

Ten Frequently Asked Questions about Latent Class Analysis

Karen Nylund-Gibson, Ph.D.  
University of California, Santa Barbara  
Department of Education  
Santa Barbara, CA 93106  
knylund@education.ucsb.edu

Andrew Young Choi, M.A.  
University of California, Santa Barbara  
Department of Counseling, Clinical, and School Psychology  
Santa Barbara, CA 93106  
a Choi@ucsb.edu

Author Note

Karen Nylund-Gibson, Department of Education, University of California, Santa Barbara; Andrew Young Choi, Department of Counseling, Clinical, and School Psychology, University of California, Santa Barbara. Correspondence concerning this article should be addressed to Karen Nylund-Gibson at [knylund@education.ucsb.edu](mailto:knylund@education.ucsb.edu)

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A160157 to the University of California, Santa Barbara. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Nylund-Gibson, K., & Choi, A. Y. (In press). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*.

© 2018, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/tps0000176

## ABSTRACT

Latent class analysis (LCA) is a statistical method used to identify unobserved subgroups in a population with a chosen set of indicators. Given the increasing popularity of LCA, our aim is to equip psychological researchers with the theoretical and statistical fundamentals that we believe will facilitate the application of LCA models in practice. In this article, we provide answers to 10 frequently asked questions about LCA. The questions included in this article were fielded from our experience consulting with applied researchers interested in using LCA. The major topics include a general introduction in the LCA; an overview of class enumeration (e.g., deciding on the number of classes), including commonly used statistical fit indices; substantive interpretation of LCA solutions; estimation of covariates and distal outcome relations to the latent class variable; data requirements for LCA; software choices and considerations; distinctions and similarities among LCA and related latent variable models; and extensions of the LCA model. To illustrate the modeling ideas described in this article, we present an applied example using LCA. Specifically, we use LCA to model individual differences in positive youth development among college students and analyze demographic characteristics as covariates and a distal outcome of overall life satisfaction. We also include key references that direct readers to more detailed and technical discussions of these topics for which we provide an applied and introductory overview. We conclude by mentioning future developments in research and practice, including advanced cross-sectional and longitudinal extensions of LCA.

*Keywords:* classification, finite mixture models, latent class analysis, latent profile analysis, heterogeneity

## PUBLIC SIGNIFICANCE STATEMENT

In this article, we answer ten frequently asked questions about the technical and applied underpinnings of latent class analysis (LCA), a statistical approach to understanding unobservable within-group differences in a population. Our goal is to provide readers with an introductory and conceptual understanding of LCA to inform appropriate application and interpretation of these models in research practice.

### Ten Frequently Asked Questions about Latent Class Analysis

Latent class analysis (LCA) is an analytic technique that has become increasingly popular among psychological researchers. The main goal of LCA is to identify groups, called latent classes, based on responses to a set of observed indicators. LCA has been used to identify subgroups in a range of substantive areas, including differential diagnosis among mental disorders (Cloitre, Garvert, Weiss, Carlson, & Bryant, 2014), profiles of school readiness among youth (Quirk, Nylund-Gibson, & Furlong, 2013), and patterns of acculturation among ethnic minorities (Jang, Park, Chiriboga, & Kim, 2017). LCA belongs to a larger family of latent variable techniques called finite mixture models, which comprises a wide range of cross-sectional and longitudinal models that all involve one or more latent class variables.

In this article, we compile what we believe to be the top 10 questions that a researcher new to LCA would ask. These questions are not exhaustive, but we hope that our introductory discussion and an illustrative applied example of LCA will establish an appropriate foundation for understanding these models and the potential contribution they can make to a range of areas in psychological research. We include references of applied, statistical, and theoretical articles throughout to facilitate further learning in these respective areas.

Given the nascence of mixture models in psychological research, our responses to the questions in this article are conditioned on the current status of scholarly knowledge on these models and focus on the latent class analysis model specifically (rather than all mixture models). Although there are many areas where best practices for the application of these models have not been firmly established, active methodological research continues to advance our understanding of how LCA, and mixture models more generally, perform in data analysis conditions. Notwithstanding the infancy of LCA compared to other modeling traditions like factor analysis,

however, there has been a steady growth of publications in which LCA is applied, as illustrated in Figure 1.

### **1. What is Latent Class Analysis (LCA)?**

LCA is a statistical method that can be used to identify and describe latent classes, or “hidden groups,” within a population. Classes may be thought of as unobserved “subgroups” or “typologies” that characterize heterogeneity in a population with respect to a given phenomenon. Concretely, this method uses responses to a chosen set of indicators to identify groups of people that are alike each other. One useful way to place LCA among other common latent variable models is to think of LCA as a way to group similar people together, whereas factor analysis groups items. The groups in LCA are unobserved (latent) and are based on the individuals’ set of responses to the indicators. Thus, LCA can be thought of as a “person-centered” approach to creating empirically-derived typologies, contrasting the dominant “variable-centered” tradition that generally requires arbitrary cutoffs for classifying or differentiating among individual cases (Nylund, Bellmore, Nishina, & Graham, 2007). Furthermore, unlike other classification techniques such as cluster analysis or k-means clustering, LCA is model-based and permits a mathematical evaluation of how well a proposed LCA model represents the data. Finally, LCA is a cross-sectional model with categorical outcomes, where binary indicators are most commonly used in practice.

The statistical expression of the LCA model describes the key parameters in the model. Suppose there are  $M$  binary latent class indicators,  $u_1, u_2, \dots, u_M$  observed on  $n$  individuals. The LCA model assumes that the  $M$  indicators are all reflective measures of an underlying unordered categorical latent class variable,  $c$ , that has  $K$  latent classes. There are two parameters of interest in an LCA model without covariates (e.g., an unconditional LCA): the relative size of the latent

classes and the conditional item probabilities. First, the relative size of each class, or the proportion of individuals in a given latent class  $k$ ,  $P(c = k)$ , is denoted by  $\pi_k$ . This parameter describes the relative size of each of the emergent latent classes. The LCA model implies that the  $K$  classes are mutually exclusive and exhaustive, meaning that each individual in the population has membership in exactly one of the  $K$  latent classes and  $\sum \pi_k = 1$ .

Similar to the typical default model specification in traditional factor analysis (e.g., one without any correlated disturbances), the LCA model assumes *local independence* for the set of  $M$  indicators conditional on class membership, thus often termed the *conditional independence assumption*. This assumption implies that latent class membership explains *all* of the shared variance among the observed indicators. Stated differently, any association among the observed indicators is assumed to be entirely explained by the latent class variable, and once the latent class variable is modeled the indicators are no longer associated. The joint distribution of all the indicators and the latent class variable,  $c$ , under the conditional independence model is expressed as

$$\Pr(u_{1i}, u_{2i}, \dots, u_{Mi}) = \sum_{k=1}^K \left[ \pi_k \cdot \left( \prod_{m=1}^M \Pr(u_{mi} | c_i = k) \right) \right]. \quad (1)$$

Because we generally use binary indicators in LCA, the model estimates a logit to capture the association between the observed indicator and the latent class variable. This variable is akin to a factor loading in factor analysis, but because the variable is categorical, a logistic regression is used in LCA instead of standard regression. A logit value for each item is estimated for each latent class. The class-specific indicator probabilities, or conditional item probabilities, are parameterized as single thresholds on the inverse logit scale as follows

$$\Pr(u_{mi} = 1 | c_i = k) = \frac{1}{1 + \exp(\tau_{mk})}, \quad (2)$$

whereas the class proportions are parameterized as intercepts on the inverse multinomial scale such that

$$\pi_k = \Pr(c_{ik} = k) = \frac{\exp(\gamma_{0k})}{\sum_{j=1}^K \exp(\gamma_{0j})}, \quad (3)$$

with  $\gamma_{0K} = 0$  for identification.

The second key parameter in the LCA model, then, is the set of conditional item probability parameters,  $\tau_{mk}$ . These parameters describe the relations between the observed indicators and the latent categorical variable (similar to factor loadings in factor analysis). These conditional item probabilities can be considered the measurement parameters in the unconditional LCA model, whereas the structural parameters are the relative class size parameters,  $\gamma_{0k}$ , and describe the relative sizes of the emergent latent classes. Although it is possible to partially relax the conditional independence assumption, as sometimes done in factor analysis models by estimating correlated disturbance terms, researchers should rely on *a priori* substantive knowledge and proceed with caution as to consider the sample-specific (and potentially idiosyncratic) nature of observed covariation that are ostensibly expressed as violations of this assumption. Other types of mixture models—namely growth mixture models (B. O. Muthén & Shedden, 1999; Petras & Masyn, 2010; Ram & Grimm, 2009) and factor mixture models (Clark et al., 2013; Lubke & Muthén, 2005)—are somewhat better suited for relaxing the conditional independence assumption, but they are not without potential problems in terms of the formation and interpretation of the emergent classes in the modeling process.

## 2. How do you decide on the number of classes?

The decision of how many classes to retain in an LCA is a critical—yet often arduous—modeling step in LCA. This process is often called class enumeration and is similar to deciding on the number of factors to retain in an exploratory factor analysis. Class enumeration includes fitting several LCA models with differing numbers of latent classes, collecting and tabulating fit information for each fitted model, and studying patterns to decide on how many classes best describe the patterns observed in the data. As in conventional structural equation modeling (SEM), there is not a single fit index that is agreed upon for use in enumerating classes, but rather, we use a set of fit indices to decide. In addition, because fit indices often do not all point to a single solution, the recommended procedure for exploring and deciding on the number of classes is to jointly consider statistical fit indices, substantive interpretability and utility, and classification diagnostics, which help to illuminate how well the classes are classifying and differentiating among the individuals considered (Masyn, 2013; B. O. Muthén, 2003).

It is recommended that we begin the modeling process by estimating a 1-class LCA model, which is a model that is simply estimating the observed indicator endorsement probability for the item set (the observed item proportions in the sample statistics). This 1-class model serves as a comparative baseline for models with more than one class. Then, we increase the number ( $k$ ) of classes by one, examining whether the addition of each class results in conceptually and statistically superior solutions. We usually stop estimating additional classes when empirical under-identification (e.g., overparameterization) or convergence issues are encountered (which is generally indicated by error messages from the software being used). Once all the LCA models are fitted, we collect fit information from each one (e.g., copy/paste the information) and summarize them in a single table for ease of evaluation.

### **Fit Indices used in LCA**



This section describes guidelines for fit indices that are currently supported for use in deciding on the number of classes in LCA models (Masyn, 2013; Morgan, 2015; Morovati, 2014; Nylund, Asparouhov, & Muthén, 2007; Yang, 2006). These fit indices have been examined in simulation studies and have been shown to generally work well in identifying the correct number of classes in an LCA. In any given data analysis setting, however, we do not know the “true” correct number of classes, so it is hoped that these fit indices can help us to approximate to the correct number of classes, assuming that latent classes truly “exist” in a given population. Table 1 provides a summary of the commonly used fit indices, references, guidelines for use, and how to obtain the values.

First are information criteria (IC)—including the Bayesian Information Criterion (BIC), Sample-size adjusted Bayesian Information Criterion (SABIC), Consistent Akaike Information Criterion (CAIC), and Approximate Weight of Evidence Criterion (AWE)—which are approximate fit indices where lower values indicate superior fit. For example, in a data analysis setting we would compare the values of the BIC across the set of fitted models and would pick the model that has the lowest value of the BIC. That is, the model preferred by the BIC is the model with the lowest value among the set being considered. It can be useful to plot the values of the BIC, ABIC, CAIC, and AWE to visually display the values and provide for easy inspection (see, for example, Figure 2). In practice, it is not uncommon that the BIC continues to decrease for each additional class added (e.g., there is not a global minimum) and in these instances these plots can be particularly useful to inspect for an “elbow” of point of “diminishing returns” in model fit (e.g., small decreases in the IC for each additional latent class), akin to how one interprets a scree plot in exploratory factor analysis.

Second are the likelihood-based tests—the Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test (VLMR-LRT) and the bootstrapped likelihood ratio test (BLRT)—which provide  $p$ -values assessing whether adding a class leads to a statistically significant improvement in model fit. These likelihood-based tests compare the fit between two neighboring class models (e.g., a 2-class versus a 3-class model). A non-significant  $p$ -value for a  $k$  class solution thus lends support for the  $k - 1$  class solution. See Nylund, Asparouhov, and colleagues (2007) for more details on ICs and likelihood-based tests and their utility in enumerating classes in the LCA context.

Third, are two indices commonly used in the Bayesian modeling context that can be used to compare LCA models. The Bayes Factor ( $BF$ ) is used as a pairwise comparison of fit between two neighboring class models. The  $BF$ , which is generally not directly provided by statistical software program and thus require calculation by the user (see Table 1), represents the ratio of the probability that each model being compared is true. More specifically,  $1 < BF < 3$  suggests “weak” support for the model with less classes,  $3 < BF < 10$  suggests “moderate” support, and  $BF > 10$  suggests “strong” support (Wagenmakers, 2007; Wasserman, 1997). Finally, correct model probability ( $cmP$ ) provides an estimate of each model being “correct” out of all models considered, assuming that the “true” model is indeed among them; the model with the largest value is selected. See Masyn (2013) for more details on the  $BF$  and  $cmP$ , as well as other considerations on fit indices.

As alluded to previously, it is not uncommon that the collection of fit information described here do not converge on one single model. It is more common that the fit indices support one or two candidate models. In this case, we recommend that fit for these candidate model be presented and that the solutions be studied closely (Masyn, 2013; B. O. Muthén, 2003; Ram & Grimm, 2009). Researchers should consider how the selected models relate to each other

(e.g., is one an expanded version of the other?) and the stability of the different models, including the relative sizes of the emergent classes (e.g., are the emergent classes relatively large—say, having more than 5-8% of the sample and a decent number of individuals in the class?). It has been shown through simulation studies of a variety of different types of mixture models (Depaoli, 2013; Morgan, 2015; Morovati, 2014; Tofighi & Enders, 2008; Tueller & Lubke, 2010) that small or “rare” classes are generally difficult to recover at small sample sizes and when class prevalences are highly unequal (e.g., the classes are not the same size). As such, we encourage researchers to be mindful as to avoid selecting an overextracted and potentially unstable class solution, especially when lacking a large enough sample size that may otherwise support the reliable detection of classes with low prevalences. Additionally, when the available ICs, like the BIC, fail to reach a minimum, some have suggested that this may be an indication that the particular mixture model may not be correctly specified and that other models may be more appropriate. This may be another type of mixture model (e.g., factor mixture model), one with correlated disturbances (Asparouhov & Muthén, 2015), or perhaps not a mixture model at all, although more research is needed in this area. Taking together these considerations, as well as results of auxiliary variable analysis, a final model can be selected.

### **Applied Example: Positive Youth Development Inventory (PYDI) Analysis**

We illustrate the modeling ideas described in this article using 7 items (Table 2) from the PYDI Contribution subscale (Arnold, Nott, & Meinhold, 2012) that were administered to 1629 college and university students. The PYDI measures behavioral, psychological, and social characteristics that are theorized to indicate positive youth development, and the Contribution Subscale specifically measures the degree to which youth express values and behaviors associated with channeling their positive psychosocial strengths to contribute meaningfully to

their local community (Lerner et al., 2005). Our research question was to examine whether this construct of positive contribution was comprised of qualitatively distinct subtypes among college and university students, and to further examine whether such subtypes (if indeed were present) were meaningfully associated with a demographic predictor and a wellbeing-related outcome.

We first dichotomized the items at the midpoint of the original 4-point Likert metric (1 = *strongly disagree* to 4 = *strongly agree*), where higher scores indicated more positive development. Using the dichotomized PYDI Contribution Subscale items, we fitted a series of LCA models beginning with a 1-class model up until a 6-class model and collected their fit information as reported in Table 3. Fit indices did not converge on a single solution, and this is generally the rule rather than the exception in applied practice. The ICs and the *cmP* suggested a 3-class solution, whereas the likelihood tests supported a 4-class solution. Evaluating Figure 2, the most prominent “elbow” was at the 2-class model, whereas the lowest point for the ICs was at the 3-class model. The *BF* suggested that the 3-, 4-, and 5-class solutions were plausible. Given that the BLRT specifically has been shown to be robust across a diversity of modeling conditions (Nylund, Asparouhov, et al., 2007), we tentatively selected the 4-class solution.

### **3. How do you interpret the item probability plot?**

Although LCA is an exploratory technique, substantive theory and model utility (including the parsimony principle) should contextualize the interpretation of fit indices and the selection of the final model (Masyn, 2013), as with all latent variable models. Indeed, Morgan (2015) advises that model selection will be facilitated to the extent that researchers can anticipate the prevalence of theoretically expected classes and their degree of differentiation. As incongruent information among fit indices is common, it is crucial to gauge the conceptual meaningfulness and plausibility of each class solution when interpreting and labeling the classes

(B. O. Muthén, 2003). One approach for doing so is to visually inspect the item probability plot and to examine the qualitative differences among the classes.

After a solution is selected and interpreted, there are several guidelines for evaluating how well the classes are differentiated (Masyn, 2013). First, entropy is an omnibus index where values  $> .80$  indicate “good” classification of individual cases into classes (Clark & Muthén, 2009). Second, average posterior probabilities (AvePP) provide information about how well a given model classifies individuals into their most likely class. Individuals’ AvePP values are reported for their most likely class assigned, where values  $> .70$  indicate well-separated classes (Nagin, 2005). Third, class homogeneity reflects how similar people are to each other with respect to their item responses in each class, where conditional item probabilities  $> .70$  and  $< .30$  indicate high homogeneity. It can be useful to visually evaluate class homogeneity using the item probability plot. Fourth, class separation is how dissimilar people are across classes in their item responses, where odds ratios of item probabilities between two classes  $> 5$  and  $< .20$  indicate high separation. Notably, items may differentiate some classes well but not others. Researchers should thus consider holistically how well each item contributes to class separation for the overall model.

### **PYDI Analysis**

Figure 3 illustrates the conditional item probabilities for the selected 4-class solution, where “conditional” again refers to the likelihood of endorsing each item as a function of class membership (e.g., the probabilities are “conditioned” on class). The model indicators are labeled on the x-axis whereas the y-axis presents the metric of the item probabilities (0 to 1). The four classes are defined by the crisscrossing lines, and their preliminary labels are listed at the bottom of Figure 3 with class prevalence (e.g., relative class sizes) in parentheses. We referred to the

indicators in Table 2 in evaluating the substantive meaning of the joint patterns of item responses that emerged within each class. The first and largest class is characterized by high response probabilities for all indicators and were thus labeled the *Holistic-Collaborative* class. Youth in this class are likely to value social contribution overall and pursue pertinent activities in cooperation with others. The second class was labeled *Altruistic-Low Self-efficacy* and is likely to be comprised of youth who highly value social contribution yet do not believe in their own effectiveness for impactful engagement with their community. The third class was labeled *Low Engagement* given the characteristically low or ambivalent response probabilities for all model indicators. The fourth class was labeled *Holistic-Independent* given their similarity to the *Holistic-Collaborative* class apart from two indicators measuring interpersonal and community engagement. Youth in this class are likely to value social contribution but may prefer pursuing relevant activities in a more independent or introverted manner. Examining Figure 3, it is apparent that these two classes are quite homogenous (and thus not well-separated) in their item responses except on items 1 and 3, which may diminish the model classification statistics.

As seen in Table 3, our entropy value approached but did not meet the recommended cutoff, suggesting non-ignorable overlap among the classes. Examining Table 4, it further appears that this lack of differentiation is localized to the *Holistic-Independent* class for which AvePP is only slightly better than chance. More specifically, youth who were modally assigned to the *Holistic-Independent* class also had an approximately 44% chance of belonging in the *Holistic-Collaborative* class, which is unsurprising given their similar conditional item probability patterns. Nevertheless, the AvePP were high for the remaining classes. Finally, we evaluated class homogeneity and separation considering the statistical heuristics described above and the item probability plot. As examples, it appears that items 5, 6, and 7 (see Table 2 and

Figure 3) demonstrate high class homogeneity where youth are either highly likely or very unlikely to endorse this subset of items. In terms of separation, item 7 clearly distinguishes the Low Engagement class from the other three classes, but not among those three remaining classes.

#### 4. How do you include covariates and distal outcomes into an LCA model?

It is often of substantive interest to identify and examine the antecedents and consequences of latent class membership. Covariates and distal outcomes, often called auxiliary variables, provide a context for understanding more about the emergent latent classes and the people who comprise them. The estimation of the relations among the auxiliary variables and the latent class variable is a very active area of methodological research, where new methods and approaches are being developed. Whereas an intensive discussion is beyond the scope of this article, broadly speaking, covariates can be used to explore whether class prevalence is equivalent across levels of a predictor of class membership (e.g., treatment vs. no treatment). Distal outcomes can be used to examine whether the latent classes display statistically significant mean-level differences in the selected distal outcome variables (e.g., mental health).

The relations between latent class membership and a covariate is expressed as a multinomial logistic regression model, given by

$$\Pr(c_i = k | x_i) = \frac{\exp(\gamma_{0k} + \gamma_{1k}x_i)}{\sum_{j=1}^K \exp(\gamma_{0j} + \gamma_{1j}x_i)}, \quad (4)$$

where parameters of the last class are set to zero for identification, specifically  $\gamma_{0K} = \gamma_{1K} = 0$ , and the indicators are independent of the covariate conditional on class membership such that the joint distribution of the  $u$ 's is expressed as

$$\Pr(u_{1i}, u_{2i}, \dots, u_{Mi} | x_i) = \sum_{k=1}^K \left[ \Pr(c_i = k | x_i) \cdot \left( \prod_{m=1}^M \Pr(u_{mi} | c_i = k) \right) \right]. \quad (5)$$

For distal outcomes, effects across classes are examined by estimating class-specific mean and variance estimates for each distal outcome and then conducting pairwise comparisons to determine where among the classes the distal outcomes are significantly different. Examples of how to estimate these relations in *Mplus* are presented in Appendices 1 and 2.

Ideally, the joint estimation of the latent class measurement model and the association among the latent class variable and auxiliary variables would be conducted in a 1-step method as commonly done in a general SEM context. In mixture modeling, however, this is not as straightforward because including an auxiliary variable can and may unintentionally influence the formation of the latent class variable, both in relative class *size* and *type* (Asparouhov & Muthén, 2014; Nylund-Gibson, Grimm, Quirk, & Furlong, 2014; Vermunt, 2010). Specifically, the joint estimation of the measurement model and auxiliary relations can lead to vastly different latent class solutions, which has led to the recent research activity around three-step methods which circumvent these issues. Recent simulation studies have recommended that we enumerate classes prior to estimating auxiliary variable relations (Nylund-Gibson & Masyn, 2016).

Over the last 12 years, at least 7 different approaches on how to include auxiliary analyses have been proposed as a way of dealing with these issues where, generally, the latent classes are estimated independently from the association among the latent class variable and auxiliary variables (Grimm, Nylund-Gibson, & Masyn, 2018). Without discussing each method extensively, we focus on the current best practice, which is to use either the three-step (Vermunt, 2010) or Bolck, Croons, & Hagnaars (BCH) method (Asparouhov & Muthén, 2014; Bolck, Croon, & Hagnaars, 2004; Vermunt, 2010). More specifically, the current recommendation is to separate the class enumeration step from any subsequent structural analyses, wherein classes are enumerated solely with the chosen latent class indicators measuring the substantive domain of



interest (Nylund-Gibson & Masyn, 2016). Once a final class solution is determined, the 3-step or the BCH method can be used to estimate covariate and/or distal outcome relations. In these approaches the measurement parameters of the latent classes are held fixed while accounting for classification error, and then auxiliary variables are subsequently included and their relation to the latent class variable is estimated. Nylund-Gibson, Grimm, Quirk, and Furlong (2014) present a detailed discussion about using the three-step method. For more guidance on the BCH method as implemented in *Mplus*, see Asparouhov and Muthén (2014). Both articles include *Mplus* syntax for using these methods with LCA and advanced mixture models.

### **PYDI Analysis**

We demonstrate the use of the manual 3-step method by examining a binary covariate of race (e.g., Hispanic vs. non-Hispanic) and a continuous distal outcome of life satisfaction as they relate to the PYDI Contribution Subscale latent classes (see Figure 4). It should be noted that we included the direct relation between life satisfaction and race to control for the direct association among these variables; this path was statistically non-significant. See the Appendices 1 and 2 for *Mplus* syntax for the specification of the model in Figure 4.

Following the three-step procedure, the logit values of the classification probabilities for the 4-class solution for the first step were copied to be used in the third step. We used these values in a second set of analyses to fix the measurement parameters of the latent classes, and the auxiliary variables relations were estimated thereafter (see Appendix 2). We used multinomial logistic regression to evaluate whether the relative proportions of Hispanic and non-Hispanic youth were equal across the four classes, and the results are reported in Table 5. Notably, Hispanic youth (vs. non-Hispanic) were more likely to be in the *Holistic-Collaborative* class compared to both the *Altruistic-Low Self-efficacy* class ( $OR = 2.71, p = .001$ ) and the *Holistic-*

*Independent* class ( $OR = 1.62, p = .009$ ). No other covariate-class relations were statistically significant. Simultaneously, we estimated class-specific means of life satisfaction for each of the four classes. For the interpretation of distal outcomes, we centered the covariate such that the distal outcome mean differences across the latent classes accounted for the relative proportion of Hispanic vs. non-Hispanic youth in the entire sample. Pairwise Wald tests revealed that life satisfaction in the *Holistic-Collaborative* class was significantly higher compared to each of the other three classes. No other distal mean comparisons were statistically significant.

### **5. What is the required sample size for LCA?**

A common question with any latent variable technique is about the required sample size. As in all SEM, there is not one fixed minimum sample size in LCA as this depends strongly on multiple study conditions that vary in practice, including the number of indicators included in the model, how well the indicators differentiate classes, how well-separated the classes are, and the relative sizes of the classes, among other aspects. To date, there are no concrete guidelines for sample size for LCA models, although a heuristic of  $N \geq 500$  has been suggested based on two simulation studies (Finch & Bronk, 2011). More generally, however, simulation study results have been mixed and many LCA applications have a much smaller sample size.

Drawing from previous methodological research, it appears that  $N \approx 300-1000$  is roughly the range in which most commonly used fit indices for mixture models can be expected to function adequately (Morgan, 2015; Morovati, 2014; Nylund, Asparouhov, et al., 2007; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2008; Yang, 2006). For simple LCA models with a pair of well-separated classes, a sample size as small as 30 may be sufficient. For larger and more commonly encountered LCA models with many indicators and many classes, larger samples should be used to ensure that there is sufficient power to estimate all model parameters and

recover the true number of classes (Lubke & Luningham, 2017). Inadequate sample size can lead to convergence issues, insufficient power to detect classes, and improper and unstable solutions. Further, small samples may mask small, yet substantively relevant classes that are hard to detect because of their low prevalence (Masyn, 2013). Although simulation studies can be used to help decide on the necessary sample size for a given model, they require that the “true” model be specified, which is often unknown given the exploratory nature of LCA (see L. K. Muthén & Muthén, 2002). For an example of how to set up simulation study in *Mplus* using an LCA model, see example 12.4 in the *Mplus* user’s guide (L. K. Muthén & Muthén, 1998-2017).

When deciding on sample size, researchers should consider both the prevalence for any given class and the coverage for the class-specific parameters that are being estimated (Lubke & Luningham, 2017). Researchers often have a substantive interest in rare classes but, as aforementioned, these cannot be detected without a sufficient overall sample size to estimate classes with low relative prevalences (say, 1-8%), especially if a large number of indicators have been included and many class-specific parameters are freely estimated (Masyn, 2013). For multiple types of mixture models, simulation studies have shown that beyond sample size per se, the likelihood of accurately recovering the correct number of mixture classes depends heavily on class separation (Depaoli, 2013; Lubke & Muthén, 2007; Lubke & Neale, 2006; Tein et al., 2013; Tofighi & Enders, 2008; Tueller & Lubke, 2010). Adding well-separating items to an LCA has been shown to improve correct latent class recovery and to help mitigate the destabilizing effects of small sample size (Wurpts & Geiser, 2014). However, the nature and number of classes are often difficult to anticipate in applied modeling scenarios, including the selection of items that can effectively identify the subgroups that are hypothesized to exist in the population.

As in other statistical approaches, item choice in an LCA should be motivated by well-defined research questions. It has been suggested that researchers select items that as a set covers the full range of the construct theorized to drive classification among individual cases in a population (Lubke & Luningham, 2017). If we have “good” items that separate classes well, then we do not need as many items (Collins & Wugalter, 1992). Again, however, we rarely have *a priori* knowledge of which items will differentiate classes well. Although there are ways to identify items that have poor measurement qualities (for a discussion on class homogeneity and separation, see Masyn, 2013), these approaches are based on a specified model and may not generalize outside the parameters of that model. As such, the problem of item selection in LCA should be approached substantively where choices are thoughtfully grounded in theory and prior research to the extent possible.

#### **6. Which statistical analysis software can you use to estimate LCA models?**

There are several statistical software packages that can be used to fit LCA models. Some of the more common software packages include *Mplus* (L. K. Muthén & Muthén, 1998-2017), Latent Gold (Vermunt & Magidson, 2013), *MplusAutomation* in R (Hallquist & Wiley, 2018), LAVAAN package in R (Rosseel, 2012), GLLAMM for Stata (Rabe-Hesketh, Skrondal, & Pickles, 2004), and PROC LCA in SAS (Lanza, Collins, Lemmon, & Schafer, 2007), among others. Each software package has different strengths. Whereas some are lower in cost or free (e.g., *MplusAutomation* and LAVAAN), they may not provide as much direct user support services or may require that users are proficient in a specific coding language. Some are very flexible yet may not have an intuitive programming language. Latent Gold provides an array of graphic outputs and an intuitive language. *Mplus* also has an intuitive language and allows researchers to specify the latent class variable in a larger latent variable framework (e.g., in the

context of a mediation model, see for example McLarnon & O'Neill, 2018) and offers an active online support website and one-on-one customer support. For SAS or Stata users, PROC LCA and GLLAAMM may be the preferred computing environments. MplusAutomation in R (Hallquist & Wiley, 2018) provides useful ways to compile many syntax files, tabulate results, and create useful plots for interpreting results, and user support can be found online and through the package documentation. For simple LCA models, there will not be major differences across software packages; this choice is more relevant when researchers are estimating more complex mixture models, using different estimation techniques, or seeking specific graphical output.

### **7. Which estimation technique is commonly used? What are random starts?**

In most software packages used to estimate mixture models, the LCA model parameters are estimated by maximum likelihood (ML) using the expectation-maximization (EM) procedure. ML estimates have many features that make them desirable and thus are the preferred estimation method. Related to the point estimates is the issue of confidence in a solution. For ML estimation, item-level missingness is easily accommodated and assumed missing at random (MAR). That is, individual cases are not excluded from the analysis unless they are missing data on all the observed items.

For mixture models, there is a known sensitivity of the likelihood function to converge on a local, instead of a global, solution (McLachlan & Peel, 2000). To circumvent this problem, the suggestion is to use a set of random start values, estimate the model many times, and see if across the set of random start values, there is convergence on a similar solution (Berlin, Williams, & Parra, 2014; Masyn, 2013). This process is automated in some software packages. In *Mplus*, for example, researchers can specify how many random starts to consider.

Bayesian estimation methods are gaining popularity in the social sciences (Kaplan, 2014) and can be used as an alternative to ML to estimate LCA models. This estimation technique is easily implemented in *Mplus* (as well as other software packages) and can be useful when modeling assumption are not met (e.g., conditional independence). For example, by using non-informative priors, researchers can specify small within-class item correlations that better approximate the within-class dependency without significantly reshaping the emergent classes (Asparouhov & Muthén, 2011). In other mixture modeling contexts, simulation studies have shown that if accurate and informative priors are available, including them in the model estimation process can reduce parameter bias and enhance mixture class recovery (Depaoli, 2012, 2013; Depaoli, Yang, & Felt, 2017). See Li, Lord-Bessen, Shiyko, and Loeb (2018) for an introductory tutorial on how to conduct Bayesian LCA, which also has an applied example that illustrates the method. In addition, van de Schoot and colleagues (2014) provide a tutorial on how to specify some basic Bayesian models and describe some of the modeling options. Also, they help guide the reader in how to interpret modeling results, since these are interpreted quite differently in the Bayesian context (e.g., they do not report  $p$ -values, among many other things).

### **PYDI Analysis**

In our PYDI example, we estimated the set of LCA models in *Mplus* 8.0 (L. K. Muthén & Muthén, 1998-2017) using full information maximum likelihood (FIML) with robust standard errors. We used a large number of random starts (over 500 initial random starts) to establish global maxima and avoid local solutions (McLachlan & Peel, 2000). Examination of the random start output indicated that the log-likelihood value was replicated 82 times out of the 500 requested, which is more than sufficient (e.g., some recommend that 3-10% is needed) to have confidence that the solution we selected was a global, and not local, solution.

### **8. How is LCA different from latent profile analysis?**

LCA is conceptually akin to latent profile analysis (LPA) as both approaches aim to estimate an underlying categorical latent variable. Although the modeling process for both LCA and LPA share similarities, there are some important differences as well (Masyn, 2013). LPA uses observed indicators that are continuous in nature, whereas in LCA, the indicators are categorical and most often binary. As a result, instead of estimating conditional item probabilities across classes, LPA estimates conditional means and variances of the continuous indicators. In LPA, the conditional independence assumption implies that the indicators are uncorrelated given latent profile membership, although this assumption can be relaxed and items may be allowed to freely correlate within profiles. The variance terms in LPA—both the within class variance and covariance of the observed items—add a layer of complexity because they can influence the nature of the heterogeneity being modeled. In fact, freely estimated variances and covariances (either class-varying or invariant) become a modeled source of heterogeneity that contributes to the formation of the classes. There have been controversial discussions in the literature about the sensitivity of the within-class variance assumptions and class separation (Bauer & Curran, 2003; B. O. Muthén, 2003), where its implications for the class enumeration process in growth mixture models (Tofighi & Enders, 2008) and LPA models (Peugh & Fan, 2013) have been subject to investigation. Although there are recommendations for which models to consider, the consequences of misspecifying the within-class variance and covariance structure in LPA and growth mixture models are not well understood and are an active area of research.

### **9. What is the difference between LCA and factor analysis?**

One intuitive way to compare LCA with factor analysis is by understanding that the former groups similar individuals (people, clinics, schools, etc.) whereas the latter groups similar items. More technically, factor analysis defines groups of variables that are similar to each other by estimating a latent factor that represents the shared covariation among the chosen set of indicators (e.g., communality) and is thus termed the “common factor” model. Rather than homogeneity, LCA groups people based on heterogeneity where individuals are differentiated according to systematically diverging patterns of responses to a set of observed indicators. Whereas there are clear differences in these models, hybrid models can be specified that include both a factor and a latent class model, such as in a factor mixture model (Lubke & Muthén, 2005). These models can accommodate “level” differences (e.g., severity) in the factor while estimating “shape” differences, which involve heterogeneity (e.g., typology) in the item responses.

#### **10. Can you use LCA with longitudinal data?**

LCA typically uses cross-sectional data in which responses to multiple items (or measures) are collected at a single time point. However, there is no statistical necessity dictating the use of cross-sectional data. Longitudinal LCA (LLCA), sometimes called repeated measures LCA (RM-LCA; Collins & Lanza, 2010), can be used with longitudinal measures (much like growth modeling) where a latent class variable is specified to capture heterogeneity in the repeated measures over time. In these applications, the latent class variable can be used to describe change over time without having to make any assumptions about the structure or functional form of the change process whereas other longitudinal models do so, such as growth models. LLCA models can thus be specified before a growth model or growth mixture model as a baseline model to explore heterogeneity in change. An example of an applied LLCA model of



change in alcohol use across racial groups revealed heterogeneous patterns of alcohol use over time, both within (e.g., different classes) and across race (Chung, Pedersen, Kim, Hipwell, & Stepp, 2014).

### **Conclusion**

In this article, we described LCA and its versatility as an analytic approach to parameterizing unobserved within-group differences in a population with respect to a substantive domain of interest. LCA is increasingly popular among psychological researchers because they—having inherited a historical and professional tradition of seeking to better understand individual and within-group differences—regularly encounter research scenarios that call for identifying how people are systematically different in ways we cannot directly observe. By answering 10 frequently asked questions about LCA, we presented a pedagogical, introductory overview of common topics related to this approach, illustrated core modeling ideas with an applied example using the PYDI, and facilitated access to references that provide a more in-depth treatment of the issues discussed. In addition, we provided basic *Mplus* code as an introduction to the syntax of a simple LCA model. An informed and judicious application of LCA can offer psychological researchers the potential to attain a more nuanced and refined knowledge of psychosocial phenomena and, by extension, to inform interventions and policy that are more sensitive and responsive to the complex variation inherent those phenomena.

The LCA model provides a foundation for appreciating and understanding the wide-ranging possibilities afforded by the mixture modeling framework. There is a vast array of modeling extensions from the LCA model that allow researchers to identify heterogeneity in complex and dynamic psychological phenomena, and we briefly mention a select few here.

Although LCA is generally used as an exploratory technique, it can be applied in a confirmatory

manner, for instance to test for the equivalence of latent class solutions across known groupings akin to measurement invariance in factor analysis (Finch & Bronk, 2011; Laudy et al., 2005; Morin, Meyer, Creusier, & Biétry, 2016). Also, it may be of interest to model individuals' joint membership across multiple latent class variables and for these scenarios using a joint LCA or latent "co-occurrence" model can be appropriate (Jeon, Lee, Anthony, & Chung, 2017). For nested data (or data that violate the assumption of independent observations), multilevel LCA can be used where the probability of class membership is allowed to vary across units of nestedness, for example classrooms (Nylund-Gibson, Graham, & Juvonen, 2010) or communities (Henry & Muthén, 2010). Latent transition analysis (LTA) is used to examine how individuals transition in their class membership over time with respect to a latent class variable measured across multiple time points (Collins & Wugalter, 1992; Nylund, 2007). Growth mixture models (GMM), as alluded to previously, are used to model the heterogeneity in growth trajectories (e.g., often termed latent trajectory classes) over time (Petras & Masyn, 2010; Ram & Grimm, 2009). In addition, latent class variables can be embedded into larger latent variable models, such as a mediation model (McLarnon & O'Neill, 2018). All of these advanced mixture models in essence have at their foundation a latent class variable.

In addition to the wide range of mixture modeling applications, there are continuing advances in our knowledge of how to specify these models, including the specification of auxiliary variables (e.g., covariates and distal outcomes). In this article, we demonstrated the three-step method, but more novel approaches that preserve the latent class measurement model and allow for efficient and unbiased estimation of auxiliary relations are currently being developed and tested. For example, recent work on differential item functioning (DIF) in the LCA context (e.g., the direct association of a covariate and a latent class indicator) has presented

modeling strategies for permitting DIF and including it in subsequent three-step specifications (Masyn, 2017). Thus, we encourage researchers to keep abreast of the exciting changes and developments in best practices that are expected in the coming years.

LCA and its extensions are likely to expand in their central place in the larger landscape of statistical methods for psychological research in the near future. Methodological work is well underway and continues to improve our understanding of how mixture models perform in practice and their plausibility in applied contexts. We hope that this article has illustrated the flexibility of the mixture modeling framework, demonstrated the ability of this class of models to help answer questions about unobserved heterogeneity in diverse phenomena, and highlighted the substantial utility that LCA can bring to psychological research.

### References

- Arnold, M. E., Nott, B. D., & Meinhold, J. L. (2012). *The Positive Youth Development Inventory Full Version*. Corvallis, OR: Oregon State University.
- Asparouhov, T., & Muthén, B. O. (2011). Using Bayesian priors for more flexible latent class analysis. In *Joint Statistical Proceedings, Government Statistics Section* (pp. 4979–4993). Alexandria, VA: American Statistical Association.
- Asparouhov, T., & Muthén, B. O. (2012). Using *Mplus* TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes 14*. Retrieved from [statmodel.com/examples/webnotes/webnote14.pdf](http://statmodel.com/examples/webnotes/webnote14.pdf)
- Asparouhov, T., & Muthén, B. O. (2014). Auxiliary variables in mixture modeling: Three-step approaches using *Mplus*. *Structural Equation Modeling, 21*(3), 329–341.  
<https://doi.org/10.1080/10705511.2014.915181>
- Asparouhov, T., & Muthén, B. (2015). Residual associations in latent class and latent transition analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(2), 169-177.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338–363.
- Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An Introduction to latent variable mixture modeling (Part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology, 39*(2), 174–187. <https://doi.org/10.1093/jpepsy/jst084>
- Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3–27. <https://doi.org/10.2307/25791751>

- Chung, T., Pedersen, S. L., Kim, K. H., Hipwell, A. E., & Stepp, S. D. (2014). Racial differences in type of alcoholic beverage consumed during adolescence in the Pittsburgh Girls Study. *Alcoholism: Clinical and Experimental Research, 38*(1), 285–293.  
<https://doi.org/10.1111/acer.12222>
- Clark, S. L., Muthén, B., Kaprio, J., D’Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling, 20*(4), 681–703.  
<https://doi.org/10.1080/10705511.2013.824786>
- Clark, S. L., & Muthén, B. O. (2009). *Relating latent class analysis results to variables not included in the analysis* (unpublished manuscript).
- Cloitre, M., Garvert, D. W., Weiss, B., Carlson, E. B., & Bryant, R. A. (2014). Distinguishing PTSD, complex PTSD, and Borderline Personality Disorder: A latent class analysis. *European Journal of Psychotraumatology, 5*(1), 25097.  
<https://doi.org/10.3402/ejpt.v5.25097>
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons, Inc.
- Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*(1), 131–157.  
[https://doi.org/10.1207/s15327906mbr2701\\_8](https://doi.org/10.1207/s15327906mbr2701_8)
- Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. *Structural Equation Modeling, 19*(2), 178–203.  
<https://doi.org/10.1080/10705511.2012.659614>

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation:

Frequentist versus Bayesian estimation. *Psychological Methods, 18*(2), 186–219.

<http://dx.doi.org/10.1037/a0031609>

Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian statistics to model uncertainty in

mixture models: A sensitivity analysis of priors. *Structural Equation Modeling, 24*(2),

198–215. <https://doi.org/10.1080/10705511.2016.1250640>

Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using

*Mplus*. *Structural Equation Modeling, 18*(1), 132–151.

Grimm, R., Nylund-Gibson, K., & Masyn, K. E. (April, 2018). *Navigating the ever-changing*

*landscape of distals in mixture models: A road map of current approaches*. Paper

presented at the 2018 Annual Meeting of the American Educational Research Association

(AERA). New York City, NY.

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-

scale latent variable analyses in *Mplus*. *Structural Equation Modeling, 25*(4), 621–638.

<https://doi.org/10.1080/10705511.2017.1402334>

Henry, K. L., & Muthén, B. O. (2010). Multilevel latent class analysis: An application of

adolescent smoking typologies with individual and contextual predictors. *Structural*

*Equation Modeling, 17*(2), 193–215. <https://doi.org/10.1080/10705511003659342>

Jang, Y., Park, N. S., Chiriboga, D. A., & Kim, M. T. (2017). Latent profiles of acculturation and

their implications for health: A study with Asian Americans in central Texas. *Asian*

*American Journal of Psychology*, advance online publication.

<https://doi.org/10.1037/aap0000080>

- Jeon, S., Lee, J., Anthony, J. C., & Chung, H. (2017). Latent class analysis for multiple discrete latent variables: A study on the association between violent behavior and drug-using behaviors. *Structural Equation Modeling, 0*(0), 1–15.  
<https://doi.org/10.1080/10705511.2017.1340844>
- 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*(4), 671–694.  
<https://doi.org/10.1080/10705510701575602>
- Laudy, O., Zoccolillo, M., Baillargeon, R. H., Boom, J., Tremblay, R. E., & Hoijsink, H. (2005). Applications of confirmatory latent class analysis in developmental psychology. *European Journal of Developmental Psychology, 2*(1), 1–15.  
<https://doi.org/10.1080/17405620444000193>
- Lerner, R. M., Lerner, J. V., Almerigi, J. B., Theokas, C., Phelps, E., Gestsdottir, S., ... von Eye, A. (2005). Positive youth development, participation in community youth development programs, and community contributions of fifth-grade adolescents: Findings from the first wave of the 4-H Study of Positive Youth Development. *The Journal of Early Adolescence, 25*(1), 17–71. <https://doi.org/10.1177/0272431604272461>
- Li, Y., Lord-Bessen, J., Shiyko, M., & Loeb, R. (2018). Bayesian latent class analysis tutorial. *Multivariate Behavioral Research, 53*(3), 430–451.  
<https://doi.org/10.1080/00273171.2018.1428892>
- Lubke, G. H., & Luningham, J. (2017). Fitting latent variable mixture models. *Behaviour Research and Therapy, 98*, 91–102. <https://doi.org/10.1016/j.brat.2017.04.003>

- Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21–39. <http://dx.doi.org/10.1037/1082-989X.10.1.21>
- Lubke, G. H., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14*(1), 26–47. <https://doi.org/10.1080/10705510709336735>
- Lubke, G. H., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41*(4), 499–532.
- Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2: Statistical analysis, pp. 551–611). New York, NY: Oxford University Press.
- Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling, 24*(2), 180–197. <https://doi.org/10.1080/10705511.2016.1254049>
- McLachlan, G., & Peel, D. (2000). *Finite mixture modeling*. New York, NY: Wiley.
- McLarnon, M. J. W., & O'Neill, T. A. (2018). Extensions of auxiliary variable approaches for the investigation of mediation, moderation, and conditional effects in mixture models. *Organizational Research Methods*, advance online publication. <https://doi.org/10.1177/1094428118770731>
- Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling, 22*(1), 76–86. <https://doi.org/10.1080/10705511.2014.935751>



Morin, A. J. S., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods, 19*(2), 231–254.

<https://doi.org/10.1177/1094428115621148>

Morovati, D. (2014). *The intersection of sample size, number of indicators, and class enumeration in LCA: A Monte Carlo study* (unpublished dissertation). Santa Barbara, CA: University of California, Santa Barbara.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 369–377.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463–469.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo simulation study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620.

[https://doi.org/10.1207/S15328007SEM0904\\_8](https://doi.org/10.1207/S15328007SEM0904_8)

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (unpublished doctoral dissertation). Los Angeles, CA: University of California, Los Angeles.

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*, 14(4), 535–569.

<https://doi.org/10.1080/10705510701575396>

Nylund, K. L., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, 78(6), 1706–1722.

Nylund-Gibson, K., Graham, S., & Juvonen, J. (2010). An application of multilevel LCA to study peer victimization in middle school. *Advances and Applications in Statistical Sciences*, 3(2), 343–363.

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling*, 21(3), 439–454.

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*, 23(6), 782–797.

<https://doi.org/10.1080/10705511.2016.1221313>

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). Springer, New York, NY.

[https://doi.org/10.1007/978-0-387-77650-7\\_5](https://doi.org/10.1007/978-0-387-77650-7_5)

Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling*, 20(4), 616–639.

<https://doi.org/10.1080/10705511.2013.824780>

- Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with Grade 2 achievement. *Early Childhood Research Quarterly, 28*(2), 437–449.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM manual. In *U.C. Berkeley Division of Biostatistics Working Paper Series* (working paper 160). Berkeley, CA: Berkeley Electronic Press.
- Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565–576. <https://doi.org/10.1177/0165025409343765>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), online publication. <https://doi.org/10.18637/jss.v048.i02>
- Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling, 20*(4), 640–657.
- Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age Publishing.
- Tueller, S., & Lubke, G. H. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling, 17*(2), 165–192. <https://doi.org/10.1080/10705511003659318>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*(3), 842–860. <https://doi.org/10.1111/cdev.12169>

- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*(4), 450–469. <https://doi.org/10.2307/25792024>
- Vermunt, J. K., & Magidson, J. (2013). *LG-syntax user's guide: Manual for Latent GOLD 5.0 syntax module*. Belmont, MA: Statistical Innovations, Inc.
- Wagenmakers, E. (2007). A practical solution to the pervasive problems of p values. *Psychometric Bulletin & Review, 14*(5), 779-804.
- Wasserman, L. (August, 1997). *Bayesian model selection and model averaging*. Paper presented at Mathematical Psychology Symposium on “Methods for Model Selection,” Bloomington, IN.
- Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology, 5*, article 920. <https://doi.org/10.3389/fpsyg.2014.00920>
- Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*(4), 1090–1104. <https://doi.org/10.1016/j.csda.2004.11.004>

Table 1

*Guidelines for Commonly Used Fit Indices for Latent Class Analysis*

Fit Index	Reference	Guideline	Computation	Where it is Obtained
<b>Information Criterion</b>				
Bayesian Information Criterion (BIC) <sup>1</sup>	Schwarz (1978)	<ul style="list-style-type: none"> <li>Select the model with the lowest BIC value. Or, pick the model where there is a diminishing decrement in value for each added class (e.g., like a scree plot).</li> </ul>	$BIC = -2(LL) + d\log(n)$	Provided by <i>Mplus</i> in output
Sample-size Adjusted Bayesian Information Criterion (SABIC)	Sclove (1987)	<ul style="list-style-type: none"> <li>Same guideline as BIC</li> </ul>	$SABIC = -2(LL) + d\log(n + 2/24)$	Provided by <i>Mplus</i> in output
Consistent Akaike Information Criterion (CAIC)	Bozdogan (1987)	<ul style="list-style-type: none"> <li>Same guideline as BIC</li> </ul>	$CAIC = -2(LL) + d[\log(n) + 1]$	Computed by researcher using LL value, number of parameters ( <i>d</i> ), and sample size, all found in the output.
Approximate Weight of Evidence (AWE)	Banfield & Raftery (1993)	<ul style="list-style-type: none"> <li>Same guideline as BIC</li> </ul>	$AWE = -2(LL) + d[\log(n) + 1.5]$	Computed by researcher using LL value, number of parameters ( <i>d</i> ), and sample size, all found in the output.
<b>Relative Fit Indices</b>				
Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT)	Lo, Mendell, & Rubin (2001)	<ul style="list-style-type: none"> <li>Compares a specified <i>K</i>-class model to <i>K</i>-1 class model and provides a <i>p</i>-value indicating if the additional class (<i>K</i>-class model) significantly improves model fit compared to the <i>k</i>-1 class model.</li> <li>If <i>p</i>-value for a <i>K</i> class model is non-significant, that indicates that the <i>K</i>-1 class model is supported by the <i>VLMR-LRT</i>. (e.g., a non-significant <i>p</i>-value for a 5-class model suggests the 4-class model is supported by the <i>VLMR-LRT</i>)</li> </ul>		Computed by <i>Mplus</i> by requesting Tech11
Bootstrap Likelihood Ratio Test (BLRT)	McLachlan & Peel (2000)	<ul style="list-style-type: none"> <li>Same guideline as VLMR-LRT</li> </ul>		Computed by <i>Mplus</i> by requesting Tech14; user may need to specify more random starts for BLRT

				by specifying K-1STARTS (see Asparouhov & Muthén, 2012 for more).
Bayes Factor ( <i>BF</i> )	Wagenmakers (2007); Wasserman (1997)	<ul style="list-style-type: none"> <li>• Compares fit between two models: <i>K</i> and <i>K</i>+1.</li> <li>• A <i>BF</i> less than 3 is considered weak evidence for Model <i>K</i> over Model <i>K</i>+1, <math>3 &lt; BF &lt; 10</math> is moderate evidence for Model <i>K</i>, and <math>BF &gt; 10</math> is strong evidence for Model <i>K</i>.</li> </ul>	The <i>BF</i> comparing Model A (model <i>K</i> ) and Model B (model <i>K</i> +1) is $BF_{A,B} = \exp[SIC_A - SIC_B]$ where <i>SIC</i> , the Schwartz Information Criterion, is defined as $SIC = -.05(BIC)$ ; (e.g., <i>BF</i> of 5 provides moderate evidence for a 3-class model compared to the 4-class model).	Computed by researcher using BIC value provided by <i>Mplus</i>
Approximate Correct Model Probability ( <i>cmP</i> )	Schwarz (1978)	<ul style="list-style-type: none"> <li>• Estimates the probability that each model out of a given set of fitted LCA models is correct, assuming the “true” model is in the set. The <i>cmP</i> values across the given set of models sum to 1. The model with the largest <i>cmP</i> value is the model that is picked.</li> </ul>	We calculate a <i>cmP</i> for each model in a set of <i>J</i> models specified by the user. We compute it as follows, $cmP_A = \frac{\exp[SIC_A - SIC_{max}]}{\sum_{j=1}^J \exp[SIC_A - SIC_{max}]}$ where $SIC_{max}$ is the maximum <i>SIC</i> score of the <i>J</i> models that were considered in the given application.	Computed by researcher using BIC value provided by <i>Mplus</i>

- *Note.* <sup>1</sup>The BIC is the most commonly used and trusted fit index for model comparison (Masyn, 2013; Nylund, Asparouhov, et al., 2007). LL = log likelihood; *d* = number of parameters; *n* = sample size.

Table 2

*Indicator Labels: Positive Youth Development Inventory—Contribution  
Subscale Latent Class Analysis Model*

Item	Description
1:	I take an active role in my community.
2:	I am someone who gives to benefit others.
3:	I like to work with others to solve problems.
4:	I have things I can offer to others.
5:	I believe I can make a difference in the world.
6:	I care about contributing to make the world a better place for everyone.
7:	It is important for me to try and make a difference in the world.

Table 3

*Fit Statistics and Classification Coefficients: Positive Youth Development Inventory—Contribution Subscale Latent Class Analysis Models*

K	LL	BIC	SABIC	CAIC	AWE	BLRT <i>p</i>	VLMR-LRT <i>p</i>	Entropy	<i>BF</i>	<i>cmP</i>
1	-3905.892	7863.555	7841.317	7841.267	7877.751	-	-	-	.000	.000
2	-3439.483	6989.902	6942.250	6942.145	7020.324	<.001	<.001	.790	.000	.000
3	-3394.950	<b>6960.001</b>	<b>6886.934</b>	<b>6886.774</b>	<b>7006.648</b>	<.001	<.001	.718	<b>&gt;15.000</b>	<b>1.000</b>
4	-3379.436	6988.140	6889.658	6889.442	7051.011	<b>&lt;.001</b>	<b>.002</b>	.753	<b>&gt;15.000</b>	.000
5	-3370.534	7029.501	6905.604	6905.333	7108.598	.200	.429	.782	<b>&gt;15.000</b>	.000
6	<b>-3364.745</b>	7077.089	6927.778	6927.450	7172.411	.500	.472	<b>.803</b>	-	.000

*Note.* K = number of classes; LL = log-likelihood; BIC = Bayesian Information Criterion; SABIC = Sample-size adjusted BIC; CAIC = Consistent Akaike Information Criterion; AWE = Approximate Weight of Evidence Criterion; BLRT = bootstrapped likelihood ratio test; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; *p* = *p*-value; *BF* = Bayes Factor; *cmP* = correct model probability. **Bolded** values indicate “best” fit for each respective statistic. Entropy is included in the table for brevity but should not be used as a model selection statistic (Masyn, 2013).



Table 4

*Classification Probabilities: Positive Youth Development Inventory—Contribution Subscale 4-class Model*

Class	1	2	3	4
1: Holistic-Collaborative	<b>.969</b>	.003	.000	.028
2: Altruistic-Low Self-Efficacy	.069	<b>.910</b>	.021	.000
3: Low Engagement	.007	.167	<b>.752</b>	.074
4: Holistic-Independent	.439	.000	.003	<b>.558</b>

*Note.* Values indicate probabilities of most likely class membership (column) by latent class modal assignment (row). **Bolded** values indicate average posterior probabilities (AvePP).

Table 5

*Covariate Logits and Odds Ratios for Race (Hispanic): Positive Youth Development—Contribution Subscale 4-class Model*

Class membership	Reference class							
	1		2		3		4	
	Logit	OR	Logit	OR	Logit	OR	Logit	OR
1: Holistic-Collaborative	-	-	.998**	2.713	.589	1.802	.485*	1.624
2: Altruistic Low Self-efficacy	-.998**	.369	-	-	-.408	.665	-.513	.599
3: Low Engagement	-.589	.555	.408	1.504	-	-	-.104	.901
4: Holistic-Independent	-.485*	.616	.513	1.670	.104	1.110	-	-

*Note.* \* $p < .05$ . \*\*  $p \leq .001$ .

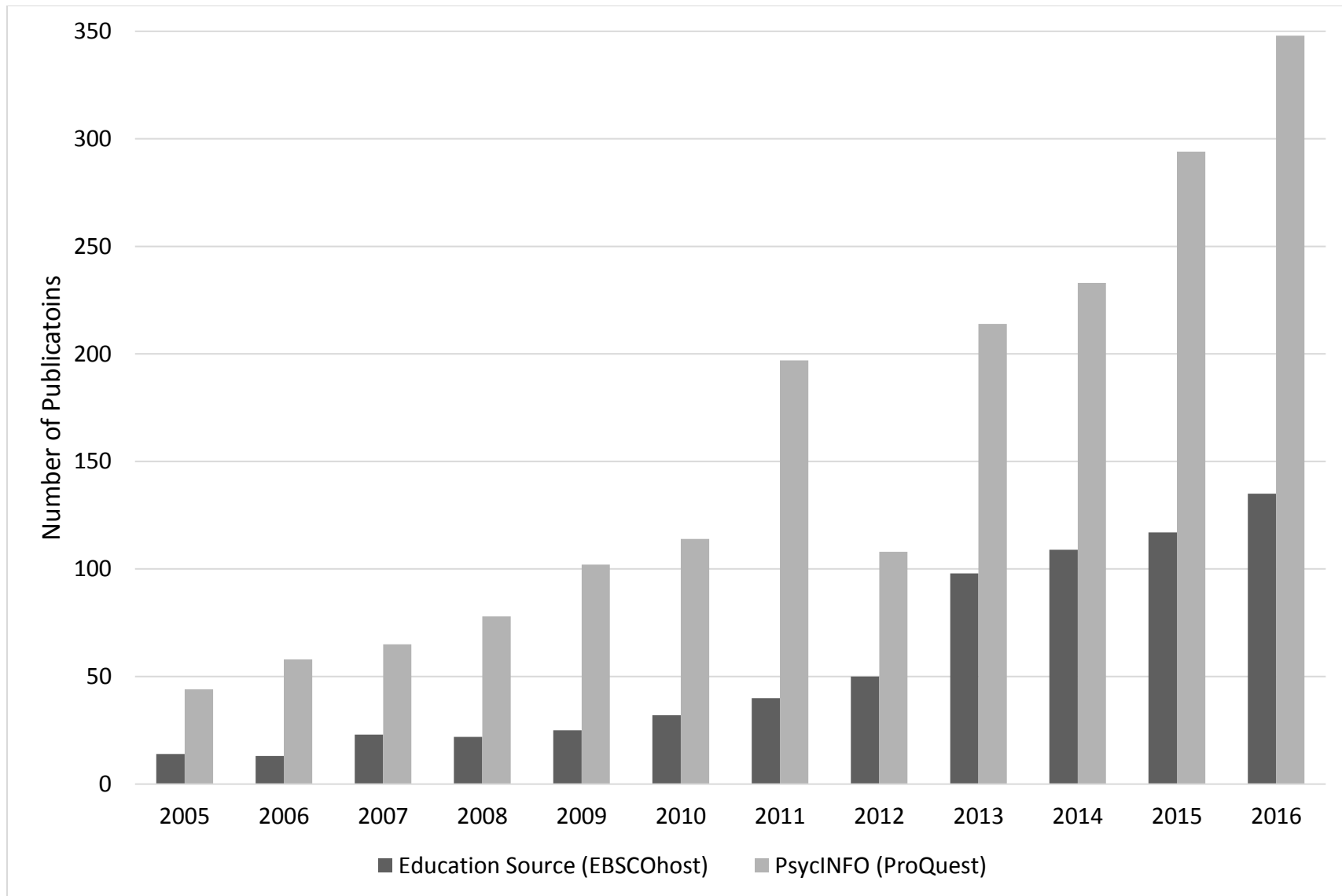


Figure 1. Number of Publications in Education and Psychology Journals between 2005 and 2016 that use Latent Class Analysis.

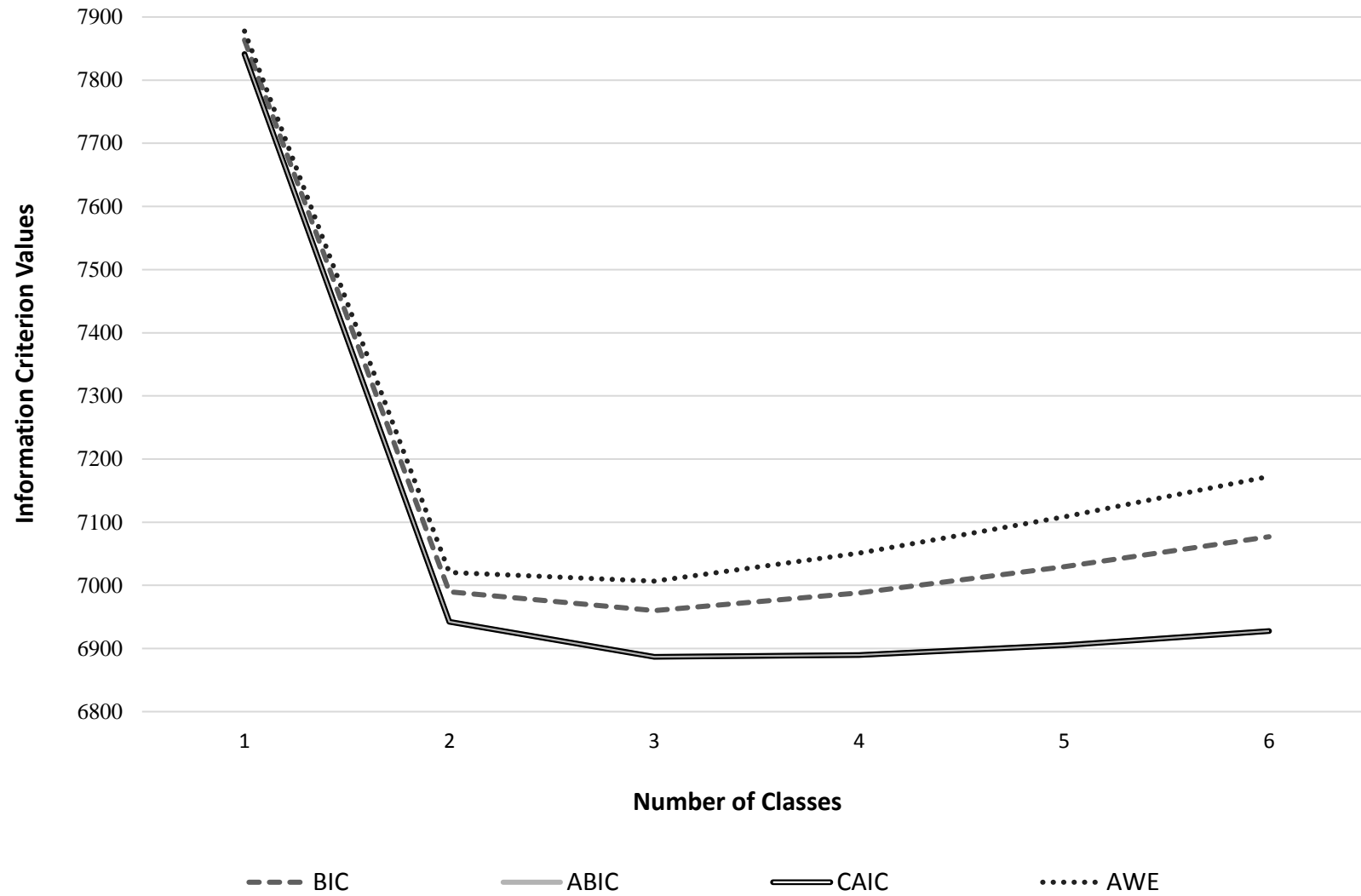


Figure 2. Plot of Information Criterion values: Positive Youth Development—Contribution Subscale Latent Class Analysis Models.  
 Note. The ABIC and CAIC lines are overlapping.

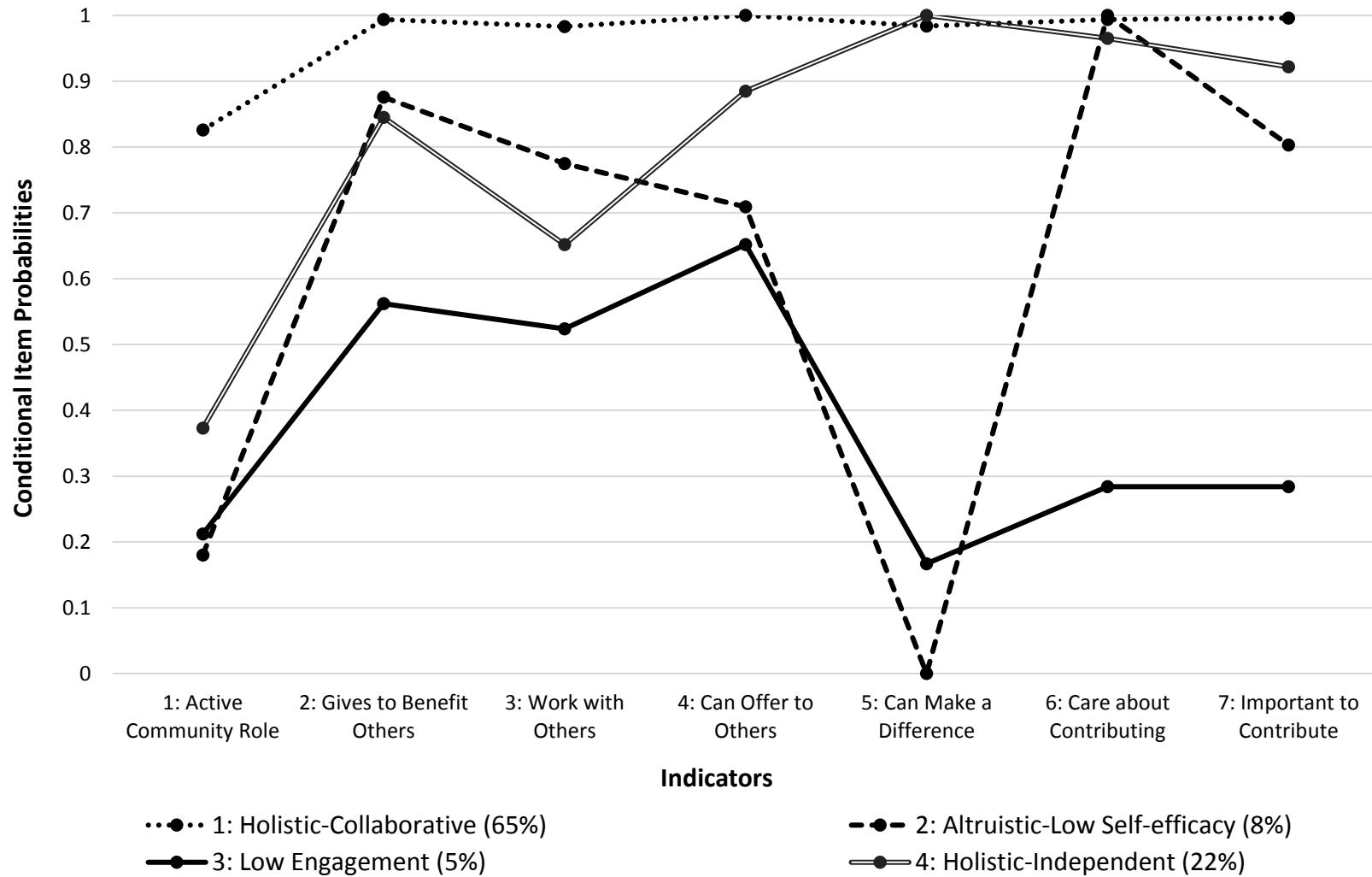


Figure 3. Conditional Item Probability Plot: Positive Youth Development Inventory—Contribution Subscale 4-class Model. Note. See Table 2 for full indicator labels. Class prevalence are in parentheses.

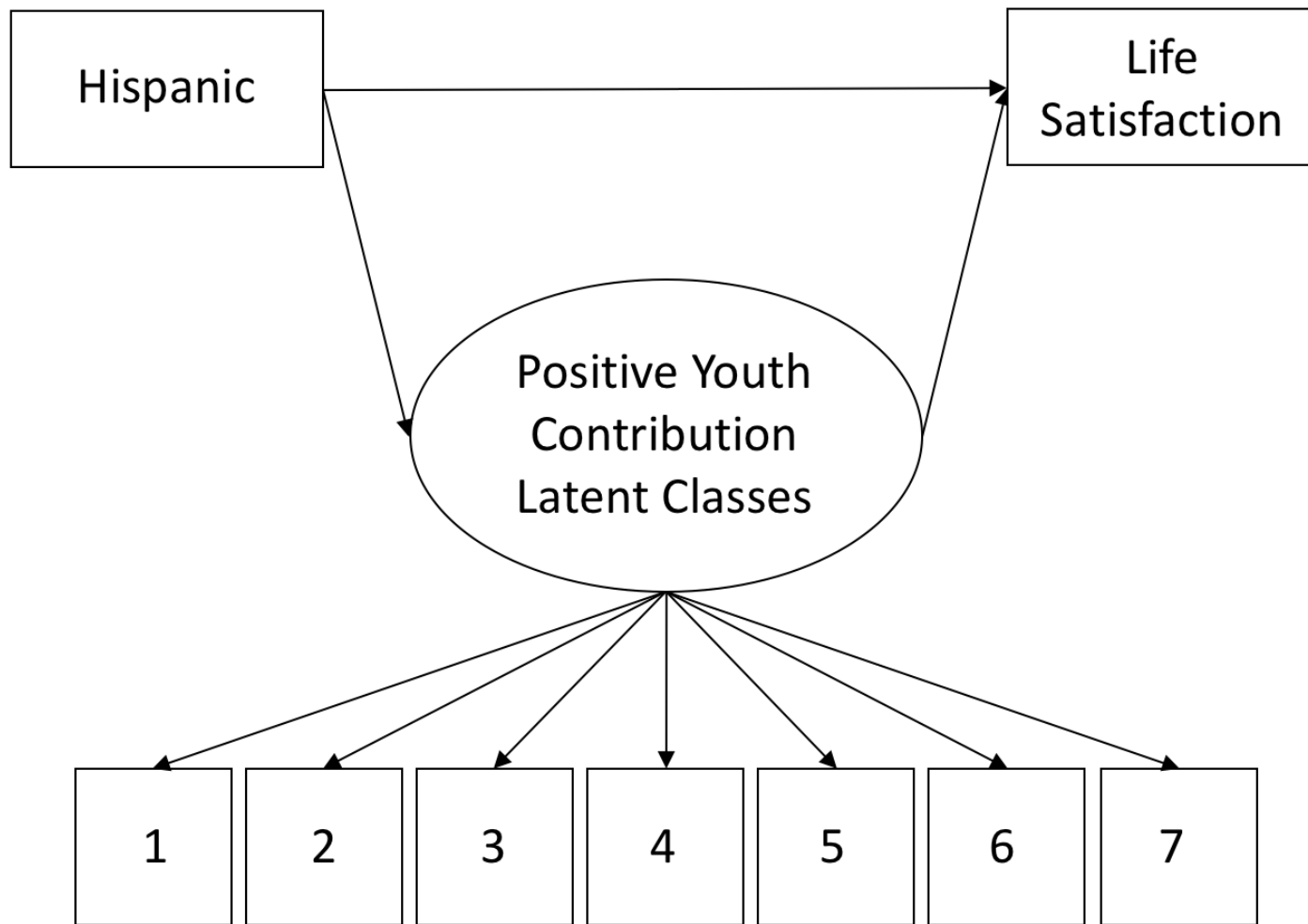


Figure 4. Path Diagram of the Positive Youth Development—Contribution Subscale Latent Class Analysis Model with Covariate and Distal Outcome. *Note.* See Table 2 for indicator labels. Please note that the diagrammed relation between the latent class variable and the distal outcome represents class-specific conditional means rather than a *bona fide* regressive path.

Appendix 1: *Mplus Syntax for Step 1 – Unconditional LCA*


---

**TITLE:** Positive Youth Development Inventory LCA (enumeration);  
**DATA:** FILE IS PYDI.dat;  
**VARIABLE:**  
 NAMES ARE hispanic LifeSat PYDI1 PYDI2 PYDI3 PYDI4 PYDI5 PYDI6 PYDI7;  
 MISSING ARE ALL (9999);  
 USEVARIABLES = PYDI1 PYDI2 PYDI3 PYDI4 PYDI5 PYDI6 PYDI7;  
 CATEGORICAL = PYDI1 PYDI2 PYDI3 PYDI4 PYDI5 PYDI6 PYDI7;  
 AUXILIARY = hispanic LifeSat;  
 CLASSES = C1(4);  
**ANALYSIS:**  
 TYPE = mixture;  
 STARTS = 500 100;  
**OUTPUT:**  
 tech11 tech14;  
**PLOT:**  
 TYPE = plot3;  
 SERIES = PYDI1-PYDI7(\*);  
**SAVEDATA:**  
 FILE = PYDI3step.dat;  
 SAVE = cprob;  
 MISSFLAG = 9999;

---

*Note.* The default estimator for mixture models in *Mplus* is maximum likelihood with robust standard errors (MLR). In MISSING ARE ALL, specify how missing data is coded in the dataset. For binary indicators, the CATEGORICAL command must be included to specify which items are categorical. In AUXILIARY, include all covariates and distal outcomes to be used in step 3. In CLASSES, we name the latent class variable, C1, and specify the determined number of classes (e.g., in this step we have already decided that the 4-class solution is our final unconditional model). To specify a different number of classes, replace 4 in C1(4) with the desired number of classes. In STARTS, 500 indicates the number of random starts and 100 indicates the number of replications that will go through full optimization. In OUTPUT, tech11 provides the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR-LRT) and tech14 provides the bootstrapped likelihood ratio test (BLRT). In PLOT, plot3 provides all possible plots for a specified model, including item probability plots when estimating mixture models. In SERIES, the indicators can be reordered for interpreting the item probability plot. In SAVE, cprob indicates that in PYDI3step.dat, the logits of the classification probabilities for the 4-class model will be saved.

---

The (\*) in the SERIES command indicates to *Mplus* to use the default spacing of the indicators along the x-axis which is unit separation.



Appendix 2: *Mplus Syntax for Step 3 – Estimation of Auxiliary Variable Relations*


---

**TITLE:** Positive Youth Development Inventory LCA (auxiliary);

**DATA:** FILE IS PYDI3step.dat;

**VARIABLE:**

NAMES ARE hispanic LifeSat

PYDI1 PYDI2 PYDI3 PYDI4 PYDI5 PYDI6 PYDI7

CPROB1 CPROB2 CPROB3 CPROB4 C;

MISSING ARE ALL (9999);

USEVARIABLES = C hispanic LifeSat;

NOMINAL = C;

CLASSES = C1(4);

**DEFINE:**

CENTER hispanic (GRANDMEAN);

**ANALYSIS:**

TYPE = mixture;

STARTS = 0;

**MODEL:**

%OVERALL%

C1 LifeSat ON hispanic;

[LifeSat];

LifeSat;

%C1#1%

[C#1@3.547];

[C#2@-2.222];

[C#3@-4.641];

[LifeSat] (d1);

LifeSat;

%C1#2%

[C#1@11.149];

[C#2@13.721];

---

---

```
[C#3@9.942];
[LifeSat] (d2);
LifeSat;
```

```
%C1#3%
[C#1@-2.419];
[C#2@0.818];
[C#3@2.321];
[LifeSat] (d3);
LifeSat;
```

...continued on next page ...

```
%C1#4%
[C#1@-0.241];
[C#2@-13.232];
[C#3@-5.155];
[LifeSat] (d4);
LifeSat;
```

#### MODEL TEST:

```
0 = d1-d2;
0 = d2-d3;
0 = d3-d4;
```

#### MODEL CONSTRAINT: NEW(diff12 diff13 diff14 diff23 diff24 diff34);

```
diff12 = d1-d2;
diff13 = d1-d3;
diff14 = d1-d4;
diff23 = d2-d3;
diff24 = d2-d4;
diff34 = d3-d4;
```

---

*Note.* The default estimator for mixture models in *Mplus* is maximum likelihood with robust standard errors (MLR). In this run, we are using the dataset saved in the previous run (namely, PYDI3step.dat). In MISSING ARE ALL, specify how missing data is coded in the dataset. In DEFINE, the covariate is grand mean-centered to ease interpretation of class-specific distal outcomes. In ANALYSIS, STARTS are fixed to 0 in Step 3 as we are fixing all measurement parameters in this step and thus do not need random starts. In %OVERALL%, specify model commands that apply to all classes, including regressing the latent class variable and the distal outcome onto the covariate (e.g., C1 LifeSat ON hispanic;). We mention the mean and variance of life

---

satisfaction in the overall statement (e.g., [LifeSat]; and LifeSat;, respectively). In the class-specific model commands (e.g., “%C1#1%,” “%C1#2%,” etc.), classes are held constant by fixing the logits of the classification probabilities for the three-step specification (e.g., [C#1@3.547]; for more on this, see Nylund-Gibson, Grimm, Quirk, & Furlong, 2014). Additionally, in the class specific statements, we estimate the conditional distal outcome means (e.g., we estimate a mean of the distal outcome for each of the latent classes). The statement “[LifeSat];” freely estimates the mean of the distal outcome life satisfaction and the command “LifeSat;” freely estimates variances for this distal outcome for each of the latent classes (e.g., class-specific variances of the distal outcome). We labeled the distal outcome means d1, d2, etc. to allow for the pairwise testing of the equality of these means to be specified in a later section of the syntax. In MODEL TEST, by mentioning statements testing the equality of all the distal means, *Mplus* will simultaneously test if all these statements are true. This provides an omnibus test (similar to the omnibus *F*-test in ANOVA) to see there are any differences among the distal outcome means. In MODEL CONSTRAINT, we specify and individually test the significance of all pairwise comparisons of the distal outcome means across classes to see where the significant differences occur.