treatment engagements among adolescents with severe psychiatric problems, or change over time in nursing patterns (Roberts & Ward 2011).

In most instances, similarly to other latent variable models (such as factor analysis or item response theory), the interest of the researchers is not merely in clustering, but also in relating the clustering to antecedents and consequences in larger models. Some examples of modelling the consequences of the clustering include assessing recidivism rate among

*Author to whom correspondence should be addressed.
[1]Department of Economics and Business, University of Catania, Corso Italia 55, Catania 95128, Italy. e-mail: roberto.dimari@unict.it
[2]Statistics Unit, Methodology Institute of Psychology, Leiden University, Wassenaarseweg 52, Leiden 2333 AK, The Netherlands.



(a) Latent class model with distal outcome

(b) Latent class regression model

(c) Latent class cluster-weighted model

Figure 1. The three competing models: LC distal outcome (a), LC regression (b) and LC cluster-weighted (c) models.

clusters of juvenile offenders (Mulder *et al.* 2012), or predicting distal pain outcomes from clusters of pain management (Roberts & Ward 2011).

Historically, in LC analysis, predictors are added to the model using simultaneous estimation based on the approach introduced by Dayton & Macready (1988). The inclusion of consequences, usually called *distal outcomes* (for a graphical representation see Figure 1a), is more problematic due to often strong distributional assumptions about the outcome variables. When such assumptions are violated, the LC solution can be distorted, and thus comparison with other models may become meaningless (Vermunt 2010; Bakk, Tekle, & Vermunt 2013). Such circumstances may as well lead to over-extraction of classes (Bauer & Curran 2003).

To overcome these problems, stepwise estimators have been proposed in literature, which allow separating the estimation of the measurement from structural model (Bakk, Tekle, & Vermunt 2013; Lanza, Tan, & Bray 2013; Asparouhov & Muthén 2014). While stepwise approaches are currently the best practice in literature, they hinder detection and testing of direct effects between the indicators of the LC and external variables.

Typically, in distal outcome models (LCdist), the distal outcome Z and $\mathbf{Y}$ are assumed to be conditionally independent given the latent variable X (Bakk, Tekle, & Vermunt 2013; Lanza, Tan, & Bray 2013). Direct effects of Z on $\mathbf{Y}$ are therefore not allowed for—neither their presence tested. Notably, maximum likelihood (ML) estimation of latent variable models is subject to severe bias when unmodelled direct effects are present in LC and latent trait models (Asparouhov & Muthén 2014), regression mixture models (Kim *et al.* 2016; Nylund-Gibson & Masyn 2016), and latent Markov models (Di Mari & Bakk 2018; Di Mari *et al.* 2022) when unmodelled direct effects are present in LC and latent trait models, and these are not accounted for.

Given the restrictiveness of the conditional independence assumption, and the possible severity of its violation, we showcase a more general model that can account for complex interdependencies between the external variable, LC membership and the indicators of the LC model. In regression mixtures, a 'circular' relation among $\mathbf{Y}$-X-Z is typically considered

in the cluster-weighted modelling approach (Ingrassia, Minotti, & Vittadini 2012; Ingrassia *et al.* 2015; Mazza, Punzo, & Ingrassia 2018). That is, a more general model is specified, where next to modelling the class-specific distribution of Z (distal outcome situation), also the direct effect of Z on $\mathbf{Y}$ is modelled (LC regression). Then, standard inference tools can assess the statistical significance of each effect.

In LC regression models (LCreg; Kamakura & Russel 1989; Wedel & DeSarbo 1994), although the assumption of conditional independence of $\mathbf{Y}$ and Z can be relaxed (see Fig. 1b), the distal outcome's distribution is not of interest, and hence not modelled. Therefore, in the traditional latent class analysis (LCA) approach, an external variable enters the model either as a predictor (LC regression) or as a distal outcome, but never as both at the same time. The cluster-weighted approach, in the context of LCA, allows embedding the models in Figure 1a,b into a single one, as depicted in Figure 1c. Although this idea is not new in LCA (Di Mari, Bakk, & Punzo 2020), the present paper demonstrates that the assumptions of both distal outcome and LC regression models can be tested within LC cluster-weighted (LCcw) umbrella.

Di Mari, Bakk, & Punzo (2020) already proposed the LC cluster-weighted approach, presenting both simultaneous as well as two-step estimators. They provided a detailed simulation study under different levels of violation of the assumption of local independence, and of the distributional assumptions of the distal outcome (that are known to affect the distal outcome model estimates). The simulation results showed the superiority of the LCcw approach when compared to the distal outcome and the LC model with covariates, when the underlying model assumptions of the latter two models are violated.

In this paper, we broaden the scope of cluster-weighted modelling in LCA by bringing out a significant theoretical and practical connection with the LC regression model—the primary competitor for assessing the relationship between $\mathbf{Y}$ and Z. Furthermore, we deal with the important issue of model selection, especially relevant when the underlying model assumptions are violated. We focus on how to handle direct effects with the different approaches—whereby recommending readers to Di Mari, Bakk, & Punzo (2020) for the treatment of violations of distributional assumptions of Z in distal outcome models.

We will show evidence, based on a set of artificial examples, an extensive simulation study and an empirical application, that (1) if direct effects are present, our approach, contrary to the distal outcome model, yields unbiased estimates of the distal outcome cluster specific means and variances; and (2) if the most suitable model is one between the distal outcome model or the LC regression model, the relative class sizes and compositions will be the same as the ones delivered under the proposed cluster-weighted modelling approach.

The paper proceeds as follows. In Section 2, we give model definitions and details on the parameterisations. We illustrate the proposed modelling approach through three artificial examples for both one and two external variable cases, for a total of six artificial examples, in Section 3. We present the design and the results of the simulation study in Section 4. In Section 5, we analyse data from the PSYCONES survey (Isaksson *et al.* 2003), and conclude with some final remarks in Section 6.

2. Latent class analysis with external variables: three different approaches

2.1. The latent class cluster-weighted model

Let $\mathbf{Y} = (Y_1, \ldots, Y_J)^\top$ be the vector of the full response pattern and $\mathbf{y}$ its realisation. Let us assume also one continuous external variable Z is available, and we denote as z its

realisation. Let us denote as X the categorical latent variable, with LCs $s = 1, \ldots, S$. A general form of association between $\mathbf{Y}$, X and Z, involves modelling the following joint probability

$$\Pr(Z = z, X = s, \mathbf{Y} = \mathbf{y}) = \Pr(Z = z, X = s) \Pr(\mathbf{Y} = \mathbf{y} | Z = z, X = s), \tag{1}$$

where the common assumption in LCA of $\mathbf{Y}$ and Z being conditionally independent given the latent process is relaxed. From (1), several submodels can be specified (predictor model, distal outcome model, etc). If substantive theoretical arguments postulate the latent variable to be a predictor of the external variable Z, the LC cluster-weighted model specifies the probability of observing a response pattern $\mathbf{y}$ as

$$\Pr(\mathbf{Y} = \mathbf{y}, Z = z) = \sum_{s=1}^{S} \underbrace{\Pr(X = s)}_{\text{a}} \underbrace{\Pr(Z = z | X = s)}_{\text{b}} \underbrace{\Pr(\mathbf{Y} = \mathbf{y} | Z = z, X = s)}_{\text{c}}, \tag{2}$$

which is defined by three components: the structural component (a), describing the LC variable, the external variable model (b), modelling the LC-specific distribution of Z and a measurement component (c), connecting the LC to the observed responses with a direct effect of Z. Under the assumption of local independence of response variables given the class membership and Z, the conditional distribution of the responses can be written as

$$\Pr(\mathbf{Y} = \mathbf{y} | Z = z, X = s) = \prod_{j=1}^{J} \Pr(Y_j = y_j | Z = z, X = s). \tag{3}$$

For estimating the model in (2), we assume each Y_j to be conditionally Bernoulli distributed, with success probability π_{sj}, and parametrise the conditional response probabilities through the following log-odds

$$\log\left(\frac{\pi_{js}}{1 - \pi_{js}}\right) = \alpha_{js} + Z\beta_{js}, \tag{4}$$

whereby Z is assumed to be conditionally Gaussian with mean μ_s and variance σ_s^2, for $1 \leq s \leq S$.

The model in (2) can be used to assign observations to clusters based on the posterior membership probabilities

$$\Pr(\mathrm{X} = s | \mathbf{Y} = \mathbf{y}, Z = z) = \frac{\Pr(X = s) \Pr(Z = z | X = s) \Pr(\mathbf{Y} = \mathbf{y} | Z = z, X = s)}{\Pr(\mathbf{Y} = \mathbf{y}, Z = z)}, \tag{5}$$

according to, for instance, modal or proportional assignment rules.

The LC unconditional probabilities can as well be parametrised using logistic regressions. We opt for the following parametrisation

$$\log\left(\frac{\Pr(X = s)}{\Pr(X = 1)}\right) = \theta_s, \tag{6}$$

for $1 < s \leq S$, where we take the first category as reference, and we set to zero the related parameter. The total number of free parameters to be estimated is therefore $J \times S$ (random

intercepts) + $J \times S$ (random slopes) for the measurement model, S (means) + S (variances) for the external variable model, and $S - 1$ random intercepts for the structural model.

Notice that, by setting the β_{js}s of (4) to zero, the LC cluster-weighted model reduces to a standard LC with distal outcome model. In contrast, given that the external variable component is completely missing, the LC regression is not formally nested in the LC cluster-weighted model, although it can be thought of as a sub-model in which the conditional distribution of $\mathbf{Y}|Z$ is modelled, and Z is taken as fixed value rather than a random variable to be modelled as well.

2.2. The LC with distal outcome model

It is common, in LCA, to consider a less general version of the joint distribution of (1), by assuming the responses and Z to be conditionally independent given the latent process. If, again, the LC variable is taken to be a predictor of the external variable Z, this yields the following LC with distal outcome model

$$\Pr(\mathbf{Y} = \mathbf{y}, Z = z) = \sum_{s=1}^{S} \Pr(X = s) \Pr(Z = z|X = s) \Pr(\mathbf{Y} = \mathbf{y}|X = s). \tag{7}$$

Under the local independence assumption of the indicators given the LC variable, the response conditional probabilities can be written, similarly to (3), as

$$\Pr(\mathbf{Y} = \mathbf{y}|X = s) = \prod_{j=1}^{J} \Pr(Y_j = y_j|X = s), \tag{8}$$

and parametrised through the following log-odds

$$\log\left(\frac{\pi_{js}}{1 - \pi_{js}}\right) = \alpha_{js}. \tag{9}$$

The model of (7) can be used to cluster observations, according to modal or proportional assignment rules, based on the following posterior membership probabilities

$$\Pr(\mathrm{X} = s|\mathbf{Y} = \mathbf{y}, Z = z) = \frac{\Pr(X = s) \Pr(Z = z|X = s) \Pr(\mathbf{Y} = \mathbf{y}|X = s)}{\Pr(\mathbf{Y} = \mathbf{y}, Z = z)}. \tag{10}$$

The external variable Z is assumed, conditional to the LC, to be Gaussian with mean μ_s and variance σ_s^2, $s = 1, \ldots, S$, whereby the LC unconditional probabilities are parametrised as in (6). This yields $J \times S + 2S + S - 1$ free parameters to be estimated. The only difference with model (2) is that $\mathbf{Y}$, in the measurement component, is conditional only on X, not on Z. That is, $\mathbf{Y}$ is assumed to be independent of Z given X, which is a standard, and rather very strong, assumption of LCA.

2.3. The LC regression model

Rather than modelling the joint distribution $\Pr(Z, X, \mathbf{Y})$ in (1), the LC regression models the conditional distribution of $\mathbf{Y}$ given Z and the LC variable, specifying the following model for $\mathbf{Y}$:

Table 1. Summary of different modelling assumptions and number of free parameters to be estimated.

	Dir. Eff.	Z modelled	#par
Latent class regression	✓	×	$2(J \times S) + (S - 1)$
Latent class with distal outcome	×	✓	$(J \times S) + 2S + (S - 1)$
Latent class cluster weighted	✓	✓	$2(J \times S) + 2S + (S - 1)$

$$\Pr(\mathbf{Y} = \mathbf{y}|Z = z) = \sum_{s=1}^{S} \Pr(X = s) \Pr(\mathbf{Y} = \mathbf{y}|Z = z, X = s). \tag{11}$$

In this case, the conditional response probabilities depend on the external variable Z and, under local independence of the responses given the latent variable, the measurement model can be written as in (3), and parametrised as in (4). The posterior membership probabilities, computed based on the model in (11), are as follows

$$\Pr(\mathrm{X} = s|\mathbf{Y} = \mathbf{y}, Z = z) = \frac{\Pr(X = s) \Pr(\mathbf{Y} = \mathbf{y}|Z = z, X = s)}{\Pr(\mathbf{Y} = \mathbf{y}, Z = z)}. \tag{12}$$

With LC unconditional probabilities parametrised as in (6), the total number of free parameters to be estimated is $2(J \times S) + S - 1$.

Table 1 summarises how Z enters each of the three models, and the total number of free parameters to be estimated. Intuitively, this shows that the first two models can be seen as special cases of the third model, which therefore models the relationship between the three sets of variables in the most exhaustive manner.

2.4. Maximum likelihood estimation

From (2), (7) and (11), given an observed sample of n units, it is possible to specify the log-likelihood functions of the cluster-weighted, the distal outcome and the LC regression models. Iterative procedures like the expectation–maximisation or Newton-type algorithms, or a combination of both as implemented in Latent GOLD 5.1 (Vermunt & Magidson 2013), can be used to maximise each lo–likelihood function with respect to the model parameters in order to find the maximum likelihood estimates.

Specifically, Latent GOLD starts the model fitting with the EM algorithm, and then switches to Newton–Raphson for the last few iterations. This is done to take the best of both worlds. Namely, to exploit the stability of the EM, while retaining the speed of Newton–Raphson when close to convergence (Vermunt & Magidson 2016).

Latent GOLD's EM algorithm does posterior mode estimation with iterative proportional fitting. The M step is instead a Newton-type step, therefore the log-likelihood is increased (and not maximised). In models including multivariate Gaussian variables (such as the distal outcome model), the Fisher scoring algorithm is only used for M steps when the covariance-structure parameters have no closed-form solution. Readers interested in more details of the estimation approach used by Latent GOLD can refer to (Vermunt 1997; Vermunt & Magidson 2013). Alternative EM implementations are also available for LC models with external variables. See, for instance, Durante, Canale, & Rigon (2019).

The log-likelihood surface, as in all mixture models, presents many local maxima (McLachlan & Peel 2004). Thus, appropriate starting values can be crucial to find a

meaningful maximiser. We use Latent GOLD 5.1 for model fitting, which implements a well-refined initialisation strategy (details can be found in Vermunt & Magidson 2016, Sec. 7.8).

2.5. Evaluating the latent class solution

A commonly used measure for class separation in LCA, and thus also for classification error, is the entropy-based R^2 (Magidson 1981). For response pattern i, this quantifies how much the posterior membership probabilities deviate from uniformity by using the principle of entropy as follows

$$E_i = \sum_{s=1}^{S} - \Pr(X_i = s|\mathbf{Y}_i = \mathbf{y}_i) \log \Pr(X_i = s|\mathbf{Y}_i = \mathbf{y}_i), \tag{13}$$

where $\Pr(X_i = s|\mathbf{Y}_i = \mathbf{y}_i)$ are the posterior class membership probabilities for the simple LC model. An analogous measure of entropy can be computed based on the posteriors from the LC regression, the LC with distal outcome and the cluster-weighted LC models.

The total entropy, when no information on the response (and potentially the external variables) is used to predict X, is defined as

$$E_{\text{TOT}} = \sum_{s=1}^{S} - \widehat{p}_s \log \widehat{p}_s, \tag{14}$$

where $\widehat{p}_s$ is an estimate of $\Pr(X = s)$. Equation (14) considers only the estimated marginal class probabilities rather than the posterior class membership probabilities as in (13). The proportional reduction of entropy when $\mathbf{y}$ is available compared to the situation in which $\mathbf{y}$ is unknown is an (entropy based) R^2 measure for class separation—as well as for the quality of the classification of a sample—and is defined as

$$\text{R}^2_{\text{entr}} = \frac{E_{\text{TOT}} - \bar{E}}{E_{\text{TOT}}}, \tag{15}$$

where $\bar{E} = n^{-1} \sum_{i=1}^{n} E_i$.

3. Artificial data examples

To substantiate the benefits of the cluster-weighted modelling approach in LCA, in this Section we propose an artificial data analysis on some exemplary LCA contexts. In particular, we analyse six large data sets (30,000 sample units), three with 1 (block 1) and the other three with two external variables (block 2). Under both blocks, we generate data from each of the three models in Figure 1a–c.

In either block, we set the number of LCs to $S = 2$, and to begin with we fit all three models assuming this value to be known. Then, we also show results on estimation of the number of LCs based on BIC. The data were generated in R (R Core Team 2017), and parameter estimation was carried out with Latent GOLD 5.1 (Vermunt & Magidson 2016).

To get approximately equal (realistic) conditions on class separation, we generated the data such that the entropy-based R^2 (Magidson 1981) for the correctly specified model is

Table 2. *LCreg* data.

	Class proportions		Entr. R^2	#par
LCreg	**0.7010**	**0.2990**	**0.7675**	**25**
LCdist	0.7357	0.2643	0.8639	17
LCcw	0.7018	0.2982	0.7681	29

Notes: Estimated class proportions, entropy-based R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table 3. *LCreg* data.

	Means		Wald(=) P	Variances		Wald(=) P
LCdist	0.0525***	−0.1640***	0.0000	1.0301	0.8846	0.0000
	(0.0071)	(0.0122)		(0.0100)	(0.0158)	
LCcw	−0.0010	−0.0134	0.9000	0.9966	1.0105	0.6100
	(0.0086)	(0.0163)		(0.0114)	(0.0208)	

Notes: Estimated means (***P-value<0.001, **P-value<0.01, *P-value<0.05) and variances, and P-values from Wald test of equality of component means and variances for the LCdist model and the LCcw model. Standard errors in parentheses. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

about 0.7 in all the three data sets—which is the minimum class separation to get a good LC model (Vermunt 2010; Asparouhov & Muthén 2014). Below we present results from Block 1. Results obtained with two external variables are comparable, and reported in the Appendix.

3.1. LC regression (*LCreg*) data

The *LCreg* data set was generated from a two-class LCreg model, with class memberships of 0.7 and 0.3, six dichotomous indicators ($J = 6$) and one continuous Z—drawn from a standard normal distribution—loaded on all six indicators. The external variable Z is loaded on the indicators with a coefficient of −0.5, if the most likely response is on the first class, or 1, if the most likely response is on the second class, giving a large effect size. In classical LCA this is known as differential item functioning (DIF), a violation of the conditional independence assumption often ignored (Masyn 2017; see also Kankaraš, Moors, & Vermunt 2010; Lee, Bulut, & Suh 2017; Suh & Cho 2014; Woods 2009).

We observe in Table 2 that the LCdist model overinflates the mixing proportion on the bigger class, whereas the LCcw model yields nearly equivalent class proportions as in the correctly specified case. This at the cost of four more parameters to be estimated.

Table 3 reports estimated means and variances for the variable Z based on the LCdist and LCcw models, along with standard errors and P-values of the Wald tests of equality of the means and the variances. Nothing is reported for LCreg, as Z is not modelled. In the LCdist model, both means are wrongly estimated to be statistically different from zero. Moreover, based on the reported Wald tests, we reject the nulls of equal means and equal variances (with P-values smaller than 0.01). These findings for the LCdist model can be explained by the fact that it wrongly predicts a clustered distribution on Z in order to accommodate for a direct effect of Z on the indicators which is not accounted for. This creates an additional source of entropy in the class solution (as displayed by the relatively higher value of the entropy-based R^2).

Table 4. *LCdist* data.

	Class proportions		Entr. R^2	#par
LCreg	0.5850	0.4150	0.2781	25
LCdist	**0.7006**	**0.2994**	**0.7274**	**17**
LCcw	0.7026	0.2974	0.7320	29

Notes: Estimated class proportions, entropy R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table 5. *LCdist* data.

	Means		Wald(=) *P*	Variances		Wald(=) *P*
LCdist	**−0.9911*****	**1.0156*****	**0.0000**	**1.0072**	**0.9886**	0.4000
	(0.0084)	(0.0145)		(0.0119)	(0.0196)	
LCcw	−0.9889***	1.0242***	0.0000	1.0075	0.9810	0.2700
	(0.0096)	(0.0176)		(0.0125)	(0.0207)	

Notes: Estimated means (****P*-value<0.001, ***P*-value<0.01, **P*-value<0.05) and variances, and *P*-values from Wald test of equality of component means and variances for the LCdist model and the LCcw model. Standard errors in parentheses. Results from correctly specified model in bold font. Reported *P*-values are function of the unobserved LC variable, and are therefore approximate.

3.2. LC distal outcome (*LCdist*) data

The *LCdist* data set was generated from a two-class LCdist model, with class memberships of 0.7 and 0.3, six dichotomous indicators ($J = 6$) and one continuous Z, drawn from a mixture of two normal distributions with means of -1 and 1 and common variance of 1.

The LCreg model yields a completely distorted class solution, whereby both the LCdist and LCcw models yield almost identical (correct) solutions (Table 4). Interestingly, the misspecified response-Z relation in the LCreg model yields a solution with relatively smaller class separation (as measured by the entropy-based R^2).

Next, we compare estimates of class-specific means and variances of Z as obtained by the LCdist and LCcw models (Table 5).

The LCdist and the LCcw models estimate almost identical means and variances of Z, both correctly not rejecting the null of common variance across LCs. We observe that the SE's for the less parsimonious LCcw model are systematically larger than those of the correctly specified model: this is not surprising, as having less degrees of freedom corresponds, all else equal, to slightly more variable estimates.

3.3. LC cluster-weighted (*LCcw*)data

The *LCcw* data were generated from a two-class LCcw model, with class memberships of 0.7 and 0.3, six dichotomous indicators ($J = 6$) and one continuous Z, drawn from a mixture of two normal distributions with means of -1 and 1 and common variance of 1.

Both the LCreg and the LCdist models deliver distorted class solutions (Table 6). Although with a higher entropy-based R^2, the residual dependence among the indicators due to the exclusion of the direct effect causes a more severe distortion in the LCdist model compared to the LCreg model. Also relative class sizes are overturned for LCdist. Equally we observe (Table 7) that the means and variance(s) of Z are both biased in the LCdist

Table 6. *LCcw* data.

	Class proportions		Entr. R^2	#par
LCreg	0.8899	0.1101	0.5611	25
LCdist	0.4373	0.5627	0.6441	17
LCcw	**0.6993**	**0.3007**	**0.7045**	**29**

Notes: Estimated class proportions, entropy-based R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table 7. *LCcw* data.

	Means		Wald(=) *P*	Variances		Wald(=) *P*
LCdist	−1.4544***	0.4159***	0.0000	0.7044	1.2096	0.0000
	(0.0102)	(0.0120)		(0.0111)	(0.0159)	
LCcw	**−1.0029*****	**0.9955*****	**0.0000**	**1.0215**	**0.9818**	**0.0380**
	(0.0104)	(0.0122)		(0.0146)	(0.0160)	

Notes: Estimated means (***P-value<0.001, **P-value<0.01, *P-value<0.05) and variances, and P-values from Wald test of equality of component means and variances for the LCdist model and the LCcw model. Standard errors in parentheses. Results from correctly specified model in bold font. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

Table 8. Adjusted Rand indexes, computed between clustering with correctly specified models—LCreg, LCdist and LCcw models—and clustering with the other two models.

		Fitted model		
Data	Correct model	LCreg	LCdist	LCcwm
LCreg	LCreg	1	0.9732	0.9997
LCdist	LCdist	0.2125	1	0.9889
LCcw	LCcw	0.1604	0.3101	1

model. Contrary to the correctly specified model, in LCdist the Wald test cannot reject the equal variances hypothesis (at 1 % level).

Table 8 reports adjusted Rand indexes (ARIs) (Hubert & Arabie 1985), arranged in a three-by-three table, comparing the hard partitions obtained with each fitted model under the three data generating model scenarios. The results are in line with what was observed above. When the data are generated with a LCreg model, the LCcw model delivers an almost identical partition to that of the correctly specified model, followed close up by the LCdist model—with an ARI of ≈ 0.97. In the *LCdist* data set as well, the LCcw model's partition is nearly as in the correctly specified model (ARI of ≈ 0.99), whereby the ARI drops to ≈ 0.21 when the comparison is with the LCreg partition. In the latest scenario—*LCcw* data set—fitting both the LCreg and the LCdist models delivers in both cases quite different partitions (≈ 0.16 and ≈ 0.31 ARIs) compared to the correctly specified model.

Based on the above data sets, in Table 9 we report also results on BIC values for the three models in all three scenarios, for $S = 1, \ldots, 5$. Although BIC values can be compared for LCdist and LCcw, selecting a model among the three with BIC cannot be done as Z in LCreg is not modelled and the model likelihoods are therefore not comparable. In both the *LCreg* and *LCdist* data sets, BIC for the LCcw model selects, together with the correctly specified model in the first two scenarios, the correct number of classes.

Table 9. Model selection with BIC computed for each model at each data generating model—LCreg, LCdist and LCcw—for $S = 1, \ldots, 5$.

Data		Number of components				
		$S = 1$	$S = 2$	$S = 3$	$S = 4$	$S = 5$
LCreg	**LCreg**	235,957.23	**191,309.94**	191,411.62	191,539.73	191,642.90
	LCdist	321,109.39	283,926.34	278,042.32	277,453.71	**277,117.53**
	LCcw	321,137.31	**276,509.33**	276,636.54	276,766.45	276,888.98
LCdist	LCreg	233,652.06	231,508.46	231,251.75	231,041.07	**230,893.17**
	LCdist	348,714.25	**331,953.90**	332,037.14	332,112.24	332,189.69
	LCcw	337,204.58	**332,067.91**	332,198.43	332,329.23	332,449.70
LCcw	LCreg	213,395.81	207,054.91	206,149.45	205,996.43	**205,922.50**
	LCdist	321,792.35	310,231.51	306,431.47	305,466.02	**304,613.18**
	LCcw	316,998.91	**303,623.67**	303,759.81	303,910.04	304,025.80

Notes: Data generating model and minimum BIC value, for each model at each scenario, in bold.

Interestingly, however, misspecifying the indicators-Z relation causes, in both the LCreg and LCdist models, a severe overstatement of the number of classes (*LCcw* data set).

4. Simulation study

4.1. Design

This simulation study is designed to assess the cluster-weighted LC approach under varying conditions on sample size, class size and class separation on both the indicators and the continuous variable means. Using the same terminology of the artificial examples, we generate data from a two-class LC model with six indicators ($J = 6$) and one external variable under the three model specifications *LRreg*, *LCdist* and *LCcw*. This allows us to give a fair account of each model's performance both in the correct and in the incorrect model specification cases.

As target measures, we consider estimated class proportions to assess the clustering performance, and estimated means for the external variable (only for LCdist and LCcw). We manipulate class separation through two channels: (1) by altering the response probabilities that control the strength of the relationship between indicators and LCs; and (2) by varying the class conditional means of the external variable.

Regarding (1), we consider two levels, 0.65 and 0.9, corresponding to entropy-based R^2 of about 0.7 and 0.9. Related to (2), the external variable, under the *LCdist* and *LCcw* data generating models, is sampled from a clustered normal distribution with unit variance and mean depending on class membership: the values are set to -1 and 1, -2 and 2 and -3 and 3, corresponding to an increasing separation level.

Under the *LCreg* data generating model, Z is sampled from a standard normal. The sample sizes used are 500 and 2000, and the relative class sizes equal to 0.5 (equal class sizes) and 0.3 (unequal class sizes). The regression coefficient β_{js} (*LCreg* and *LCcw* data generating models) is set to 0.5, which corresponds to a moderate–strong magnitude on the logistic scale. For the resulting 56 simulation crossed scenarios, obtained by combining the conditions on sample and class sizes, class separation on the indicators and the external variable and data generating models, we generate 250 data sets. Data generation is done within R (R Core Team 2022), whereas model estimation is carried out using Latent GOLD 5.1 (Vermunt & Magidson 2016) in combination with R.

Table 10. *Relative absolute bias* for the *class proportion* of the first (smallest) class, with average Monte Carlo standard deviations in parentheses, per data generating model, averaged over simulation condition.

	Model Specification		
Data	LCreg	LCdist	LCcwm
LCreg	0.025	0.119	0.033
	(0.023)	(0.046)	(0.028)
LCdist	0.243	0.003	0.008
	(0.080)	(0.018)	(0.026)
LCcw	0.310	0.098	0.009
	(0.084)	(0.018)	(0.026)

4.2. Summary of the results

Under *LCreg*, the correctly specified model and LCcw do well (in terms of estimated class sizes) in all conditions. LCdist has the worst performance, exacerbated by small separation and equal class sizes.

With the *LCdist* data generating model, again the correctly specified model and LCcw do well in all conditions. LCreg reaches its own best performance in conditions with small separation on both indicators and external variable means (conditions 1–6). Larger separation and equal cluster size conditions is where LCreg does worst. Both LCdist and LCcw estimate well the external variable means in all conditions, with SEs recovering well the true underlying variability.

Under *LCcw*, both classical approaches LCdist and LCreg are misspecified. Interestingly, with the exception of the first four conditions (small separation on both indicators and external variable means), LCdist performs relatively better (smallest bias in estimated class sizes and smaller variance) than LCreg. LCdist estimates well the class conditional means of the external variable only in large separation conditions. More generally, we observe that class separation on the indicators has a relatively stronger impact on LCdist outcome.

In Table 10 we present relative absolute bias in the class proportion for the first (smallest) class, with average Monte Carlo standard deviations in parentheses, averaged over all the levels of separation between classes and sample size per data generating model for the three models. Except for the correctly specified case, LCreg delivers the most distorted class composition overall, with up to 30% bias under the LCcw data generating process. When a LCdist model is fitted, under *LCreg* and *LCcw*, the bias is relatively lower—between 10% and 12%. The aggregate numbers show that, on average, class distortion for the cluster-weighted model specification is below 3%. In contrast, the distal outcome model's class proportion output is estimated with relatively smaller variance than its competitors. This is not surprising, as LCdist's specification with a single distal outcome requires estimating a smaller number of free parameters than LCreg and LCcw.

Similarly, Table 11 presents the aggregate relative absolute bias for the first-class estimated distal outcome mean, averaged over simulation rounds and conditions, per data generating process. In line with what we commented above, the distal outcome's means are estimated, on average, with bias by LCdist when there is model misspecification. In our settings, this bias is, on average, of the order of 8% relative to the true parameter value.

Table 11. *Relative absolute bias* for the first (smallest) class *distal outcome mean*, with average Monte Carlo standard deviations in parentheses, per data generating model, averaged over simulation condition.

	Model specification	
Data	LCdist	LCcwm
LCdist	0.002	0.003
	(0.059)	(0.054)
LCcw	0.077	0.001
	(0.059)	(0.047)

The detailed simulation output (stacked barplots and linegraphs) can be found in the Data S1.

5. Real data application: relating LC model of psychological contract types to job insecurity

We analysed data from the Dutch and Belgian sample of the Psychological Contracts across Employment Situation (PSYCONES) project (European Commission, 2006). The sample consisted of $n = 1353$ respondents. We selected $J = 8$ indicator variables, measuring psychological contract type: the first four refer to employee obligations (whether a promise was made or not), and the last 4 to employer obligations, where each set of four indicators contained two items for relational and two for transactional obligations. Examples of the wording of indicators are: 'This organization promised me a reasonably secure job' and 'This organization promised me a good pay for the work I do'. The typology on these indicators was first proposed by Cuyper *et al.* (2008), who identified a LC model with four classes corresponding to mutual high obligations, employee over-obligation, employee under-obligation and mutual low obligations in the Belgian and German sample of the PSYCONES data. The four-class model was replicated on the Dutch and Belgian sample by Bakk, Tekle, & Vermunt (2013). We replicated the LC model proposed by Bakk, Tekle, & Vermunt (2013) and related it to perceived job insecurity measured using a scale developed by De Witte *et al.* (2000), that consists of four indicators with five categories and had a Cronbach's alpha value of 0.88. While Bakk, Tekle, & Vermunt (2013) used a stepwise distal outcome model, and thus ignored possible direct effects between the eight indicators of the LC membership and job insecurity, we re-analyse the data using cluster-weighted LC and LC regression models that allow also for possible direct effects.

In Table 12 we report the simple LC model with four classes. The respondents belonging to the mutual high obligation class have a high probability to have all employee and employer obligations, while in the over-obligation class employees have a high probability to have all obligations, while they perceive that the employer scores low on their side. The under-obligation class shows a reverse pattern: employers score high on all obligations while employees are disengaged. In the mutual low class both parties have low obligations. More detailed description of the model is available in Cuyper *et al.* (2008).

Subsequently, we investigated whether there is a difference with regard to perceived job insecurity among the four classes. To verify if the same number of classes would be

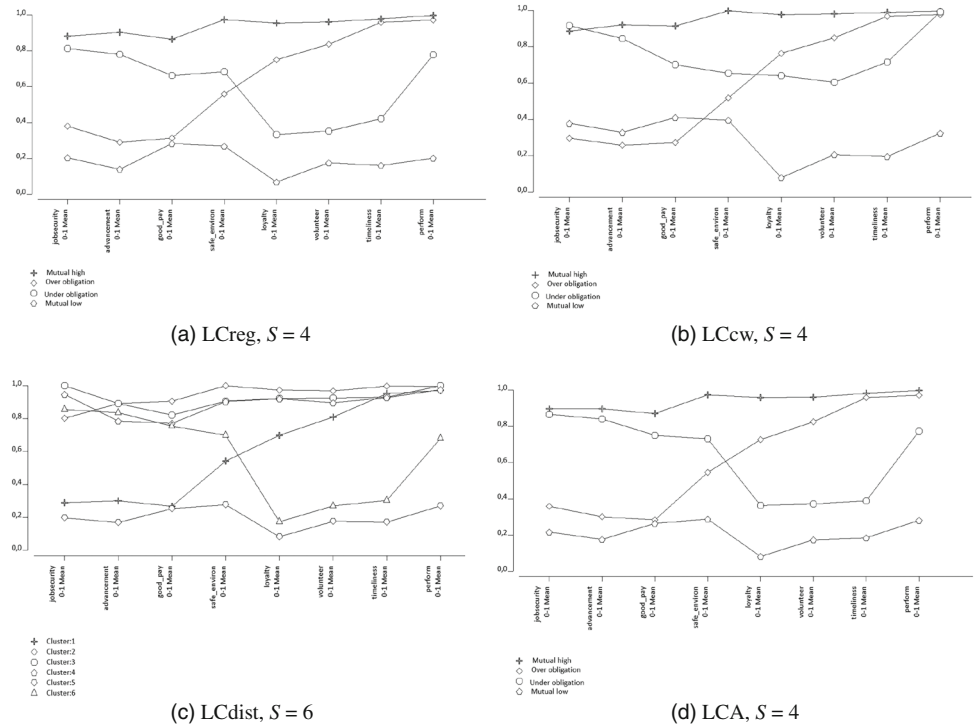
Table 12. Class proportions and class-specific probabilities of a positive response for the four-class model estimated for the PSYCONES data (Belgian and Dutch combined sample).

	Class 1 Mutual low	Class 2 Under-obligation	Class 3 Over-obligation	Class 4 Mutual high
Class proportion	0.090	0.100	0.290	0.520
Employers' obligations				
Secure job	0.210	0.870	0.360	0.900
Advancement	0.180	0.850	0.300	0.900
Good pay	0.260	0.750	0.280	0.870
Safe work environment	0.290	0.730	0.550	0.970
Employees' obligations				
Loyalty	0.080	0.360	0.730	0.980
Volunteer	0.170	0.370	0.830	0.980
On time	0.180	0.390	0.960	0.980
Good performance	0.280	0.770	0.970	0.990

Table 13. Number of components (Ncomp), BIC, number of parameters (#par), expected classification error (Class. Err.), and entropy-based (Entr.) R^2 for each of the three models, from 1 to 6 (seven in case of LCdist) components.

	S	BIC	#par	Class. Err.	Entr. R^2
LCreg	1	11,637.241	16	0	1
	2	10,447.216	33	0.074	0.724
	3	10,161.689	50	0.113	0.714
	4	10,139.188	67	0.123	0.718
	5	10,212.570	84	0.178	0.669
	6	10,296.546	101	0.116	0.764
LCdist	1	15,621.130	10	0	1
	2	14,280.901	21	0.066	0.742
	3	13,933.870	32	0.101	0.737
	4	13,847.386	43	0.189	0.682
	5	13,829,068	54	0.186	0.687
	6	13,803.998	65	0.186	0.721
	7	13,836.384	76	0.212	0.705
LCcw	1	15,491.408	18	0	1
	2	14,280.162	37	0.063	0.758
	3	13,999.164	56	0.089	0.756
	4	13,989.909	75	0.169	0.700
	5	14,006.917	95	0.147	0.739
	6	14,050.696	113	0.206	0.710

selected applying the different approaches, first we looked at model fit indices. In Table 13 we report overall fit statistics for the LCreg, LCdist and LCcw models with one to six classes for LCcw and LCreg, and till seven classes for LCdist. While with LCcw and LCreg the best fitting model would be the four-class model (with class definitions very similar to the simple LC model), with LCdist a six-class model is selected by BIC as best fitting. The six-class model breaks down the mutual high class into three smaller classes to account for the unmodelled direct effects of job insecurity and to better model its class-specific distribution. However, when using LCcw we allow for modelling the mild direct effects between job insecurity and some of the indicators, and the four-class model (that is validated in literature) stays the best fit.



(a) LCreg, $S = 4$ (b) LCcw, $S = 4$

(c) LCdist, $S = 6$ (d) LCA, $S = 4$

Figure 2. Probability profile plot for LCreg ($S = 4$), LCcw ($S = 4$), LCdist ($S = 4$) and LCA $S = 3$).

As Figure 2 shows, the definition of the class-specific response pattern on the eight indicators is comparable between the simple LCA, LCreg and LCcw, while using LCdist we obtain three classes that split up the mutual high class to better model the distribution of the distal outcome.

In Table 14, the direct effects as modelled with LCcw are reported. While these effects are mild, still modelling them improves the estimation of the relationship between the LC variable and the distal outcome. This leads to a model with four classes that, differently to the distal outcome model, has comparable classes to the simple LC model. Most noticeable is the direct effects on the openness to volunteer and the perception of a safe working environment. Higher job insecurity is associated with lower levels of volunteering in all classes but the mutual low class, while higher level of job insecurity is also associated with lower levels of perceived safe environment in the under obligation and mutual high classes. Indeed, both direct effects show relevant substantive insights.

Furthermore, in Table 15 we report the class-specific means and variances obtained using the LCcw approach. The job insecurity is the highest in the over-obligation class (with also the highest variance), while lowest in the under-obligation class (with also the lowest variance, mostly due to within-class homogeneity). The main conclusions are comparable to the initial results by Bakk, Tekle, & Vermunt (2013), where the authors used a three-step distal outcome model. That is job insecurity is the largest using both approaches in the over-obligation class and lowest in the under-obligation class, but effect

Table 14. Estimated direct effects of job insecurity on each variable per LC.

	Coefficients					
Insecurity on	Mutual high	Over-ob	Under-Ob	Mutual low	Wald(0) *P*	Wald(=) *P*
Job security	−0.813	−0.148	−0.138	−0.595	0.000	0.130
Advancement	−0.322	0.347	0.677	0.172	0.040	0.090
Good pay	−0.203	−0.105	−0.042	−0.258	0.040	0.920
Safe environ	−0.492	0.191	−1.085	−0.056	0.630	0.020
Loyalty	−1.135	−0.469	−0.936	−0.057	0.050	0.250
Volunteer	−0.720	−0.503	−1.092	0.102	0.000	0.030
Timeliness	−1.080	−0.867	−0.862	−0.372	0.000	0.650
Perform	−0.636	−0.220	1.696	−0.259	0.010	0.600

Notes: ***P-value<0.01, **P-value<0.05, *P-value<0.1, P-values from Wald test of joint equality of each variable's direct effect to zero— Wald(0)—and from Wald test of equality of effects across LCs— Wald(=)—for the LCcw model. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

Table 15. The class-specific means, variances and corresponding Wald test statistics for the LCcw model on the PSYCON data.

	Mutual high	Over-ob	Under-Ob	Mutual low	Wald test statistic
Means	2.515	2.9089	1.5864	2.3691	304.66
	(0.0832)	(0.088)	(0.0442)	(0.0996)	
Variances	0.8145	1.0532	0.2407	0.8857	172.23
	(0.0684)	(0.0906)	(0.0273)	(0.1137)	

sizes differ. Nevertheless using the stepwise distal outcome approach the direct effects were ignored.

6. Conclusion

In this paper, we have brought modelling ideas from the regression mixture literature into LCA. Our focus has been to motivate the use of the cluster-weighted modelling approach as a general specification for the joint relationship of the response variables, the external variable and the latent variable. Six artificial data analyses, and a comprehensive simulation study have been used to illustrate this idea, and the actual advantage of the proposed approach was showed through an application using data from the Dutch and Belgian sample of the PSYCONES project (European Commission, 2006).

The cluster-weighted modelling approach, contrary to the simpler distal outcome model, was able to bring to light, meaningful insights about the relationship between job insecurity and some indicators of the psychological contract type LCs. By allowing for a more general LCcw type of specification, we were able to show the presence of substantively interesting direct effect of job insecurity on volunteering and on perceived safe work environment in some of the classes.

In the applied researcher perspective, the proposed approach has several advantages. By relaxing the conditional independence assumption, cluster-weighted modelling allows to model direct effects when these are of substantive interest, as well as when they are not. In other words, if direct effects are present, our approach, contrary to the distal outcome

model, is able to correctly estimate the distal outcome cluster-specific means and variances. Practically, by starting from the most general model and then testing the model assumptions of both the distal outcome and the LC regression models, it guarantees a more flexible option.

The approach we suggest has some limitations as well. First, it can be unstable if some of the response patterns are unobserved, or their sample frequency is too small (for standard sufficient conditions for identifiability of LCA see, for instance, Bandeen-Roche *et al.* 1997). In such cases, simpler models can be more attractive. Second, in exploratory contexts where the goal of the analysis is not always clear in mind, the *circularity* of cluster-weighted modelling might be relatively harder to interpret and simpler models might be preferable.

Appendix: Results of data analysis on Block 2 of section 3 with two external variables

Data were generated with entropy-based R^2 for the correctly specified model is about 0.7 in all the three data sets. To avoid the well-known issue of degeneracy of Gaussian mixtures (McLachlan & Peel 2004; see also Di Mari, Rocci, & Gattone 2017; García-Escudero *et al.* 2018 and references therein), we impose homoscedasticity in the class conditional variances for both external variables.

A.1. *LCreg* data

The *LCreg* data set was generated like in 3.1 for the LC model with two continuous external variables—each drawn from a standard normal distribution—loaded on all six indicators. The external variables Z_1 and Z_2 are loaded on the indicators with a coefficient of 0.5, if the most likely response is on the first class, or −0.5, if the most likely response is on the second class, giving a moderate/large effect size (Tables A1 and A2).

Table A1. *LCreg* data.

	Class proportions		Entr. R^2	#par
LCreg	**0.3004**	**0.6996**	**0.5967**	**27**
LCdist	0.4456	0.5544	0.6583	28
LCcw	0.2996	0.7004	0.5971	44

Notes: Two external variables. Estimated class proportions, entropy-based R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table A2. *LCreg* data.

	Z_1				Z_2			
	Means		Wald(=) P	Variance	Means		Wald(=) P	Variance
LCdist	−0.1809***	0.1512***	0.0000	0.9766	−0.1729***	0.1551***	0.0000	0.9734
	(0.0095)	(0.0094)		(0.0081)	(0.0095)	(0.0094)		(0.0081)
LCcw	0.0048	0.0025	0.9328	1.0038	0.0146	0.0132	0.7438	1.0000
	(0.0143)	(0.0081)		(0.0082)	(0.0144)	(0.0081)		(0.0082)

Notes: Two external variables. Estimated means (***P-value<0.001, **P-value<0.01, *P-value<0.05) and variances, and P-values from Wald test of equality of component means for the LCdist model and the LCcw model. Standard errors in parentheses. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

A.2. LC distal outcome (*LCdist*) data

The *LCdist* data set was generated from a two-class LCdist model, with class memberships of 0.7 and 0.3, six dichotomous indicators ($J = 6$) and two continuous external variables, both drawn from a normal distribution with unit variance and mean depending on class membership of either -1 or 1 (Tables A3 and A4).

Table A3. *LCdist* data.

	Class proportions		Entr. R^2	#par
LCreg	0.5010	0.4990	0.1599	27
LCdist	**0.6937**	**0.3063**	**0.6818**	**28**
LCcw	0.6934	0.3066	0.6787	44

Notes: Two external variables. Estimated class proportions, entropy R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table A4. *LCdist* data.

	Z_1				Z_2			
	Means		Wald(=) P	Variance	Means		Wald(=) P	Variance
LCdist	−1.0114***	0.9785***	0.0000	1.0161	0.4992***	−0.5113***	0.0000	1.0065
	(0.0087)	(0.0146)		(0.0112)	(0.0077)	(0.0127)		(0.0090)
LCcw	−1.0116***	0.9775***	0.0000	1.0166	0.4990***	−0.5099***	0.0000	1.0071
	(0.0097)	(0.0164)		(0.0118)	(0.0081)	(0.0139)		(0.0092)

Notes: Two external variables. Estimated means (***P-value<0.001, **P-value<0.01, *P-value<0.05) and variances, and P-values from Wald test of equality of component means for the LCdist model and the LCcw model. Standard errors in parentheses. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

A.3. *LCcw* data

The *LCcw* data set was generated from a two-class LCdist model, with class memberships of 0.7 and 0.3, six dichotomous indicators ($J = 6$) and two continuous external variables, both drawn from a normal distribution with unit variance and mean depending on class membership of either -1 or 1. Both Z_1 and Z_2 are loaded on the indicators with a coefficient of 0.5, if the most likely response is on the first class, or -0.5, if the most likely response is on the second class, giving a moderate/large effect size (Tables A5–A8).

Table A5. *LCcw* data.

	Class proportions		Entr. R^2	#par
LCreg	0.5500	0.4500	0.3224	27
LCdist	0.5004	0.4996	0.5877	28
LCcw	**0.3018**	**0.6982**	**0.7353**	**44**

Notes: Two external variables. Estimated class proportions, entropy R^2 and number of parameters for each of the three estimated models. Results from correctly specified model in bold font.

Table A6. *LCcw* data.

	Z_1				Z_2			
	Means		Wald(=) P	Variance	Means		Wald(=) P	Variance
LCdist	0.0473**	−0.8517***	0.0000	1.6225	−0.4047***	0.7991***	0.0000	0.8454
	(0.0172)	(0.0117)		(0.0153)	(0.0104)	(0.0133)		(0.0091)
LCcw	0.9993***	−1.0075***	0.0000	0.9760	−0.5079***	0.5012***	0.0000	0.9931
	(0.0142)	(0.0075)		(0.0094)	(0.0113)	(0.0077)		(0.0087)

Notes: Two external variables. Estimated means (***P-value<0.001, **P-value<0.01, *P-value<0.05) and variances, and P-values from Wald test of equality of component means for the LCdist model and the LCcw model. Standard errors in parentheses. Reported P-values are function of the unobserved LC variable, and are therefore approximate.

Table A7. *Adjusted Rand indexes*, computed between clustering with correctly specified models—LCreg, LCdist and LCcw models—and clustering with the other two models. Two external variables.

		Fitted model		
Data	Correct model	LCreg	LCdist	LCcwm
LCreg	LCreg	1	0.3592	0.9976
LCdist	LCdist	0.0597	1	0.9870
LCcw	LCcw	0.3312	0.3051	1

Table A8. *Model selection with BIC* computed for each model at each data generating model—LCreg, LCdist and LCcw—for $S = 1, \ldots, 5$. Data generating model and minimum BIC value, for each model at each scenario, in bold. Two external variables.

		Number of components				
Data		$S = 1$	$S = 2$	$S = 3$	$S = 4$	$S = 5$
LCreg	LCreg	241,594.70	**216,434.30**	216,598.30	216,731.50	216,870.90
	LCdist	413,838.40	394,636.90	389,691.80	389,371.80	**388,883.70**
	LCcw	412,023.70	**386,883.80**	387,028.10	387,242.80	387,361.30
LCdist	LCreg	239,603.70	239,120.30	239,080.80	239,100.10	**239,073.10**
	LCdist	441,991.90	**428,871.60**	428,956.10	429,033.30	429,128.00
	LCcw	434,547.40	**429,095.60**	429,278.90	429,456.00	429,599.80
LCcw	LCreg	242,531.80	234,943.10	234,472.40	234,281.10	**234,076.40**
	LCdist	443,603.30	429,953.50	424,364.30	421,796.10	**421,160.30**
	LCcw	436,546.50	**417,642.30**	417,820.20	418,007.40	418,201.30

Supporting information

Additional supporting information may be found in the online version of this article at http://wileyonlinelibrary.com/journal/anzs.

Data S1. Supplementary Material.

REFERENCES

ASPAROUHOV, T. & MUTHÉN, B. (2014). Auxiliary variables in mixture modeling: three-step approaches using Mplus. *Structural Equation Modeling*, **21**, 329–341.

BAKK, Z., TEKLE, F.B. & VERMUNT, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, **43**(1), 272–311.

BANDEEN-ROCHE, K., MIGLIORETTI, D., ZEGGER, S.L. & RATHOUZ, P. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, **92**, 1375–1386.

BAUER, D.J. & CURRAN, P.J. (2003). Distributional assumptions of growth mixture models: Implication for overextraction of latent trajectory classes. *Psychological methods*, **8**, 338–363.

CUYPER, N.D., RIGOTTI, T., WITTE, H.D. & MOHR, G. (2008). Balancing psychological contracts: validation of a typology. *The International Journal of Human Resource Management*, **19**, 543–561.

DAYTON, C.M. & MACREADY, G.B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, **83**, 173–178.

DI MARI, R. & BAKK, Z. (2018). Mostly harmless direct effects: A comparison of different latent Markov modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, **25**(3), 467–483.

DI MARI, R., BAKK, Z. & PUNZO, A. (2020). A random-covariate approach for distal outcome prediction with latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, **27**, 351–368.

DI MARI, R., DOTTO, F., FARCOMENI, A. & PUNZO, A. (2022). Assessing measurement invariance for longitudinal data through latent markov models. *Structural Equation Modeling: A Multidisciplinary Journal*, **29**, 381–393.

DI MARI, R., ROCCI, R. & GATTONE, S.A. (2017). Clusterwise linear regression modeling with soft scale constraints. *International Journal of Approximate Reasoning*, **91**, 160–178.

DURANTE, D., CANALE, A. & RIGON, T. (2019). A nested expectation–maximization algorithm for latent class models with covariates. *Statistics & Probability Letters*, **146**, 97–103.

GARCÍA-ESCUDERO, L.A., GORDALIZA, A., GRESELIN, F., INGRASSIA, S. & MAYO-ISCAR, A. (2018). Eigenvalues and constraints in mixture modeling: geometric and computational issues. *Advances in Data Analysis and Classification*, **12**, 203–233.

HUBERT, L. & ARABIE, P. (1985). Comparing partitions. *Journal of Classification*, **2**, 193–218.

INGRASSIA, S., MINOTTI, S. & VITTADINI, G. (2012). Local statistical modeling via a cluster–weighted approach with elliptical distributions. *Journal of Classification*, **29**, 363–401.

INGRASSIA, S., PUNZO, A., VITTADINI, G. & MINOTTI, S.C. (2015). The generalized linear mixed cluster-weighted model. *Journal of Classification*, **32**, 85–113.

ISAKSSON, K., BERNHARD, C., PEIRÓ, J.M., *et al.* (2003). *Psychological contracts across employment situations (PSYCONES). Results from the pilot study*. Stockholm: National Institute for Working Life & Saltsa.

KAMAKURA, W.A. & RUSSEL, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, **26**, 379–390.

KANKARAŠ, M., MOORS, G. & VERMUNT, J.K. (2010). *Testing for measurement invariance with latent class analysis*. NY: Routledge New York, pp. 359–384.

KIM, M., VERMUNT, J.K., BAKK, Z., JAKI, T. & VAN HORN, M.L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, **23**, 601–614.

LANZA, T.S., TAN, X. & BRAY, C.B. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, **20**, 1–26.

LEE, S., BULUT, O. & SUH, Y. (2017). Multidimensional extension of multiple indicators multiple causes models to detect dif. *Educational and Psychological Measurement*, **77**, 545–569.

MAGIDSON, J. (1981). Qualitative variance, entropy, and correlation ratios for nominal dependent variables. *Social Science Research*, **10**(2), 177–194.

MASYN, K.E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **24**, 180–197.

MAZZA, A., PUNZO, A. & INGRASSIA, S. (2018). flexCWM: a flexible framework for cluster-weighted models. *Journal of Statistical Software*, **86**, 1–30.

MCCUTCHEON, A.L. (1985). A latent class analysis of tolerance for nonconformity in the American public. *Public Opinion Quarterly*, **49**, 474–488.

MCLACHLAN, G.J. & PEEL, D. (2004). *Finite Mixture Models*. New York: John Wiley & Sons.

MULDER, E., VERMUNT, J., BRAND, E., BULLENS, R. & VAN MERLE, H. (2012). Recidivism in subgroups of serious juvenile offenders: different profiles, different risks? *Criminal Behaviour and Mental Health*, **22**, 122–135.

NYLUND-GIBSON, K. & MASYN, K.E. (2016). Covariates and mixture modeling: results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, **23**, 782–797.

OKAZAKI, S., CAMPO, S., ANDREU, L. & ROMERO, J. (2015). A latent class analysis of Spanish travelers' mobile internet usage in travel planning and execution. *Cornell Hospitality Quarterly*, **56**, 191–201.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

ROBERTS, T.J. & WARD, S.E. (2011). Using latent transition analysis in nursing research to explore change over time. *Nursing Research*, **60**, 73–79.

SUH, Y. & CHO, S.J. (2014). Chi-square difference tests for detecting differential functioning in a multidimensional IRT model: a Monte Carlo study. *Applied Psychological Measurement*, **38**, 359–375.

VERMUNT, J.K. (1997). *Log-Linear models for event histories*. New Delhi: SAGE Publications, India.

VERMUNT, J.K. (2010). Latent class modeling with covariates: two improved three-step approaches. *Political Analysis*, **18**, 450–469.

VERMUNT, J.K. & MAGIDSON, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced and Syntax*. Belmont Massachusetts: Statistical Innovations Inc.

VERMUNT, J.K. & MAGIDSON, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

WEDEL, M. & DESARBO, W. (1994). *A Review of Recent Developments in Latent Class Regression Models*. Cambridge, Massachusetts: Blackwell Publishers, pp. 352–388.

DE WITTE, H., BOUWEN, R., DE WITTE, K., DE WITTE, H. & TAILLEU, T. (2000). Arbeidsethos en jobonzekerheid: meting en gevolgen voor welzijn, tevredenheid en inzet op het werk. In *Van groep naar gemeenschap. Liber Amicorum Prof. Dr. Leo Lagrou*, eds., Bouwen, R., De Witte, K., De Witte, H., & Tailleu, T.G., Leuven: Garant, pp. 325–350.

WOODS, C.M. (2009). Evaluation of mimic-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, **44**, 1–27.