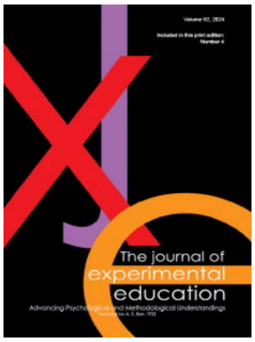

The Journal of Experimental Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/vjxe20

Class Enumeration in Mixture Modeling with Nested Data: A Brief Report

Rashelle J. Musci, Joseph M. Kush, Elise T. Pas & Catherine P. Bradshaw

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2024.2386983

MEASUREMENT, STATISTICS, AND RESEARCH DESIGN

Class Enumeration in Mixture Modeling with Nested Data: A Brief Report

Rashelle J. Musci[a], Joseph M. Kush[b], Elise T. Pas[a], and Catherine P. Bradshaw[c]

[a]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [b]James Madison University, Harrisonburg, VA, USA; [c]University of VA, Charlottesville, VA, USA

ABSTRACT
Given the increased focus of educational research on what works for whom and under what circumstances over the last decade, educational researchers are increasingly turning toward mixture models to identify heterogeneous subgroups among students. Such data are inherently nested, as students are nested within classrooms and schools. Yet there has been limited guidance on which specifications are most appropriate for enumerating latent classes when data are nested. This study utilized longitudinal, state-collected student data to demonstrate the impact of different specifications (i.e., ignoring nested data, using a post-hoc adjustment, and a parametric and non-parametric approach) of a latent class model when analyzing nested data. The overarching goal of this study was to provide the implications of four different model specifications commonly used to adjust for clustering in the context of mixture modeling. We highlight factors that may influence researchers' decisions to employ one approach over another when conducting multilevel mixture modeling. We conclude with a set of recommendations that may be particularly helpful for the use of these methods in educational settings, where nested data is common.

KEYWORDS
Multilevel mixture modeling; multilevel latent class analysis; subgroup; nested data; model specification

Introduction

Nested data are ubiquitous in educational research, as most educational studies include multiple levels of clustering (e.g., students nested in classrooms and/or teachers, staff and students nested in schools, schools nested within districts, observations nested within time). Further, with evolving theories regarding the importance of contextual factors at multiple levels, as well as the increasing availability of multilevel educational data, more researchers are interested in exploring multilevel outcomes and relationships among variables. This type of modeling, when combined with mixture modeling, requires the use of multilevel latent class analysis (Henry et al., 2011). Mixture modeling is so common that it has been mentioned in over 200 hundred manuscripts published in the *Journal of Experimental Education*. Unfortunately, the development of sophisticated methodological tools to appropriately handle mixture modeling within the context of nested data lags behind its widespread analysis. In fact, discussions surrounding ignoring the nested structure of data is not novel, with Miyazaki et al. (2019) recently discussing the implications in Rasch/IRT models. In fact, relatively few studies have systematically explored which methods are most appropriate for addressing

CONTACT Rashelle J. Musci rmusci1@jhu.edu Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Baltimore, MD 21205, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00220973.2024.2386983.

clustered data in the context of mixture modeling. The Huber-White sandwich estimator is highly relied upon, but there is lacking evidence of how accurately it estimates different model types (Freedman, 2006). Best practices to adjust for clustering is particularly unclear within a latent class framework, especially when the latent class indicators vary significantly across level two units.

The current study aimed to address these gaps by exploring the impact of various modeling specifications (i.e., ignoring nested data, using a post-hoc adjustment, and a parametric approach) within a latent class modeling framework on both the number of classes extracted and the estimated parameters. We provide an empirical case example examining student achievement and behavioral outcomes among eighth graders across a state using multilevel latent class analysis. This fills research gaps regarding mixture modeling with nested data and informs the field on the degree to which a range of model specifications can and should be interpreted.

Mixture modeling in educational research

Mixture modeling, particularly latent class/profile analysis (LCA), growth mixture modeling (GMM), and latent transition analysis (LTA), is common in educational research. Mixture modeling enables researchers to identify subgroups of individuals who demonstrate similarities in their data patterns. These approaches are being used with increased frequency, often in the context of (classroom- or school-level) intervention studies to determine for whom interventions are most impactful, or alternatively to identify individuals who may be less responsive to the program or intervention (e.g., Bradshaw et al., 2015). For example, LCA has been used to elucidate subgroups of students across a variety of content areas such as math and reading, classroom behavior, and bullying (Bettencourt et al., 2013; Giang & Graham, 2008; Williford et al., 2011). As an illustration, Lovegrove and Cornell (2014) examined subgroups of bullying behavior in a large sample from Virginia. Bradshaw and colleagues used LPA to identify children with patterns of teacher-rated behavior problems that predicted students' responsivity to the widely used positive behavior support model within the context of a randomized controlled trial (Bradshaw et al., 2015).

While there are challenges and research gaps surrounding the estimation of single-level latent classes, our focus here is on the implications of estimating multilevel latent class models with nested data. Multilevel latent class analysis (MLCA) extends standard LCA, by considering the impact of higher level contextual (e.g., classroom-level, school-level) predictors. Like a single-level LCA, MLCA includes level 1 (e.g., student-level) covariates to predict the probability that an individual belongs to a particular latent class at level 1. Additionally, MLCA allows for the estimation of latent classes at higher levels, such as the school-level (e.g., level 2) and examination of how these higher-level classes relate to individual outcomes. For example, Larson et al. (2020) found that the percentage of students in a school who qualified for free or reduced-price meals, a group or school-level variable, was related to disciplinary outcomes at the student-level. While the impact of group membership likely varies across both construct and group level, the impact is likely not trivial.

Mixture modeling with nested data

The movement toward multilevel analysis emerged from early work by Cronbach (1976) suggesting that much education data was collected and analyzed at the individual level, ignoring cluster effects (e.g., on standard errors), and therefore likely resulted in inaccurate conclusions. Specifically, the intraclass correlation (ICC) is essential to a model with nested data and accounting for cluster effects. Within a multilevel framework, the ICC is defined as the proportion of variance in the outcome that is at the cluster level, or rather it indicates the degree of relatedness between groups, where larger ICCs indicate higher levels of between-cluster variability. On the other hand, ICC estimates are near zero when nested data are independent of each other and there is little to no between-cluster variability. Conceptually, the ICC demonstrates how individuals (e.g., students) who are members of a group (e.g., school) are likely to be more similar to one another than they are to those outside of their group. Ignoring the ICC in a multilevel framework may result in a misspecification of the model, leading to underestimated variance parameters and biased results (Chen et al., 2017; Nagelkerke et al., 2017; Park & Yu, 2016). Therefore, it is important to understand the impact the data structure can have on the analytic model, regardless of whether the researcher is using a model-based or design-based approach. While work related to non-mixture models exists in the literature and standards for proper modeling of nested data has been laid out by experts (e.g., Raudenbush & Bryk, 2002), little is known about the extent to which ignoring or mis-specifying the higher level model can influence the class enumeration process in a mixture model.

In general, extant research demonstrates uncertainty on how to handle nested data during the class enumeration process. Beyond the enumeration process, there are many ways to model nested data once the number of classes at level 1 has been decided (i.e., for parameter estimation). Deciding on the model specification for a multilevel LCA is less straightforward, requiring the consideration of multiple fit statistics and conceptual issues (e.g., hypotheses regarding relationships among variables at different levels; Henry & Muthén, 2010; Park & Yu, 2016). Consider the following two-level logistic regression model:

$$\text{Level 1}: logit[\Pr(y_{ij} = 1)] = \beta_{0j} + \beta_1 x_{ij} \tag{1a}$$

$$\text{Level 2}: \beta_{0j} = \gamma_0 + \gamma_1 w_j + u_{0j}, \tag{1b}$$

in which the probability of the observed binary outcome y_{ij} for unit i in cluster j follows a logit link function, β_{0j} represents a random intercept, β_1 represents the regression coefficient associated with the level 1 covariate x_{ij}, γ_0 represents the overall grand mean, γ_1 represents the regression coefficient associated with the level 2 covariate w_j, and u_{0j} is a residual error term representing random deviation from the grand mean. Combining Equations 1a and 1b, the model can equivalently be expressed using the logistic (logit^{-1}) function:

$$Pr(y_{ij} = 1) = \frac{e^{\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)}}{1 + e^{\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)}}. \tag{2}$$

The specification of this model is important as the very nature of nested data violates conditional independence, a key assumption of standard LCA. This assumption asserts that latent class indicators are independent of each other given the latent class variable. In response, many researchers adopted a multilevel analytic approach to handle nested data. Henry and Muthén (2010) describe the myriad ways to specify a multilevel LCA. For the purposes of this paper, we will give a brief overview of some of the ways educational researchers may model nested data at level 2.

One possible multilevel latent class model specification includes a parametric approach as described in Henry and Muthén (2010). This specification allows for the random means from latent classes at level 1 to be estimated and modeled at level 2 and modeled as a continuous variable that varies across level 2 units. With this approach, N-1 means are modeled at level 2, where N is equal to the number of latent classes at level 1. Because this approach is often computationally intense, researchers have proposed to use a common factor to model the random means at level 2. The comparison of model fit between a level 2 model with the random means, or a common factor model, can and should be explored when deciding how to specify a multilevel LCA.

In contrast, an example of a nonparametric approach was initially described by Bijmolt et al. (2004) and allows for between-unit latent classes to be modeled at level 2, measured by the random means from the level 1 LCA solution. Procedures for deciding the number of latent classes to model at level 2 remain unclear and deserve more focus in future methods research. The implications of varying specifications at level 2 on class enumeration and parameter estimation remain unknown.

Once the measurement at both levels of the multilevel LCA is decided, the inclusion of class predictors at multiple levels is relatively straightforward albeit computationally intense.

Multilevel interventions in education research

Understanding and explicitly modeling what occurs at multiple levels is becoming increasingly important with the movement toward implementing multi-level prevention and intervention programming and with the emergence of more complex student data systems. There has been a major shift in education toward utilizing multi-tiered systems of support (MTSS) frameworks, which invokes the tiered prevention approach. The MTSS framework focuses on the universal provision of behavioral, social, emotional, and/or academic supports for all students; the use of data to identify when students are non-responsive to the universal supports; and provision of selective interventions targeting groups with moderate levels of risk and indicated interventions that are individualized for students at high levels of risk (O'Connell et al., 2009). Studying MTSS adds a level of complexity to educational research by introducing multi-level interventions. Further, more sophisticated and linked student data tracking systems have emerged in response to calls and investments from the federal government (i.e., initiated in the Educational Technical Assistance Act of 2002 and bolstered by the American Recovery & Reinvestment Act of 2009). These multivariate, longitudinal data systems introduce additional complexity to available education data. This makes the methodological gaps even more urgent to address and exemplifies the importance of having a clear understanding of how to specify multilevel models with latent variables.

Current study

The current study utilized longitudinal, state-collected student records and testing data to demonstrate the impact of different specifications (i.e., ignoring nested data, using a post-hoc adjustment, and a parametric approach) of a latent class model when using nested data. A randomly-selected subset of individual student-level data from a state-wide database of educational outcomes was utilized. While the content area and latent construct modeled is important to the educational research, the focus of this paper will be to understand the implications of varying specifications in a multilevel LCA on both fit statistics and the specific class parameters that result from the model and is expected to transcend the education field. These specifications included: (1) completely ignoring the nested nature of the data and modeling a standard LCA; (2) using the easily accessible post-hoc Huber-White adjustment by accounting for the clustering across level 2 units; (3) the parametric approach described by Henry and Muthén (2010) wherein random means from the latent classes at level 1 are modeled at level 2; and (4) a non-parametric approach that uses the random means from level 1 as indicators of latent classes at level 2. Specification three and four move beyond the post-hoc adjustment and considers the true structure of the data. We hypothesized that model specification will impact the distribution of latent classes and individuals in each class based on posterior class probabilities. The overarching goals of this study were to (1) examine the implications of use of each of these four specifications, (2) call attention to the need for future research related to multilevel mixture modeling, particularly in the educational setting where research often utilizes both nested data as well as mixture models, and (3) generate guidance and recommendations on how to specify multilevel LCA models.

Method

Data source

Utilizing a state-wide academic administrative dataset of students nested within schools, we explored a latent class model, performing class membership under varied conditions. The full administrative dataset includes about 850,000 students in elementary through high school, nested within approximately 1,400 schools across one state, representing the state population of public-school students. For the purposes of this analysis, we focus on a random subset of 63,440 8th grade students nested within 319 schools during the 2011 school year. Eighth grade students were chosen to capitalize on heterogeneity in available outcomes because middle school is often a time during which we see increases in suspensions and truancy, therefore ensuring that we would not experience many rare outcomes. See Table 1 for descriptive statistics about latent class indicators, sex, race/ethnicity, and year.

Outcome measures

The following four binary indicators were used in the latent class model: (1) reading proficiency, (2) math proficiency, (3) student suspension, and (4) student truancy. Reading and math proficiency were captured *via* standardized achievement tests wherein students were coded as either 0 = "basic" or 1 = "proficient" or "advanced" based on their original scaled standardized test score. Standard practice for achievement testing is either binary proficiency (as employed here) or categorical achievement levels; continuous scores of the achievement testing lack meaning as there is no standard mapping of a score to proficiency across grade levels. Regarding suspension, students were coded as 0 = not suspended in the given academic year, or 1 = suspended at least once during the academic year. Truancy was calculated within year and a student was considered truant (= 1) if they missed twenty or more days of the school year. Binary coding of these variables are also substantively meaningful, addresses non-normality when considering attendance as counts (e.g., days absent), and is parsimonious within this study by keeping all latent class indicators as the same variable type.

Analytic plan

All models were conducted utilizing the M*plus* version 8 software (Muthén & Muthén, 2017). Consistent with the process for model fitting outlined by Nylund et al. (2007), the analytic plan

Table 1. Fit statistics for different LCA aodels.

No. of Classes	Log Likelihood	SSA BIC	VLMR *p*-value	Entropy	Smallest Class
Model 1: Single level LCA					
1	−98,511.52	197,054.56	NA	NA	NA
2	−90,169.34	180,409.59	< .001	0.681	26.1%
3	−89,942.49	179,995.30	< .001	0.800	4.8%
4	−89,905.15	179,960.02	< .001	0.894	5.0%
Model 2: Single level LCA with cluster adjusted standard errors					
1	−98,511.52	197,054.56	NA	NA	NA
2	−90,169.34	180,409.59	< .001	0.681	26.1%
3	−89,942.49	179,995.30	< .001	0.800	4.8%
4	−89,905.15	179,960.02	< .001	0.894	5.0%
Model 3: Multilevel LCA with classes at level 1					
1	−98,511.52	197,054.56	NA	NA	NA
2	−90,169.34	180,409.59	< .001	0.681	26.1%
3	−89,942.70	179,995.72	< .001	0.794	5.0%
4	−89,905.68	179,961.07	< .001	0.568	6.9%
Model 4: Multilevel LCA with classes at level 1 and level 2					
1	−93,158.74	186,388.41	NA	NA	NA
2	−86,627.57	173,404.86	NA	0.828	12.5%
3	−86,398.09	173,024.69	NA	0.878	1.8%
4	−86,360.57	173,028.46	NA	0.836	1.2%

Note. SSA BIC = the sample size adjusted Bayesian Information Criterion; VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test.

began with class enumeration for a series of unconditional single-level models while assessing standard mixture modeling fit indices (e.g., the sample size adjusted Bayesian Information Criterion [BIC] and the Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test [VLMR-LRT]). In models that do not allow these standard fit indices, we will rely on the log-likelihood and sample size-adjusted Bayesian information criteria (SSA BIC). Models with larger negative log-likelihood values (i.e., closer to zero) are considered indicative of better model fit. In this first set of models (model set 1), the nested structure of the data is ignored and reading, math, suspension and truancy variables are included as indicators of the latent class. In a second set of models (model set 2), the nested nature of the data is accounted for in the class enumeration process utilizing the cluster command in M*plus*, with the cluster variable being the school identifier. The other change in syntax included adding "complex" to the analysis portion of the code, to adjust the standard errors of parameter estimates due to non-independence. In the next set of models (model set 3), we present a parametric approach to multilevel LCA. This approach required both a change to the analysis command and to the model statement. Random latent class means from level 1 were used as indicators to level 2 and allowed to correlate with each other. Example annotated code for this parametric approach can be found in the supplemental material of Henry and Muthén (2010). In the final model set (model set 4), latent classes are estimated at both level 1 and level 2. In this non-parametric approach, random means from the level 1 latent class are used as indicators of a latent class model at level 2 (Henry & Muthén, 2010). All models are compared in terms of standard fit statistics for the set of enumerated models (1 class through 4 classes). Alongside fit statistics, estimated parameters for the best fitting class solution based on the unconditional fixed effects model were compared across each set of models. Missing data was minimal given the use of administrative data. We utilized full information maximum likelihood to address any missing data on the latent class indicators.

Results

Class enumeration

Fit statistics for all four sets of models can be found in Table 1. As expected, fit statistics for model set 1 and 2 are identical. In comparing these sets, the only difference in model estimation is the post-hoc adjustment of the standard errors, therefore there are no differences in standard latent class fit statistics. The set 3 results differ from sets 1 and 2 in the 3- and 4-class solutions (i.e., not for the two-class solution), whereas the model set 4 differs for the 1 class through four class solutions for models 1–3. Both the log-likelihood and BIC shift in model set 4 and we see a notable reduction in the size of the smallest class in both the 3- and 4-class solutions for model 4. When deciding on the number of classes, we must rely on a limited number of fit indices for model set 4 as the LMR-LRT is not available when you estimate a latent class at level 2. Because of the available fit statistics, we move forward with a three-class solution and now present the average latent class probabilities and parameter estimates across each model set. The ICC values for these ranged from 0.02 to 0.29.

Table 2 displays the average latent class probabilities for each model set. While this information should not necessarily be used for deciding the number of classes, it can be useful to understand how well the model estimation is placing individuals within each latent class. Table 2 demonstrates that there are some minor changes in average latent class probabilities when estimating a model at level 2 (model set 3) and when estimating a latent class at level 2 (model set 4). Supplemental Table 2 provides the cross tabulation in both absolute number for the latent classes across the four model sets. While the cross tabulations for model set 1 through 3 indicated that differing measurement models do not necessarily hard classify individuals in different latent classes, the class tabulation with model set 4 indicates substantial changes in hard classification

Table 2. Average latent class probabilities for most likely latent class membership (row) by latent class.

	Class 1	Class 2	Class 3
Model 1			
Class 1	0.974	0.007	0.018
Class 2	0.113	0.755	0.132
Class 3	0.207	0.088	0.705
Model 2			
Class 1	0.974	0.007	0.018
Class 2	0.113	0.755	0.132
Class 3	0.207	0.088	0.705
Model 3			
Class 1	0.974	0.008	0.018
Class 2	0.112	0.759	0.129
Class 3	0.211	0.093	0.679
Model 4			
Class 1	0.754	0.102	0.110
Class 2	0.034	0.871	0.093
Class 3	0.004	0.111	0.880

(see Supplemental Table 3). For example, when comparing the hard classification in model set 1 to model set 4, individuals from class 2 of model set 1 are split across two latent classes in model set.

Parameter estimates

Parameter estimates for the 3-class unconditional model across the 4 model sets are similar (see Supplemental Table 4). As expected based on model specifications, parameter estimates are the same across model sets 1 and 2 except for the standard errors in model set 2 as they are adjusted to account for the clustering of students within schools. Parameter estimates vary slightly in model set 3 when estimating a latent class model in a multilevel framework. These slight changes will likely not lead to alterations in model interpretation. The parameter estimates change significantly in model set 4, these alterations have the potential to change the interpretation of the model.

Discussion

Latent class modeling is incredibly common in educational research, as it is useful and often quite intuitive to explain to stakeholders. Researchers use these models to understand the underlying heterogeneity of a construct in a population. Results from education-based mixture models can be used to inform policy, intervention development and adaptation, and to understand etiology of certain educational outcomes and behaviors. With these important implications, educational researchers must take care in their modeling specifications to account for the structure of their data. This is often fundamentally ignored with mixture modeling as the methods with which to handle nested data is seen as overly time consuming and not impactful on model results. Therefore, this study examined whether modeling specifications are ignorable; our results showed that alterations in the methodology used to handle nested data can have a potentially meaningful impact on model parameters and class proportions, particularly when using a non-parametric approach (i.e., as compared to the more standard approaches of ignoring nesting or accounting for with a cluster command as well as the less prevalent parametric approach). This is not to say that either a parametric or non-parametric approach is the correct model in all cases, but that changes in model specification can alter interpretation of the measurement model. Therefore, more work must be done to both educate researchers on how to handle nested data in mixture modeling, and to understand the data conditions in which differences in model estimation can have the most significant impact on mixture models.

The results from this exemplar study suggest that even when the same number of latent classes are extracted from the measurement model using different specifications, there are differences in the class proportions and parameter estimates. When exploring the cross-tabulations between model set 1 and model set 4, we found that individuals assigned to latent class 2 in model set 1 are evenly split across latent classes 2 and 3 in model set 4. These differences could have significant impacts on the interpretation of findings. Even slight alterations in the individuals hard classified to smaller classes have significant downstream effects. Not surprisingly, we found that parameter estimates differed with notable impacts on interpretations in the non-parametric model (model 4). This suggests caution should be taken when deciding on the number of classes and interpreting parameter estimates in a latent class model with nested data. In fact, prior research by Chen et al. (2017) suggest that standard fit statistics do not perform well during the class enumeration process while ignoring the structure of the data. It is important to note here that we do not suggest a one-size fits all approach to multilevel latent class analysis. When specifying the measurement model, one should consider the ICCs as well as any hypotheses regarding heterogeneity across level two units. While there remains more methodological work to be done, we recommend careful exploration of data prior to determining the measurement model to be used for the latent class analysis.

Limitations and directions for future research

The overall goal of this study was to demonstrate whether and what differences arise when estimating a mixture model with real-world nested education data. Because we utilized real data, we were unable to alter specific parameters of the data structure, therefore limiting the conclusions that we can make from the present results. Future work should utilize simulated data that alters the ICCs of the indicators across level 2 units such that some datasets have high ICCs and some have very low ICCs. Worth noting, the ICCs for the current data ranged from 0.06 to 0.15. These numbers are on the smaller side when comparing ICCs in other studies. For example, Hedges and Hedberg (2007) found ICCs that ranged from 0.17 to 0.27. While some may suggest that use of multilevel models is not needed when ICCs are small, this has been demonstrated to be incorrect (Huang, 2018). However, it is possible that because our real-world data has relatively low ICCs, the impact of model misspecification is not large. Additionally, we also estimated our latent class models using real-world data, without any knowledge of the "true" latent classes and therefore we are unable to determine which latent class estimation was most accurate to the underlying structure of the data. Simulation studies should focus on the creation of datasets with a known latent class structure so that we can determine which model fits the data best. A second goal of this study was to demonstrate how parameter estimates of the unconditional model differed across varying measurement models. However, we did not explore the potential impact of either categorical or continuous covariates on latent class membership. Future research should expand these analyses to identify the impact of covariates in various model estimation strategies in both simulated data and real-world data. Further, we also only used a small number of possible specifications for the level 2 model. We did not present the non-parametric approach that estimates a latent class model at both level 1 and level 2. Future work should explore other specifications, and perhaps novel ones, to fully understand multilevel latent class estimation. While it is possible that these other approaches would yield different results, the ultimate goal of this study was to call attention to the fact that model specification can significantly impact the class enumeration and model interpretation process. The addition of other model specification would likely add support to this concept.

Conclusions and recommendations

Taken together, these findings suggest that, similar to Rasch/IRT models, the modeling approach to accounting for the structure of the data when enumerating latent classes can have consequences on the estimated parameters and the number of classes extracted (Park & Yu, 2016). Most significantly, a non-parametric approach yields a notably different outcome to all other three approaches to handling the clustering (i.e., ignoring nested data, employing a *post hoc* adjustment for clustering, and a parametric approach). These findings contribute to the relatively small body of literature on multilevel mixture modeling and the impact modeling decisions have on the interpretation of the findings. We hope that this demonstrative analysis aids researchers interested in using mixture modeling with multilevel data, and more specifically encourages them to think critically about model estimation decisions We are not suggesting that any of the model specifications presented here are superior, without simulation studies we are unable to determine which specification was able to extract the latent class structures correctly. We do, however, recommend that researchers carefully consider the measurement model when estimating mixture models in nested data. We also urge quantitative researchers to develop easy to follow protocols for substantive researchers to follow when deciding on model specification. General rules of thumb could be developed based on size of ICCs such that researchers are pointed to a set of model specifications to explore in their own data. With improved software, MLCA is easier to implement and will likely be the appropriate analytic model for many seeking to identify subgroups in educational data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H150027 (PI: C. Bradshaw) to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education or the Maryland State Department of Education.

ORCID

Rashelle J. Musci http://orcid.org/0000-0001-7267-5822
Joseph M. Kush http://orcid.org/0000-0003-0183-494X
Elise T. Pas http://orcid.org/0000-0001-5000-4681
Catherine P. Bradshaw http://orcid.org/0000-0003-2048-3225

References

American Recovery and Reinvestment Act of 2009. (2009). Public Law 111-5. https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf

Bettencourt, A., Farrell, A., Liu, W., & Sullivan, T. (2013). Stability and change in patterns of peer victimization and aggression during adolescence. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *42*(4), 429–441. https://doi.org/10.1080/15374416.2012.738455

Bijmolt, T. H. A., Paas, L. J., & Vermunt, J. K. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, *21*(4), 323–340. https://doi.org/10.1016/j.ijresmar.2004.06.002

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology*, *107*(2), 546–557. https://doi.org/10.1037/a0037630

Chen, Q., Luo, W., Palardy, G. J., Glaman, R., & McEnturff, A. (2017). The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. *SAGE Open*, *7*(1), 215824401770045. https://doi.org/10.1177/2158244017700459

Cronbach, L. J. (1976). Research on classrooms and schools: formulation of questions, design and analysis.

Freedman, D. A. (2006). On the so-called "huber sandwich estimator" and "robust standard errors. *The American Statistician*, *60*(4), 299–302. https://doi.org/10.1198/000313006X152207

Giang, M. T., & Graham, S. (2008). Using latent class analysis to identify aggressors and victims of peer harassment. *Aggressive Behavior*, *34*(2), 203–213. https://doi.org/10.1002/ab.20233

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., & Dymnicki, A. B. (2011). Influence of school-level variables on aggression and associated attitudes of middle school students. *Journal of School Psychology*, *49*(5), 481–503. https://doi.org/10.1016/j.jsp.2011.04.007

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 193–215. https://doi.org/10.1080/10705511003659342

Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, *33*(3), 492–499. https://doi.org/10.1037/spq0000272

Larson, N., Fulkerson, J. A., Berge, J. M., Eisenberg, M. E., & Neumark-Sztainer, D. (2020). Do parents perceive that organized activities interfere with family meals? Associations between parent perceptions and aspects of the household eating environment. *Journal of the Academy of Nutrition and Dietetics*, *120*(3), 414–423. https://doi.org/10.1016/j.jand.2019.11.008

Lovegrove, P. J., & Cornell, D. G. (2014). Patterns of bullying and victimization associated with other problem behaviors among high school students: A conditional latent class approach. *Journal of Crime and Justice*, *37*(1), 5–22. https://doi.org/10.1080/0735648X.2013.832475

Miyazaki, Y., Chungbaek, Y., Shropshire, K. O., & Hedeker, D. (2019). Consequences of ignoring nested data structure on item parameters in Rasch/1P-IRT model. *Behaviormetrika*, *46*(2), 401–434. https://doi.org/10.1007/s41237-019-00090-8

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nagelkerke, E., Oberski, D. L., & Vermunt, J. K. (2017). Power and type I error of local fit statistics in multilevel latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 216–229. https://doi.org/10.1080/10705511.2016.1250639

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

O'Connell, M. E., Boat, T., & Warner, K. E, Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions; Institute of Medicine; National Research Council. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. The National Academies Press.

Park, J., & Yu, H.-T. (2016). The impact of ignoring the level of nesting structure in nonparametric multilevel latent class models. *Educational and Psychological Measurement*, *76*(5), 824–847. https://doi.org/10.1177/0013164415618240

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (Vol. 1). Sage.

Williford, A. P., Brisson, D., Bender, K. A., Jenson, J. M., & Forrest-Bank, S. (2011). Patterns of aggressive behavior and peer victimization from childhood to early adolescence: A latent class analysis. *Journal of Youth and Adolescence*, *40*(6), 644–655. https://doi.org/10.1007/s10964-010-9583-9