Structural Equation Modeling, 18:525–536, 2011

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2011.607071

Psychology Press
Taylor & Francis Group

A Note on Cluster Effects in Latent Class Analysis

David Kaplan and Bryan Keller
Department of Educational Psychology
University of Wisconsin–Madison

This article examines the effects of clustering in latent class analysis. A comprehensive simulation study is conducted, which begins by specifying a true multilevel latent class model with varying within- and between-cluster sample sizes, varying latent class proportions, and varying intraclass correlations. These models are then estimated under the assumption of a single-level latent class model. The outcomes of interest are measures of bias in the Bayesian Information Criterion (BIC) and the entropy R^2 statistic relative to accounting for the multilevel structure of the data. The results indicate that the size of the intraclass correlation as well as between- and within-cluster sizes are the most prominent factors in determining the amount of bias in these outcome measures, with increasing intraclass correlations combined with small between-cluster sizes resulting in increased bias. Bias is particularly noticeable in the BIC. In addition, there is evidence that class separation interacts with the size of the intraclass correlations and cluster sizes in producing bias in these measures.

Keywords: BIC, cluster effects, entropy, latent class analysis, model adequacy, multilevel latent class analysis

Statistical modeling of theoretical relationships in the social and behavioral sciences often requires the use of latent variables. The most common type of latent variable used in statistical modeling is the continuous common factor model obtained via the methods of factor analysis. Although the use of continuous latent variables arguably dominates most applications of latent variable modeling, it is often useful to hypothesize the existence of categorical latent variables. Such categorical latent variables are presumed to explain response frequencies among dichotomous or ordered categorical variables. The method used to estimate categorical latent variables is *latent class analysis* (LCA).

Conventional LCA was introduced by Lazarsfeld and Henry (1968) for the purposes of deriving latent attitude variables from responses to dichotomous survey items. Important contributions to LCA have been made by Clogg (1995). For a review see Magidson and Vermunt (2004). In a traditional LCA it is assumed that an individual belongs to one and only one

Correspondence should be addressed to David Kaplan, Department of Educational Psychology, University of Wisconsin–Madison, 1025 W. Johnson Street, Madison, WI 53706, USA. E-mail: dkaplan@education.wisc.edu

latent class, and that given an individual's latent class membership, the observed responses are independent of one another—referred to as the assumption of *local independence*. The latent classes arise from the patterns of response frequencies to categorical items, where the response frequencies play a role similar to that of the correlation matrix in factor analysis (Lanza, Collins, Lemmon, & Schafer, 2007). The analog of factor loadings are parameters that estimate the probability of a particular response on the manifest indicators given membership in the latent class. Unlike continuous latent variables (i.e., factors), categorical latent variables (latent classes) divide individuals into mutually exclusive and exhaustive groups.

Until recently, applications of LCA were limited to single-level problems, ignoring possible clustering of observations due to sample design and research considerations. However, utilizing the finite mixture modeling framework of LCA (McLachlan & Peel, 2000) implemented in software packages such as M*plus* (L. K. Muthén & Muthén, 1998–2007) and Latent GOLD (Vermunt & Magidson, 2000), more general frameworks been developed that allow latent class models to be extended to multilevel contexts.

A review of the extant literature on multilevel LCA has revealed important and interesting applications (see, e.g., Kaplan, Kim, & Kim, 2009; Van Horn et al., 2008). In a recent paper, Van Horn et al. (2008) examined multilevel mixture models for the evaluation of interventions in group randomized trials. The motivation for using mixture models to evaluate interventions is that an intervention might not have a "one-size-fits-all" effect. That is, an intervention might be more or less beneficial for subpopulations of individuals, and the goal would be to identify those subpopulations for which the intervention is most beneficial.

The Van Horn et al. (2008) article pays special attention to evaluating intervention effects in group randomized trials. Group randomized trials figure prominently in educational interventions in which classrooms (and sometimes whole schools) are randomly assigned to receive an intervention. Proper statistical treatment of group randomized trials requires random-effects models that assume that individuals are independent of each other conditional on group membership.

The application presented by Van Horn et al. (2008) examined the efficacy of a group randomized trial of an intervention aimed at mitigating problem behavior among youth. The specific goal was to examine whether the intervention had differential affects among those students engaged in experimenting with illicit substances versus students engaged in more problematic behaviors such as violence. The intervention took place at the community level, but students were nested within schools. The two-level model that Van Horn et al. tested consisted of a within-school latent class model, with a single latent class variable consisting of four and five latent classes. In a fixed-effects model, it would normally be assumed that the proportion of students within each latent class is constant across the schools. Van Horn and colleagues assumed that these proportions vary across schools and are affected by the community-level intervention. On the whole, Van Horn et al. found strong evidence that the probability that a student belonged to a specific latent class of problem behavior varied across schools. In terms of treatment and control equivalency, no difference was found in the proportion of students in the latent classes prior to the intervention focusing on youth behaviors. Additionally, Van Horn et al. showed how power analyses could be used to examine intervention effects in the multilevel latent class model.

In a recent didactic paper, Kaplan et al. (2009) applied multilevel LCA (Vermunt, 2003) to reading data from the Early Childhood Longitudinal Study (ECLS; National Center for Education Statistics, 2001). Specifically, they used the fall first-grade data of the ECLS–K, focusing on five reading subtests. Given the pattern of response probabilities, labels for the latent classes were generated. The Alphabet Knowledge (AK) class consisted of individuals with moderate or high probabilities of passing the letter recognition subtest and low probabilities of passing the rest of the subtests. The Phonemic Awareness (PA) class consisted of individuals with a very high probability of passing the letter recognition subtest, moderately high probabilities of passing the beginning and ending sounds subtests, and very low probabilities for the rest of the subtests. The Word Knowledge (WK) class consisted of very high probabilities of passing all of the subtests except the words in context subtest. In comparison to the conventional latent class model, which ignored clustering, the results showed virtually no difference in response probabilities or latent class proportions when taking into account clustering. However, a dramatic decrease in model selection measures (the Akaike's Information Criterion [AIC] and Bayesian Information Criterion [BIC]) was observed, suggesting that accounting for the multilevel structure of the data improved predictive accuracy.

A common practice when using multilevel modeling methods is to provide an assessment of the amount of variation in the outcomes of interest that lies among clusters. This assessment provides the analyst with a sense of the impact that clustering might have on the results. In considering the application of multilevel LCA to substantive problems in the social and behavioral sciences, it is also crucial to determine the impact of clustering. If the amount of variance that lies at the cluster level is substantively small, then perhaps not much is lost by modeling the disaggregated data under the assumption of independent observations. However, if the amount of between-cluster variance is large, then it might not be reasonable to consider the data as nonindependent, and explicit multilevel modeling must take place.

Although there have been many studies examining the impact of clustering in the multilevel regression modeling context (see, e.g., Goldstein, 2003; Mass & Hox, 2004), a review of the extant literature suggests that the effects of clustering have not been studied in the LCA setting. This article, therefore, adds to the methodological literature by examining the impact of clustering within the latent class modeling framework, with a unique focus on model selection and classification adequacy.

The organization of this article is as follows. Next we provide the specification of the multilevel latent class model. This is then followed by a discussion of the design of the simulation study. The results then follow. Finally, the article concludes with a summary and implications of our findings for the design and analysis of studies intended to utilize LCA.

MULTILEVEL LATENT CLASS ANALYSIS

We assume that the reader is familiar with the specification of the conventional latent class model (e.g., Clogg, 1995). Our specification of the multilevel latent class model is in line with Vermunt (2003). To contextualize this problem, consider students as the individuals and schools as the clusters. To begin, let $\mathbf{y}_{ig}$ be the vector of responses for student i in school g, where $i = 1, 2, \ldots, n_g$; $g = 1, 2, \ldots, G$, and let $\mathbf{s}$ be a possible response pattern vector. Furthermore, let the response vector $\mathbf{y}_{ig}$ represent dichotomous proficiency scores on, say, a mathematics competency assessment. Let K be the number of indicators, where $k = 1, 2, \ldots, K$. A specific outcome level for indicator k is denoted as s_k and the total number of categories is S_k. For example, in the case of a dichotomous item k then $S_k = 2$. Further, let C_{ig} be a latent class variable with specific latent class c, $c = 1, 2, \ldots, M$, where M is the total number of latent classes. Following Vermunt (2003), the multilevel latent class model can be written as follows. Let

$$P(\mathbf{y}_{ig} = s) = \sum_{c=1}^{M} P(C_{ig} = c) P(\mathbf{y}_{ig} = \mathbf{s} | C_{ig} = c) \quad (1)$$

$$= \sum_{c=1}^{M} P(C_{ig} = c) \prod_{k=1}^{K} P(y_{igk} = s_k | C_{ig} = c). \quad (2)$$

We see that the probability of a particular response pattern on the mathematics competency assessment for student i in school g is conditional on membership in a specific latent class. These conditional probabilities in the second part of Equation 1 are used to name the categorical factor in a manner similar to the use of factor loading patterns in factor analysis. The weights given by $P(C_{ig} = c)$ ensure that the probabilities sum to one.

As in conventional LCA, two model probabilities must be obtained. The first is the probability that the categorical latent variable C_{ig} takes on a particular value c.

$$P(C_{ig} = c) = \frac{\exp(\gamma_{cg})}{\sum_{r=1}^{M} \exp(\gamma_{rg})}, \quad (3)$$

The second is the probability that the response pattern on the indicators y_{igk} is observed at s_k written as

$$P(y_{igk} = s_k | C_{ig} = c) = \frac{\exp(\beta^k_{s_k cg})}{\sum_{u=1}^{S_k} \exp(\beta^k_{ucg})}, \quad (4)$$

where γ_{cg} and $\beta^k_{s_k cg}$ are logit parameters, with the restrictions that, say, $\gamma_{1g} = \beta^k_{1cg} = 0$.

Vermunt (2003) proposed two approaches to multilevel LCA based on including either parametric or nonparametric random effects in the multilevel latent class model. The two approaches differ conceptually in how the between-cluster heterogeneity is explained. The parametric approach assumes that cluster effects originate from a certain probability distribution, whereas the nonparametric approach assumes the existence of a discrete number of mixture components. We focus on the parametric approach because it allows for the calculation of intraclass correlations. Under the parametric approach the values of γ_{cg}, other than the one constrained for identification, are assumed to follow normal distributions such that $\gamma_{cg} = \gamma_c + \tau_c \cdot u_g$, where u_g is standard normal. In particular, for the three-class model, which is the focus of the simulation study in the next section, the random effects are

$$\gamma_{1g} = 0$$

$$\gamma_{2g} = \gamma_2 + \tau_2 \cdot u_g$$

$$\gamma_{3g} = \gamma_3 + \tau_3 \cdot u_g.$$

The Intraclass Correlation

In conventional multilevel modeling for continuous variables, a logical first step is to assess amount of variance in Level 1 variables accounted for by cluster effects. The measure typically used for this assessment is the intraclass correlation coefficient (ICC). An analog of the intraclass correlation extended to the multinomial logistic model for random effects was given by Hedeker (2003) and can be written as

$$\rho_{Ic} = \frac{\sigma_c^2}{\sigma_c^2 + \pi^2/3} \tag{5}$$

where σ_c^2 is the Level 2 variance. Equation 5 makes use of the fact that the Level 1 variance in the random effects logistic regression framework is $\pi^2/3 \approx 3.29$. For the three-class model it is possible to specify two independent intraclass correlations:

$$ICC_1 = \frac{\tau_2^2}{\tau_2^2 + \pi^2/3} \quad \text{and} \quad ICC_2 = \frac{\tau_3^2}{\tau_3^2 + \pi^2/3}.$$

The focus of the simulation study is the manipulation of the size of the ICCs to alter the amount of total variance that lies between groups.

Measures of Model Adequacy

In LCA, proceeding to name the latent classes presumes that the model adequately describes the data. Although there are many methods for assessing the adequacy of a latent class model, for simplicity, we focus on one measure of model selection and one measure of classification adequacy. In terms of model selection, we focus on the BIC, also referred to as Schwarz criterion (Schwarz, 1978). The BIC is a measure used for selecting among a set of competing models and has its origins in model selection based on the notion of Bayes factors (Kass & Raftery, 1995). It is, arguably, the most widely used method for model selection in the LCA context (Magidson & Vermunt, 2004). The BIC can be written in a general form as

$$BIC = -2 \ln L + q[\ln(n)], \tag{6}$$

where $\ln L$ is the log-likelihood, q is the number of parameters in the model, and n represents the sample size. In terms of model comparison, the model with the lower BIC among a set of competing models is preferred from a posterior predictive point of view.

In terms of classification quality, we focus on the $R^2_{entropy}$, which starts with a general form related to the reduction of classification errors. Specifically, following Vermunt and Magidson (2000), the proportional reduction of classification errors can be written as

$$R_c^2 = \frac{\text{Error}(C = c) - \text{Error}(C = c|\mathbf{y} = y)}{\text{Error}(C = c)} \tag{7}$$

where

$$\text{Error}(C = c|\mathbf{y} = y) = \frac{\sum_{i=1}^{I} w_i \text{Error}(C_{ig} = c_{ig}|\mathbf{y}_{ig} = y_{ig})}{n} \tag{8}$$

and where w_i are case weights that can be used when analyzing data arising from a complex sampling design, such as the design implemented for the ECLS–K. From here, a number of R^2 measures can be defined (Vermunt & Magidson, 2000). One measure of the proportional reduction in classification errors is based on the concept of entropy. Entropy, in the context of LCA, was developed by Ramaswamy, Desarbo, Reibstein, and Robinson (1993) as an overall measure of the degree of "fuzziness" in class membership. Values close to zero can occur when the posterior probabilities of class membership are equal, suggesting that the latent classes are not distinct. Higher values of entropy suggest clearer distinctions among the latent classes. An $R^2_{entropy}$ measure is obtained by substituting $\text{Error}(C_{ig} = c_{ig}|\mathbf{y}_{ig} = y_{ig})$ in Equation 8 with

$$\begin{aligned}\text{Error}(C_{ig} = c_{ig}|\mathbf{y}_{ig} = y_{ig}) = \sum_{x=1}^{K} -\hat{P}(C_{ig} = c_{ig}|\mathbf{y}_{ig} = y_{ig}) \\ \times \log \hat{P}(C_{ig} = c_{ig}|\mathbf{y}_{ig} = y_{ig}).\end{aligned} \tag{9}$$

SIMULATION STUDY METHODS

A simulation study was conducted to determine the effects of ignoring clustering on commonly used measures of model adequacy in a three-class LCA. The experimental conditions included varying the size of the intraclass correlations, the within- and between-cluster sample sizes, and latent class sizes.

Intraclass Correlations

Under the specification of the three-class model it is only possible to specify two independent values of the intraclass correlation. Both of these intraclass correlations were simulated at five levels: 0.0, 0.1, 0.2, 0.3, and 0.4.

Within- and Between-Cluster Size

The following within- and between-cluster sample sizes were constructed to yield a total sample size of 1,200: 15 W/80 B, 20 W/60 B, 30 W/40 B, and 80 W/15 B, 60 W/20 B, 40 W/30 B. Note that we do not alter the total sample size. Although manipulating the total sample size might be of some interest, researchers often need to make decisions about how to gather data in the most efficient way when, for example, more groups might be examined at the cost of less cases per group or vice versa. To study the effect of this trade-off we chose to focus on examining the effects of altering the between- versus within-cluster sizes holding a reasonably large overall sample size.

Latent Class Proportions

To investigate the role of class proportions, three conditions were simulated: one in which class proportions were homogenous (.33, .33, .33) and the other two in which class proportions were heterogeneous (.7, .2, .1) and (.2, .7, .1).

Latent Class Structure

Six dichotomous items (0 = *negative response*, 1 = *positive response*) were simulated for each experimental condition. The class-specific positive response probabilities for the six items were set to (0.8, 0.8, 0.8, 0.8, 0.8, 0.8), (0.8, 0.8, 0.8, 0.2, 0.2, 0.2), and (0.2, 0.2, 0.2, 0.2, 0.2, 0.2) for latent Classes 1, 2, and 3, respectively, to create classes with a moderate degree of separation. The probability of a positive response on any of the six items, given membership in latent Class 1 is 0.8, given membership in latent Class 2 is 0.8 for the first three items and 0.2 for the last three items, and given membership in latent Class 3 is 0.2.

Our choices of experimental conditions follow closely that of Lukočiene and Vermunt (2010) and provide a reasonable degree of class separation, as measured by the $R^2_{entropy}$. The average $R^2_{entropy}$ value of the population-generating models across all experimental conditions is about 0.68 ($SD = 0.04$). As expected from Lukočiene & Vermunt, the amount of separation is neither too high nor too low and is arguably representative of the class separation found in most empirical studies. Tables of population values are available on request.

Each of the four factors (the two intraclass correlations, the set of within- and between-cluster sample size conditions, and the latent class proportions) were fully crossed to yield a total of 450 unique experimental conditions. The *MCstudy* routine in Latent GOLD (Vermunt & Magidson, 2000) was used to conduct the simulation study and all analyses utilized maximum likelihood under a combination of the Expectation Maximization (EM) algorithm and the Newton–Raphson algorithm.

The steps of the simulation study were as follows: First, for each cell of the design, 100 replications of $N = 1{,}200$ cases were generated according to the specific correct population model.[1] Second, the data were analyzed with the correct model and summary statistics were recorded—in particular the average BIC and average $R^2_{entropy}$ over the 100 replications for each cell. These average values for the true model constitute the comparison base for the misspecified models. Third, the data were analyzed ignoring clustering and the average BIC and average $R^2_{entropy}$ for the misspecified model were calculated. Relative to the population-generating model, the misspecified model was created by removing the continuous group-level factor from the population-generating model. This resulted in the elimination of the two parameters that determine intraclass correlations. Again, the outcomes of interest in this study are the percentage under or overestimation in the BIC and $R^2_{entropy}$ from the model that ignores

[1]Preliminary analyses using larger numbers of replications did not reveal substantial differences when compared to 100 replications.

TABLE 1
Difference in Bayesian Information Criterion: Class Proportions (.33, .33, .33)

		ICC_2					ICC_2					ICC_2				
		0.0	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	−13	35	103	163	238	−13	43	114	188	254	−13	59	136	215	292
	0.1	51	177	232	270	325	68	185	255	295	348	76	222	278	339	388
ICC_1	0.2	142	242	298	328	356	160	275	313	351	389	177	297	355	396	438
	0.3	229	330	334	357	385	252	353	365	389	406	275	375	396	425	449
	0.4	315	378	397	399	405	344	406	429	433	448	369	431	458	482	449
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	−13	58	142	227	298	−13	79	152	247	338	−13	76	160	248	345
	0.1	90	224	284	350	406	96	234	312	354	431	104	239	320	354	420
ICC_1	0.2	186	313	362	392	446	207	325	362	406	443	216	322	383	401	459
	0.3	301	389	431	432	473	306	425	435	449	492	320	418	451	480	470
	0.4	383	461	465	477	496	400	469	491	500	506	423	484	491	500	504

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

clustering compared to the population-generating model. For the BIC, we provide the absolute difference in the population and misspecified quantities because common practice uses the differences in BIC across models as a means of model choice.[2]

Results of Simulation Study

Tables 1, 2, and 3 show that the absolute difference in BIC between the true model and misspecified model is mostly related to the size of the ICC. It should be noted that although the differences in BIC are quite large in practical terms, when examined as a percentage of the population value, the bias is not very large, never exceeding 10%. The tables of percentage bias in BIC are available on request.

In addition to sensitivity of the BIC in terms of the size of the ICCs, we also observe differences due to cluster size. In particular, absolute differences in the BIC are noticeably larger in cases of small between-cluster size regardless of the size of the ICCs. For example, in Table 1, for the 15 W/80 B and .4/.4 ICC conditions, we find that the absolute difference in the BIC is 405, compared to 504 in the 80 W/15 B condition. This pattern holds across all ICC, cluster size, and class proportion conditions of the study.

Tables 4, 5, and 6 show that the $R^2_{entropy}$ is consistently underestimated when ignoring the multilevel structure of the data and the underestimation worsens with increasing size of the ICC. Because higher values of the $R^2_{entropy}$ imply better prediction of class membership, this result suggests that ignoring multilevel structure in the latent class model worsens the prediction of latent class membership. However, the degradation in prediction of latent class

[2]Percentage bias is defined as the percentage under- or overestimation of the measure for the misspecifed model relative to the population model and is calculated as 100 × (misspecified model − population model)/(population model).

TABLE 2
Difference in Bayesian Information Criterion: Class Proportions (.70, .20, .10)

		ICC_2					ICC_2					ICC_2				
		0.0	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	−13	0	24	65	106	−13	2	35	68	126	−13	7	45	85	129
	0.1	28	70	99	118	155	38	66	107	137	160	42	81	120	145	178
ICC_1	0.2	89	115	135	156	183	101	133	137	167	191	120	138	166	186	194
	0.3	150	161	176	201	206	169	179	197	212	213	189	212	203	235	234
	0.4	200	206	212	221	224	222	213	236	235	239	251	230	239	248	235
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	−13	11	48	98	145	−14	13	59	101	171	−13	18	63	115	160
	0.1	52	96	128	167	201	57	105	132	156	208	65	95	130	183	207
ICC_1	0.2	130	153	174	195	215	134	159	197	202	221	141	154	172	195	233
	0.3	189	208	208	230	232	221	228	239	249	256	206	186	226	237	225
	0.4	265	226	255	266	270	265	284	269	259	267	265	258	271	258	254

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

membership is partly a function of the size of the ICCs relative to the latent class membership proportion. Unlike the findings for the BIC, bias in $R^2_{entropy}$ does not appear to be a function of the cluster size conditions.

In each cluster size condition and across each latent class size distribution, the lowest and highest biases were observed in the $ICC_1 = ICC_2 = 0.0$ and $ICC_1 = ICC_2 = 0.4$ cases, respectively. Interestingly, the biases are not symmetric with respect to ICCs. That is, for a fixed class size and fixed cluster sizes, the condition in which $(ICC_1, ICC_2) = (0.0, 0.4)$

TABLE 3
Difference in Bayesian Information Criterion: Class Proportions (.20, .70, .10)

		ICC_2					ICC_2					ICC_2				
		0.0	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	−13	11	54	91	141	−13	18	68	110	154	−13	26	72	124	167
	0.1	4	55	83	127	156	7	58	94	131	165	11	69	108	143	177
ICC_1	0.2	39	77	116	129	158	48	83	126	135	190	58	96	134	151	193
	0.3	77	99	110	148	163	91	110	124	155	179	111	127	154	173	193
	0.4	118	131	162	161	184	139	145	166	181	210	151	164	192	184	220
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	−13	29	82	141	197	−13	31	80	149	201	−13	42	101	153	211
	0.1	20	68	124	155	193	19	80	132	177	210	23	84	124	160	198
ICC_1	0.2	63	101	143	165	188	72	121	137	160	219	76	125	144	163	230
	0.3	119	146	146	166	204	112	154	161	188	202	126	156	166	194	215
	0.4	168	156	208	212	223	171	183	205	220	225	196	169	220	193	255

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

TABLE 4
Percentage Bias in Entropy-R^2; Class Proportions (.33, .33, .33)

		ICC_2					ICC_2					ICC_2				
		0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	0	−3	−6	−9	−12	0	−3	−6	−10	−13	0	−4	−7	−10	−13
	0.1	−2	−6	−9	−10	−13	−2	−6	−9	−11	−14	−2	−7	−10	−12	−14
ICC_1	0.2	−4	−8	−10	−12	−13	−5	−9	−10	−12	−14	−5	−9	−11	−13	−15
	0.3	−7	−10	−11	−12	−14	−7	−11	−11	−13	−14	−8	−10	−12	−13	−15
	0.4	−9	−11	−13	−13	−14	−9	−11	−13	−14	−14	−10	−12	−13	−14	−14
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	0	−4	−7	−11	−14	0	−4	−8	−12	−15	0	−5	−8	−11	−15
	0.1	−3	−7	−10	−12	−15	−3	−7	−10	−12	−15	−3	−8	−10	−12	−15
ICC_1	0.2	−5	−9	−11	−13	−14	−6	−9	−11	−13	−14	−6	−9	−11	−12	−14
	0.3	−8	−11	−12	−13	−15	−8	−11	−12	−13	−15	−9	−11	−12	−14	−15
	0.4	−10	−12	−13	−14	−15	−10	−12	−13	−14	−15	−10	−13	−13	−14	−15

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

does not necessarily display the same percentage bias in BIC or $R^2_{entropy}$ as the condition in which $(ICC_1, ICC_2) = (0.4, 0.0)$. These conditions differ in bias by as much as 5% in $R^2_{entropy}$ and more than 1.5% (or an absolute difference of 122) in BIC. Moreover, these asymmetries manifest themselves systematically in different magnitudes across the three latent class size conditions, suggesting an interaction between latent class proportions and ICCs. Further research is necessary to determine the effect of latent class proportions and response probability distribution on measures of model and predictive adequacy.

TABLE 5
Percentage Bias in Entropy-R^2; Class Proportions (.70, .20, .10)

		ICC_2					ICC_2					ICC_2				
		0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	0	−1	−4	−7	−10	0	−2	−5	−7	−11	0	−2	−6	−8	−11
	0.1	−2	−4	−6	−8	−9	−2	−4	−6	−8	−10	−3	−4	−6	−8	−11
ICC_1	0.2	−4	−5	−7	−8	−10	−5	−6	−7	−8	−10	−5	−6	−7	−9	−11
	0.3	−6	−7	−8	−9	−10	−7	−8	−9	−9	−10	−7	−8	−8	−9	−11
	0.4	−9	−9	−9	−10	−10	−10	−9	−9	−10	−10	−10	−10	−10	−10	−11
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	0	−2	−5	−9	−12	0	−2	−5	−9	−12	0	−3	−6	−9	−13
	0.1	−3	−4	−6	−9	−11	−3	−5	−6	−9	−12	−3	−5	−7	−8	−12
ICC_1	0.2	−4	−6	−8	−9	−10	−6	−6	−8	−9	−11	−6	−6	−7	−9	−11
	0.3	−8	−8	−8	−10	−11	−8	−8	−9	−9	−10	−9	−10	−8	−9	−10
	0.4	−10	−9	−10	−11	−11	−10	−9	−10	−11	−11	−11	−10	−10	−10	−10

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

TABLE 6
Percentage Bias in Entropy-R^2; Class Proportions (.20, .70, .10)

		ICC_2					ICC_2					ICC_2				
		0.0	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.0*	*0.1*	*0.2*	*0.3*	*0.4*
		15 W/80 B					20 W/60 B					30 W/40 B				
	0.0	0	−3	−8	−11	−14	0	−4	−8	−11	−15	0	−5	−9	−12	−16
	0.1	−1	−6	−9	−11	−14	−2	−6	−9	−12	−15	−2	−6	−10	−13	−17
ICC_1	0.2	−5	−7	−10	−11	−14	−5	−8	−10	−12	−15	−5	−8	−11	−13	−15
	0.3	−7	−9	−11	−13	−15	−7	−9	−11	−13	−15	−8	−10	−12	−14	−16
	0.4	−10	−11	−13	−14	−15	−11	−12	−14	−14	−16	−11	−13	−13	−14	−15
		40 W/30 B					60 W/20 B					80 W/15 B				
	0.0	0	−5	−8	−13	−17	0	−4	−9	−13	−16	0	−5	−10	−15	−17
	0.1	−3	−7	−9	−13	−17	−3	−8	−9	−13	−17	−3	−8	−11	−13	−16
ICC_1	0.2	−6	−9	−11	−13	−16	−6	−9	−11	−13	−16	−6	−9	−11	−14	−15
	0.3	−9	−12	−11	−15	−15	−9	−12	−12	−14	−15	−10	−12	−12	−13	−15
	0.4	−12	−13	−14	−15	−16	−11	−14	−14	−14	−16	−13	−14	−15	−15	−16

Note. ICC = intraclass correlation coefficient; W = within-cluster; B = between-cluster.

CONCLUSIONS

The results of a simulation study on the problem of cluster effects in the latent class model were presented. The findings demonstrate the importance of accounting for multilevel structure in the latent class model and are generally consistent with findings in the multilevel regression modeling literature. Of the design conditions examined in this study, the most important factors contributing to problems in model selection and classification adequacy are the size of the intraclass correlation and the ratio of the within- to between-cluster sample size. In this sense, our findings are in line with Mass and Hox (2004). In the context of LCA, our findings suggest that distribution of class proportions interact with the size of the intraclass correlations and the sample sizes in producing bias in these measures.

In specific terms, the influence of clustering on the BIC suggests that ignoring clustering results in larger values of BIC, which in turn suggests that such a model would not be chosen among a set of competing models. When cluster effects are large as measured by the ICC and between cluster sample sizes are small, the absolute difference in BIC is even larger. In terms of the $R^2_{entropy}$ measure, we find that the bias in this measure begins to exceed the 10% level for an ICC of 0.20 or greater regardless of sample size or class proportion conditions. From a practical viewpoint, ICCs of approximately 0.20 are commonly encountered in studies of student achievement. Thus, our findings are particularly applicable in this setting.

In conclusion, although not every possible condition commonly encountered in applied settings was examined, we view this study as a comprehensive and generalizable examination into the problem of cluster effects in the latent class model. In the context of substantive research, our results suggest that intraclass correlations in combination with the between and within group sample sizes should be routinely examined when latent class analysis is proposed for data arising from clustered sampling. It is hoped that these findings can inform model selection in applications of latent class analysis with clustered samples.

REFERENCES

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling in the social and behavioral sciences* (pp. 81–110). San Francisco, CA: Jossey-Bass.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London, England: Edward Arnold.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine, 22,* 1433–1446.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *Handbook of quantitative methods in psychology* (pp. 592–612). Newbury Park, CA: Sage.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling, 14,* 671–694.

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

Lukočiene, O., & Vermunt, J. K. (2010). Determining the number of components in mixture models for hierarchical data. In A. Fink, L. Berthold, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence.* Berlin, Germany: Springer.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Eds.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage.

Mass, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*, 127–137.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.

Muthén, L. K., & Muthén, B. O. (1998–2007). *M*plus *user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2001). *Early childhood longitudinal study: Kindergarten class of 1998–99: Kindergarten–fifth grade public–use data files user's manual* (Tech. Rep. No. NCES 2001–029). Washington, DC: U.S. Government Printing Office.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12,* 103–124.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Van Horn, M. L., Fagan A. A., Jaki, T., Brown, E. C., Hawkins, M. W., Arthur, J. D., ... Catalano, R. F. (2008). Using multilevel mixtures to evaluate intervention effects in group randomized trials. *Multivariate Behavioral Research, 43,* 289–326.

Vermunt, J. K. (2003). Multilevel latent class models. In R. M. Stolzenberg (Ed.), *Sociological methodology* (Vol. 23, pp. 213–239). Boston, MA: Blackwell.

Vermunt, J. K., & Magidson, J. (2000). *Latent GOLD user's guide: Statistical innovations*. Belmont, MA: Statistical Innovations.