OXFORD LIBRARY OF PSYCHOLOGY

Editor-in-Chief PETER E. NATHAN

The Oxford Handbook of Quantitative Methods

Edited by

Todd D. Little

VOLUME 2: STATISTICAL ANALYSIS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press in the UK and certain other
countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
The Oxford handbook of quantitative methods / edited by Todd D. Little.
v. cm. – (Oxford library of psychology)
ISBN 978–0–19–993487–4
ISBN 978–0–19–993489–8
1. Psychology–Statistical methods. 2. Psychology–Mathematical models. I. Little, Todd D.
BF39.O927 2012
150.72′1—dc23
2012015005

9 8 7 6 5 4 3 2

Printed in the United States of America
on acid-free paper

CHAPTER

25 Latent Class Analysis and Finite Mixture Modeling

Katherine E. Masyn

Abstract

Finite mixture models, which are a type of latent variable model, express the overall distribution of one or more variables as a *mixture* of a *finite* number of component distributions. In direct applications, one assumes that the overall population heterogeneity with respect to a set of manifest variables results from the existence of two or more distinct homogeneous subgroups, or latent classes, of individuals. This chapter presents the prevailing "best practices" for direct applications of basic finite mixture modeling, specifically latent class analysis (LCA) and latent profile analysis (LPA), in terms of model assumptions, specification, estimation, evaluation, selection, and interpretation. In addition, a brief introduction to structural equation mixture modeling in the form of latent class regression is provided as well as a partial overview of the many more advanced mixture models currently in use. The chapter closes with a cautionary note about the limitations and common misuses of latent class models and a look toward promising future developments in mixture modeling.

Key Words: Finite mixture, latent class, latent profile, latent variable

Introduction

Like many modern statistical techniques, mixture modeling has a rich and varied history—it is known by different names in different fields; it has been implemented using different parameterizations and estimation algorithms in different software packages; and it has been applied and extended in various ways according to the substantive interests and empirical demands of different disciplines as well as the varying curiosities of quantitative methodologists, statisticians, biostatisticians, psychometricians, and econometricians. As such, the label *mixture model* is quite equivocal, subsuming a range of specific models, including, but not limited to: latent class analysis (LCA), latent profile analysis (LPA), latent class cluster analysis, discrete latent trait analysis, factor mixture models, growth mixture models, semi-parametric group-based models, semi-nonparametric group-mixed models, regression mixture models, latent state models, latent structure analysis, and hidden Markov models.

Despite the equivocal label, all of the different mixture models listed above have two common features. First, they are all *finite mixture* models in that they express the overall distribution of one or more variables as a *mixture of* or composite of a *finite* number of component distributions, usually simpler and more tractable in form than the overall distribution. As an example, consider the distribution of adult heights in the general population. Knowing that males are taller, on average, than females, one could choose to express the distribution of heights as a mixture of two component distributions for males and

females, respectively. If $f(height)$ is the probability density function of the distribution of heights in the overall population, it could be expressed as:

$$f(height) = p_{male} \cdot f_{male}(height) + p_{female} \cdot f_{female}(height), \quad (1)$$

where p_{male} and p_{female} are the proportions of males and females in the overall population, respectively, and $f_{male}(height)$ and $f_{female}(height)$ are the distributions of heights within the male and female subpopulations, respectively. p_{male} and p_{female} are referred to as the *mixing proportions* and $f_{male}(height)$ and $f_{female}(height)$ are the *component distribution* density functions.

The second common feature for all the different kinds of mixture models previously listed is that the components themselves are not directly observed—that is, mixture component membership is unobserved or *latent* for some or all individuals in the overall population. So, rather than expressing the overall population distribution as a mixture of *known* groups, as with the height example, mixture models express the overall population distribution as a finite mixture of some number, K, of unknown groups or components. For the distribution of height, this finite mixture would be expressed as:

$$f(height) = p_1 \cdot f_1(height) + p_2 \cdot f_2(height) + \cdots + p_K \cdot f_K(height), \quad (2)$$

where the number of components, K, the mixing proportions, $p_1, \ldots, p_K$, and the component-specific height distributions, $f_1(height), \ldots, f_K(height)$, are all unknown but can be estimated, under certain identifying assumptions, using height data measured on a representative sample from the total population.

Finite Mixture Models As Latent Variable Models

It is the unknown nature of the mixing components—in number, proportion, and form—that situates finite mixture models in the broader category of latent variable models. The finite mixture distribution given in Equation 2 can be re-expressed in terms of a latent unordered categorical variable, usually referred to as a *latent class variable* and denoted by c, as follows:

$$f(height) = \Pr(c = 1) \cdot f(height|c = 1) + \cdots + \Pr(c = K) \cdot f(height|c = K), \quad (3)$$

where the number of mixing components, K, is the number of categories or classes of c ($c = 1, \ldots, K$); the mixing proportions are the class proportions, $\Pr(c = 1), \ldots, \Pr(c = K)$; and the component distribution density functions are the distribution functions of the response variable, conditional on latent class membership, $f(height|c = 1), \ldots, f(height|c = K)$.

Recognizing mixture models as latent variable models allows use of the discourse language of the latent variable modeling world. There are two primary types of variables: (1) *latent* variables (e.g., the latent class variable, c) that are not directly observed or measured, and (2) *manifest* variables (e.g., the response variables) that are observable and are presumed to be influenced by or caused by the latent variable. The manifest variables are also referred to as *indicator* variables, as their observed values for a given individual are imagined to be imperfect indications of the individual's "true" underlying latent class membership. Framed as a latent variable model, there are two parts to any mixture model: (1) the *measurement model*, and (2) the *structural model*. The statistical measurement model specifies the relationship between the underlying latent variable and the corresponding manifest variables. In the case of mixture models, the measurement model encompasses the number of latent classes and the class-specific distributions of the indicator variables. The structural model specifies the distribution of the latent variable in the population and the relationships between latent variables and between latent variables and corresponding observed predictors and outcomes (i.e., latent variable antecedent and consequent variables). In the case of unconditional mixture models, the structural model encompasses just the latent class proportions.

Finite Mixture Modeling As a Person-Centered Approach

Mixture models are obviously distinct from the more familiar latent variable factor models in which the underlying latent structure is made up of one or more continuous latent variables. The designation for mixture modeling often used in applied literature to highlight this distinction from factor analytic models does not involve the overt *categorical* versus *continuous* latent variable scale comparison but instead references mixture modeling as a *person-centered* or *person-oriented* approach (in contrast to *variable-centered* or *variable-oriented*). Person-centered approaches describe similarities and differences *among individuals* with respect to how

variables relate to each other and are predicated on the assumption that the population is heterogeneous with respect to the relationships between variables (Laursen & Hoff, 2006, p. 379). Statistical techniques oriented toward categorizing individuals by patterns of associations among variables, such as LCA and cluster analysis, are person-centered. Variable-centered approaches describe associations *among variables* and are predicated on the assumption that the population is homogeneous with respect to the relationships between variables (Laursen & Hoff, 2006, p. 379). In other words, each association between one variable and another in a variable-centered approach is assumed to hold for all individuals within the population. Statistical techniques oriented toward evaluating the relative importance of predictor variables, such as multivariate regression and structural equation modeling, are variable-centered.

Although "person-centered analysis" has become a popular and compelling catchphrase and methods-jingle for researchers to recite when providing the rationale for selecting a mixture modeling approach for their data analysis over a more traditional variable-centered approach, the elaborated justification, beyond the use of the catchphrase, is often flawed by placing person-centered and variable-centered approaches in juxtaposition as rival or oppositional approaches when, in fact, they are complementary. To understand this false dichotomy at the conceptual level, imagine that the data matrix, with rows of individuals and columns of variables, is a demarcated geographic region. You could explore this region from the ground (person-centered), allowing you to focus on unique, salient, or idiosyncratic features across the region, or you could explore this region from the air (variable-centered), allowing you to survey general and dominant features of the full expanse (e.g., the mean and covariance structure). Perhaps you might even elect to view the region both ways, recognizing that each provides a different perspective on the *same* region and that both advance your understanding of the region. That is, the region itself doesn't change but the information that can be gleaned about the region does change according to the type of search, and determining which search is more useful depends entirely on the objectives of the exploration.

The false dichotomy can also be explained in analytic terms, as Horn (2000) does so effectively in describing the linear decomposition of a n person × m variable data array:

> "In person-centered compared with variable-centered analyses, the theorem of Eckart and Young [(1936)] indicates that the linear relationships among variables have a counterpart in relationships among people. Or, to put the matter the other way around, the relationships among people that indicate types have a counterpart in relationships among variables that indicate factors . . . Quite simply, there is no variance in person-centered types that cannot be accounted for in terms of variable-centered factors, and vice-versa" (Horn, 2000, p. 925).

Beyond the conceptual and analytic considerations, there is also a practical rejection of the dichotomy between person- and variable-centered approaches. Although a majority of applications of mixture models claim and motivate an exclusive person-centered approach, most utilize strategies that combine person-centered and variable-centered elements. For example, it is not uncommon for a study to use a person-centered analysis to identify latent classes or groups of individuals characterized by different response patterns on a subset of variables and then use a variable-centered analysis to examine predictors and outcomes (antecedent and consequent correlates) of class membership. There are also many examples of "hybrid" models, such as growth mixture models, that use both latent factors (variable-centered) and latent classes (person-centered) to describe interindividual differences in intra-individual change.

With the dichotomy between person-centered and variable-centered approaches dispelled, you may be left wondering how to determine which approach to take or whether, indeed, your choice matters at all. The fact that it is possible to represent person-centered findings in variable-centered terms does not obfuscate the choice of approach but does make the explicit consideration of the fundamental assumptions of each approach in the context of the actual research question and available data all the more important. Further, explicit consideration must also be given to the consequences of choosing to represent a construct as one or more latent factors versus latent classes for the subsequent specification and testing of relationships between the construct and its hypothesized correlates. If your planned study aims at a person-centered level, and you can reasonably assume that your target population is heterogeneous in that there are actual types or classes to be revealed by an empirical study, then you have sufficient rationale for utilizing a person-centered or combined person-/variable-centered approach, and the choice

is clear. However, these rationales are not necessary for the purposed application of mixture models and I will touch on this topic again throughout the chapter, to recapitulate what constitutes principled use of mixture models.

Chapter Scope

This chapter is intended to provide the reader with a general overview of mixture modeling. I aim to summarize the current "best practices" for model specification, estimation, selection, evaluation, comparison, interpretation, and presentation for the two primary types of cross-sectional mixture analyses: latent class analysis (LCA), in which there are observed categorical indicators for a single latent class variable, and latent profile analysis (LPA), also known as latent class cluster analysis (LCCA), in which there are observed continuous indicators for a single latent class variable. As with other latent variable techniques, the procedures for model building and testing in these settings readily extend to more complex data settings—for example, longitudinal and multilevel variable systems. I begin by providing a brief historic overview of the two primary roots of modern-day mixture modeling in the social sciences and the foci of this chapter—*finite mixture modeling* and *LCA*—along with a summary of the purposed applications of the models. For each broad type of model, the general model formulation is presented, in both equations and path diagrams, followed by an in-depth discussion of model interpretation. Then a description of the model estimation including a presentation of current tools available for model evaluation and testing is provided, leading to a detailed illustration of a principled model building process with a full presentation and interpretation of results. Next, an extension of the unconditional mixture models already presented in the chapter is made to accommodate covariates using a latent class regression (LCR) formulation. I conclude the chapter with a brief cataloging of (some of) the many extensions of finite mixture modeling beyond the scope of this chapter, some cautionary notes about the misconceptions and misuses of mixture modeling, and a synopsis of prospective developments in the mixture modeling realm.

A Brief and Selective History of Mixture Modeling

Finite Mixture Modeling

Finite mixture modeling, in its most classic form, is a cross-sectional latent variable model in which the latent variable is nominal and the corresponding manifest variables are continuous. This form of finite mixture modeling is also known as LPA or LCCA. One of the first demonstrations of finite mixture modeling was done by a father of modern-day statistics, Karl Pearson, in 1894 when he fit a two-component (i.e., two-class) univariate normal mixture model to crab measurement data belonging to his colleague, Walter Weldon (1893), who had suspected that the skewness in the sample distribution of the crab measurements (the ratio of forehead to body length) might be an indication that this crab species from the Bay of Naples was evolving to two subspecies (McLachlan & Peel, 2000). Pearson used the method-of-moments to estimate his model and found evidence of the presence of two normally distributed mixing components that were subsequently identified as crab subspecies. There weren't many other mixture model applications that immediately followed suit because the daunting moments-based fitting was far too computationally intensive for mixtures. And it would take statisticians nearly 80 years to find more viable, as well as superior, alternative estimation procedures. Tan and Chang (1972) were among the researchers of their time that proved the maximum likelihood solution to be better for mixture models than the method-of-moments. Following on the heels of this insight was the release of the landmark article by Dempster, Laird, and Rubin (1977) that explicated, in general terms, an iterative estimation scheme—the expectation-maximization (EM) algorithm—for maximum-likelihood estimation from incomplete data. The recognition that finite mixture models could be easily reconceived as missing data problems (because latent class membership is missing for all individuals)—and thus estimated via the EM algorithm—represented a true turning point in the development of mixture modeling. Since that time, there has been rapid advancement in a variety of applications and extensions of mixture modeling, which are covered briefly in the section on "The More Recent Past" following the separate historical accounting of LCA.

Before moving on, there is another feature of the finite mixture history that is worth remarking on, as it relates to the earlier discussion of person-centered versus variable-centered approaches. Over the course of the twentieth century, there was a bifurcation in the development and application of finite mixture models in the statistical community following that early mixture modeling by Pearson, both before and after the advancement of the estimation algorithms. There was a distinction that

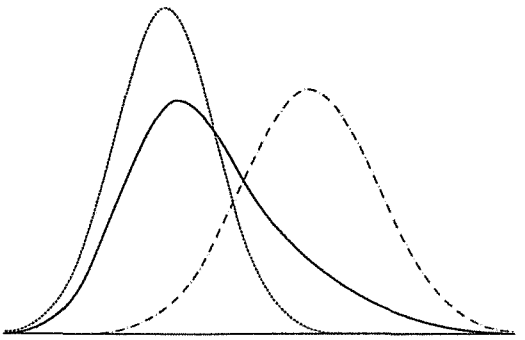

Figure 25.1 Hypothetical overall univariate non-normal population distribution (solid line) resulting from a mixing of two normally distributed subpopulations (dashed lines).

began to be made between *direct* and *indirect* applications (Titterington, Smith, & Makov, 1985) of finite mixture modeling. In direct applications, as in person-centered approaches, mixture models are used with the a priori assumption that the overall population is heterogeneous, and made up of a finite number of (latent and substantively meaningful) homogeneous groups or subpopulations, usually specified to have tractable distributions of indicators within groups, such as a multivariate normal distribution. In indirect applications, as in variable-centered approaches, it is assumed that the overall population is homogeneous and finite mixtures are simply used as more tractable, semi-parametric technique for modeling a population distribution of outcomes for which it may not be possible (practically or analytically speaking) to specify a parametric model. Mathematical work was done to prove that virtually any continuous distribution (even highly skewed, highly kurtotic, multimodal, or in other ways non-normal) could be approximated by the mixing of K normal distributions if K was permitted to be indiscriminately large and that a reasonably good approximation of most distributions could be obtained by the mixing of a relatively small number of normal distributions (Titterington, Smith, & Makov, 1985). Figure 25.1 provides an illustration of a univariate non-normal distribution that is the result of the mixing of two normally distributed components. The focus for indirect applications is then not on the resultant mixture components nor their interpretation but, rather, on the overall population distribution approximated by the mixing.

I find the *indirect* versus *direct* application distinction for mixture modeling less ambiguous than the *person-centered* versus *variable-centered* labels and, thus, will favor that language throughout the remainder of this chapter. Furthermore, the focus in this chapter is almost exclusively on direct applications of mixture models as I devote considerable time to the processes of class enumeration and interpretation and give weight to matters of classification quality, all of which are of little consequence for indirect applications.

Latent Class Analysis

Latent class models can be considered a special subset of finite mixture models formulated as a mixture of generalized linear models; that is, finite mixtures with discrete response variables with class-specific multinomial distributions. However, LCA has a rich history within the psychometric tradition, somewhat independent of the development of finite mixture models, that is worthy of remark, not unlike the way in which analysis of variance (ANOVA) and analysis of covariance (ANCOVA) models, although easily characterized as a special subset of multiple linear regression models, have their own historical timeline.

It didn't take long after Spearman's seminal work on factor analysis in 1904 for suggestions regarding categorical latent variables to appear in the literature. However, it wasn't until Lazarsfeld and Henry summarized their two decades of work on latent structure analysis (which included LCA as a subdomain of models) in 1968 that social scientists were presented with a comprehensive treatment of the theoretical and analytic features of LCA that had been in serious development since the 1950s.

Despite the expansive presentation and motivation for LCA provided by Lazarsfeld and Henry (1968), there were still two primary barriers to larger scale adoption of latent class models by applied researchers: (1) the categorical indicators could only be binary, and (2) there was no general, reliable, or widely implemented estimation method for obtaining parameter estimates (Goodman, 2002). Goodman (1974) resolved the first and part of the second problem with the development of a method for obtaining maximum likelihood estimates of latent class parameters for dichotomous and polytomous indicators. Once Goodman's estimation algorithm was implemented in readily available statistical software, first by Clogg in 1977, and Goodman's approach was shown to be closely related to the EM algorithm of Dempster, Laird, and Rubin (1977), currently the most widely utilized estimation algorithm for LCA software (Collins & Lanza, 2010), the remaining portion of the second barrier to the application of LCA was annulled. I will

return to matters related to maximum likelihood estimation for LCA parameters later in this chapter.

As with the history of finite mixture modeling, there is some comment necessary on the features of LCA history related to person-centered versus variable-centered approaches. Latent class models, with categorical indicators of a categorical latent variable, have, at different times, been described in both person-centered and variable-centered terms. For example, one of the fundamental assumptions in classical LCA is that the relationship among the observed categorical variables is "explained" by an underlying categorical latent variable (latent class variable)—that is, the observed variables are conditionally (locally) independent given latent class membership. In this way, LCA was framed as the pure categorical variable-centered analog to continuous variable-centered factor analysis (in which the covariances among the observed continuous variables is explained by one or more underlying continuous factors). Alternatively, LCA can be framed as a multivariate data reduction technique for categorical response variables, similarly to how factor analysis may be framed as a dimension-reduction technique that enables a system of m variables to be reduced to a more parsimonious system of q factors with $q \ll m$. Consider a set of 10 binary indicator variables. There are $2^{10} = 1024$ possible observed response patterns, and one could exactly represent the $n \times 10$ observed data matrix as a frequency table with 1024 (or fewer) rows corresponding to the actual observed response patterns. Essentially, in the observed data there are a maximum of 1024 groupings of individuals based on their observed responses. Latent class analysis then enables the researcher to group or cluster these responses patterns (and, thus, the individuals with those response patterns) into a smaller number of K latent classes ($K \ll 1024$) such that the response patterns for individuals within each class are more similar than response patterns across classes. For example, response patterns (1 1 1 1 1 1 1 1 1 1) and (0 1 1 1 1 1 1 1 1 1) might be grouped in the same latent class, different from (0 0 0 0 0 0 0 0 0 0) and (0 0 0 0 0 0 0 0 1 0). The classes are then characterized not by exact response patterns but by response *profiles* or typologies described by the relative frequencies of item endorsements. Because grouping the observed response patterns is tantamount to grouping individuals, this framing of LCA is more person-oriented. Thus, in both the psychometric tradition in which LCA was developed and in the classical mathematical statistics tradition in which finite mixture modeling was developed, mixture models have been used as both a person-centered and variable-centered approach, leading to some of the confusion surrounding the misleading association of mixture models as implicitly person-centered models and the false dichotomy between person-centered and variable-centered approaches.

The More Recent Past

In both finite mixture modeling and LCA, the utilization of the EM algorithm for maximum likelihood estimation of the models, coupled with rapid and widespread advancements in statistical computing, resulted in a remarkable acceleration in the development, extension, application, and understanding of mixture modeling over the last three decades, as well as a general blurring of the line that delineated latent class models from more general finite mixture models. A few of the many notable developments include the placement of latent class models within the framework of log linear models (Formann, 1982, 1992; Vermunt, 1999); LCR and conditional finite mixture models, incorporating predictors of class membership (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Dayton & Macready, 1988); and the placement of finite mixture modeling within a general latent structure framework, enabling multiple and mixed measurement modalities (discrete and continuous) for both manifest and latent variables (Hancock & Samuelson, 2008; Muthén & Shedden, 1999; Skrondal & Rabe-Hesketh, 2004). For an overview of the most recent developments in finite mixture modeling, *see* McLachlan and Peel (2000) and Vermunt and McCutcheon (2012). For more recent developments specifically related to LCA, *see* Hagenaars and McCutcheon (2002) and Collins and Lanza (2010).

There has also been conspicuous growth in the number of statistical software packages that enable the application of a variety of mixture models in real data settings. The two most prominent self-contained modeling software packages are Mplus V6.11 (Muthén & Muthén, 1998–2011), which is the software used for all the empirical examples in this chapter, and Latent GOLD V4.5 (Statistical Innovations, Inc., 2005–2011), both capable of general and comprehensive latent variable modeling, including, but not limited to, finite mixture modeling. The two most popular modular packages that operate within existing software are PROC LCA and PROC LTA for SAS (Lanza, Dziak, Huang, Xu, &

Collins, 2011), which are limited to traditional categorical indicator latent class and latent transition analysis models, and GLLAMM for Stata (Rabe-Hesketh, Skrondal & Pickles, 2004), which is a comprehensive generalized linear latent and mixed model framework utilizing adaptive quadrature for maximum likelihood estimation.

Access to software and the advancements in high-speed computing have also led to a remarkable expansion in the number of disciplines that have made use of mixture models as an analytic tool. There has been particularly notable growth in the direct application of mixture models within the behavioral and educational sciences over the last decade. Mixture models have been used in the empirical investigations of such varied topics as typologies of adolescent smoking within and across schools (Henry & Muthén, 2010); marijuana use and attitudes among high school seniors (Chung, Flaherty, & Schafer, 2006); profiles of gambling and substance use (Bray, 2007); risk profiles for overweight in adolescent populations (BeLue, Francis, Rollins, & Colaco, 2009); patterns of peer victimization in middle school (Nylund-Gibson, Graham, & Juvonen, 2010); liability to externalizing disorders (Markon & Krueger, 2005); profiles of academic self-concept (Marsh, Lüdtke, Trautwein, & Morin, 2009); profiles of program evaluators' self-reported practices (Christie & Masyn, 2010); rater behavior in essay grading (DeCarlo, 2005); mathematical ability for special education students (Yang, Shaftel, Glasnapp, & Poggio, 2005); patterns of public assistance receipt among female high school dropouts (Hamil-Luker, 2005); marital expectations of adolescents (Crissey, 2005); and psychosocial needs of cancer patients (Soothill, Francis, Awwad, Morris, Thomas, & Mclllmurray, 2004).

Latent Class Analysis

Although latent class models—mixture models with exclusively categorical indicator variables for the latent class variable—emerged more than a half-century after the inception of finite mixture models, I choose to use LCA for this initial foray into the details of mixture modeling because I believe it is the most accessible point of entry for applied readers.

Model Formulation

As with any latent variable model, there are two parts to a latent class model: (1) the measurement model, which relates the observed response variables (also called indicator or manifest variables) to the underlying latent variable(s); and (2) the structural model, which characterizes the distribution of the latent variable(s) and the relationships among latent variables and between latent variables and observed antecedent and consequent variables. In a traditional latent variable model-building process, the unconditional measurement model for each latent variable of interest is established prior to any structural model-based hypothesis testing. It is the results of the final measurement model that researchers use to assign meaning to the latent classes that are then used in the substantive interpretations of any structural relationships that emerge. Thus, the formal LCA model specification begins here with an unconditional model in which the only observed variables are the categorical manifest variables of the latent class variable.

Suppose there are M categorical (binary, ordinal, and/or multinomial) latent class indicators, $u_1, u_2, \ldots, u_M$ observed on n study participants where u_{mi} is the observed response to item m for participant i. It is assumed for the unconditional LCA that there is an underlying unordered categorical latent class variable, denoted by c, with K classes where $c_i = k$ if individual i belongs to Class k. The proportion of individuals in Class k, $\Pr(c = k)$, is denoted by π_k. The K classes are exhaustive and mutually exclusive such that each individual in the population has membership in exactly one of the K latent classes and $\Sigma\pi_k = 1$. The relationship between the observed responses on the M items and the latent class variable, c, is expressed as

$$\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}) = \sum_{k=1}^{K} \left[\pi_k \cdot \Pr(u_{1i}, u_{2i}, \ldots, u_{Mi} | c_i = k)\right]. \quad (4)$$

The above expression *is* the latent class measurement model. The measurement parameters are all those related to the class-specific response pattern probabilities, $\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi} | c_i = k)$, and the structural parameters are those related to the distribution of the latent class variable, which for the unconditional model are simply the class proportions, π_k.

The model expressed in Equation 4 can be represented by a path diagram as shown in Figure 25.2. All the path diagrams in this chapter follow the diagramming conventions used in the Mplus V6.11 software manual (Muthén & Muthén, 1998–2011): boxes to enclose observed variables; circles to enclose latent variables; single-headed arrow paths

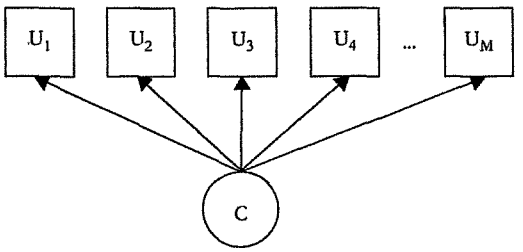

Figure 25.2 Generic path diagram for an unconditional latent class model.

to represent direct (causal) relationships; double-headed arrow paths to represent nondirection (correlational) relationships; "u" to denote observed categorical variables; "y" to denote observed continuous variables; "c" to denote latent categorical variables (finite mixtures or latent class variables); and "η" to denote latent continuous variables (factors).

Similarly to the typical default model specification in traditional factor analysis, *conditional* or *local independence* is assumed for the M items conditional on class membership. This assumption implies that latent class membership explains *all* of the associations among the observed items. Thus, the formation of the latent classes (in number and nature) based on sample data will be driven by the set of associations among the observed items in the overall sample. If all the items were independent from each other in the sample—that is, if all the items were uncorrelated in the overall sample—then it would not be possible to estimate a latent class model with more than $K = 1$ classes because there would be no observed associations to be explained by class membership. Under the local independence assumption, Equation 4 simplifies to

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

This assumption is represented in Figure 25.2 by the *absence* of any nondirectional (double-headed arrow) paths between the us that would represent item correlations *within* or *conditional on* latent class membership. The tenability of the local independence assumption can be evaluated and may also be *partially* relaxed (*see*, e.g., Huang & Bandeen-Roche, 2004). However, some degree of local independence is necessary for latent class model identification. It is not possible to fully relax this assumption for models with $K > 1$ classes—that is, an unconditional latent class model with all the items allowed to co-vary with all other items within class is not identified for $K > 1$ classes unless other parameter restrictions are imposed. I will revisit this assumption in the context of finite mixture modeling with continuous indicators. In that setting, models with $K > 1$ classes are identified even with all items co-varying within latent classes under certain other assumptions—for example, the distributional assumption of multivariate normality of the indicators within class.

Model Interpretation

As I mentioned earlier, it is the results of the final unconditional LCA, the measurement model, that are used to assign meaning to the latent classes, which augments the substantive interpretations of any structural relationships that emerge. Unless you are using mixture models in an indirect application as a semi-parametric approximation for an overall homogeneous population such that your attention will only be on parameter estimates for the overall (re)mixed population, you will focus your interpretation on the separate mixing components, interpreting each latent class based on the relationships between the classes and their indicators just as you use factor loadings and item communalities to interpret factors in a factor analysis. And just as with factor analysis, to reasonably interpret the latent class variable, you must have "good" measures of each of the classes.

A good item is one that measures the latent class variable well (i.e., reliably). A good latent class indicator is one for which there is a strong relationship between the item and the latent class variable. Strong item–class relationships must have both of the following features: (1) a particular item response—for example, item endorsement in the case of binary items, epitomizes members in at least one of the K latent classes in the model; and (2) the item can be used to distinguish between members across at least one pair of classes among the K latent classes in the model. The first quality is referred to as latent class *homogeneity* and the second quality is referred to as latent class *separation* (Collins & Lanza, 2010).

To better understand the concepts of latent class homogeneity and latent class separation, and how these concepts both relate to the parameters of the unconditional measurement model and ultimately qualify the interpretation of the resultant latent classes, consider a hypothetical example with five

binary response variables ($M = 5$) measuring a three-class categorical latent variable ($K = 3$). The unconditional model is given by

$$\Pr(u_{1i}, u_{2i}, u_{4i}, u_{4i}, u_{5i}) = \sum_{k=1}^{3}\left[\pi_k \cdot \left(\prod_{m=1}^{5} \omega_{m|k}\right)\right], \tag{6}$$

where $\omega_{m|k}$ is the probability that an individual belonging to Class k would endorse item m—that is, $\Pr(u_{mi} = 1|c_i = k) = \omega_{m|k}$.

Class Homogeneity. To interpret each of the K classes, you first need to identify items that epitomize each class. If a class has a high degree of homogeneity with respect to a particular item then there is a particular response category on that item that can be considered a response that typifies that class. In the case of binary items, strong associations with a particular class or high class homogeneity is indicated by high or low model-estimated probabilities of endorsement—that is, $\hat{\omega}_{m|k}$ or $1 - \hat{\omega}_{m|k}$close 1, with "close" defined by $\hat{\omega}_{m|k} > .7$ or $\hat{\omega}_{m|k} < .3$. For example, consider a class with an estimated class-specific item probability of 0.90. This means that in that class, an estimated 90% of individuals will endorse that particular item whereas only 10% will not. You could then consider this item endorsement as "typical" or "characteristic of" that class and could say that class has high homogeneity with respect to that item. Now consider a class with an estimated class-specific item probability of 0.55. This means that in that class, only an estimated 55% of individuals will endorse that particular item whereas 45% will not. Item endorsement is neither typical nor characteristic of that class, nor is lack of item endorsement, for that matter, and you could say that class has low homogeneity with respect to that item and would not consider that item a good indicator of membership for that particular class.

Class Separation. To interpret each of the K classes, you must not only have class homogeneity with respect to the items such that the classes are each well characterized by the item set, you also need to be able to distinguish *between* the classes—this quality is referred to as the degree of class separation. It is possible to have high class homogeneity and still have low class separation. For example, consider two classes, one of which has an estimated class-specific item probability of 0.90 and another class with an estimated class-specific item probability of 0.95. In this case, since item endorsement is "typical" for both of these classes and the two classes can be characterized by a high rate of endorsement for that item, they are not distinct from each other with respect to that item. Now consider two classes, one of which has an estimated class-specific item probability of 0.90 and another with an estimated class-specific item probability of 0.05. In this case, each class has good homogeneity with respect to the item and they also have a high degree of separation because the first class may be characterized by a high rate of item endorsement whereas the other class may be characterized by a high rate of item non-endorsement. To quantify class separation between Class j and Class k with respect to a particular item, m, compute the estimated item endorsement odds ratio as given by:

$$\hat{OR}_{m|jk} = \frac{\left(\hat{\omega}_{m|j}/1 - \hat{\omega}_{m|j}\right)}{\left(\hat{\omega}_{m|k}/1 - \hat{\omega}_{m|k}\right)}. \tag{7}$$

Thus, $\hat{OR}_{m|jk}$ is the ratio of the odds of endorsement of item m in Class j to the odds of endorsement of item m in Class k. A large $\hat{OR}_{m|jk} > 5$ (corresponding to approximately $\hat{\omega}_{m|j} > .7$ and $\hat{\omega}_{m|k} < .3$) or small $\hat{OR}_{m|jk} < .2$ (corresponding to approximately $\hat{\omega}_{m|j} < .3$ and $\hat{\omega}_{m|k} > .7$) indicates a high degree of separation between Classes j and k with respect to item m. Thus, high class homogeneity with respect to an item is a necessary but not sufficient condition for a high degree of class separation with respect to an item.

I should note here that although simply taking the ratio of the class-specific item response probabilities may seem more intuitive, the use of the odds ratio of item response rather than the response probability ratio is preferred because the odds ratio doesn't depend on whether you emphasize item endorsement or item non-endorsement separation or whether you are assessing the item-endorsement separation for classes with relatively high endorsement rates overall or low endorsement rates overall for the item in question. For example, an $\hat{OR}_{m|jk}$ of 0.44 corresponding to $\hat{\omega}_{m|j} = .80$ versus $\hat{\omega}_{m|k} = .90$ is the same as the $\hat{OR}_{m|jk}$ corresponding to $\hat{\omega}_{m|j} = .10$ versus $\hat{\omega}_{m|k} = .20$, whereas class-specific item probability ratios would be $.80/.90 = 0.87$ and $.10/.20 = 0.50$.

Class Proportions. It is possible, to a certain extent, to use the class proportion values themselves to assign meaning to the classes. Consider the case in which you have a population-based sample and one of the resultant classes has an estimated class proportion of greater than 0.50—that is, the class represents more than 50% of the overall population. Then part of your interpretation of this class

may include an attribution of "normal," "regular," or "typical" in that the class represents the statistical majority of the overall population. Similarly, if you had a resultant class with a small estimated class proportion (e.g., 0.10) part of your interpretation of that class might include an attribution of "rare," "unusual," or "atypical," ever mindful that such attribution labels, depending on the context, could carry an unintended negative connotation, implying the presence of deviance or pathology in the subpopulation represent by that class. Also remember that the estimated class proportion reflects the distribution of the latent classes *in the sample.* Thus, if you have a nonrandom or nonrepresentative sample, exercise caution when using the estimated class proportions in the class interpretations. For example, a "normal" class in a clinical sample may still be present in a nonclinical sample but may have a much smaller "atypical" representation in the overall population.

Hypothetical Example. Continuing with the hypothetical example of a three-class LCA with five binary indicators, Table 25.1 provides hypothetical model-estimated item response probabilities for each class along with the item response odds ratios calculated following Equation 7. Classes 1, 2, and 3 all have high homogeneity with respect to items u_1, u_2, and u_4 because all class-specific item response probabilities are greater than 0.70 or less than 0.30. Class 1 also has high homogeneity with respect to item u_3, whereas Classes 2 and 3 do not. Thus, Classes 2 and 3 are not well characterized by item u_3—that is, there is not a response to item u_3 that typifies either Class 2 or 3. None of the classes are well characterized by item u_5, and this might be an item that is considered for revision or elimination in future studies.

Class 1 is well separated from Class 2 by all the items except the last, with $\hat{OR}_{m|1,2} > 5$. Class 1 is not well distinguished from Class 3 by items u_1 and u_2 but is well separated from Class 3 by items u_3 and u_4. Classes 2 and 3 are well separated by items u_1 and u_2 but not by items u_3 and u_4. Thus, as a result of Classes 2 and 3 not being well characterized by item u_3, they are consequently not distinguishable from each other with respect to item u_3. Because none of the classes have a high degree of homogeneity with respect to item u_5, none of the classes are well separated from each other by that item.

The class homogeneity and separation information contained in Table 25.1 is usually depicted graphically in what is often referred to as a "profile plot" in which the class-specific item probabilities (y-values) are plotted in a line graph for each of the items (x-values). Figure 25.3 depicts a profile plot using the hypothetical model results presented in Table 25.1. I have added horizontal lines to the profile plot at 0.70 and 0.30 to assist in the visual inspection with respect to both class homogeneity and class separation. Class 1 can be interpreted as a group of individuals with a high propensity for endorsing items $u_1 - u_4$; Class 2, a group of individuals with a low propensity for endorsing items u_1, u_2, and u_4; and Class 3, a group of individuals with a high propensity for endorsing item u_1 and u_2 with a low propensity for endorsing item u_4. Notice that I do not use items with low class homogeneity for the interpretation of that class nor do I use language in the class interpretation that would imply a class separation with respect to an item that isn't meaningful.

Table 25.1. Hypothetical Example: Model-Estimated, Class-Specific Item Response Probabilities and Odds Ratios Based on a Three-Class Unconditional Latent Class Analysis

	$\hat{\omega}_{m\|k}$			$\hat{OR}_{m\|jk}$		
Item	Class 1 (70%)	Class 2 (20%)	Class 3 (10%)	Class 1 vs. 2	Class 1 vs. 3	Class 2 vs. 3
u_1	**0.90***	**0.10**	**0.90**	**81.00****	1.00	**0.01**
u_2	**0.80**	**0.20**	**0.90**	**16.00**	0.44	**0.03**
u_3	**0.90**	0.40	0.50	**13.50**	**9.00**	0.67
u_4	**0.80**	**0.10**	**0.20**	**36.00**	**16.00**	0.44
u_5	0.60	0.50	0.40	1.50	2.25	1.50

*Item probabilities >0.7 or <0.3 are bolded to indicate a high degree of class homogeneity.
**Odds ratios >5 or <0.2 are bolded to indicate a high degree of class separation.

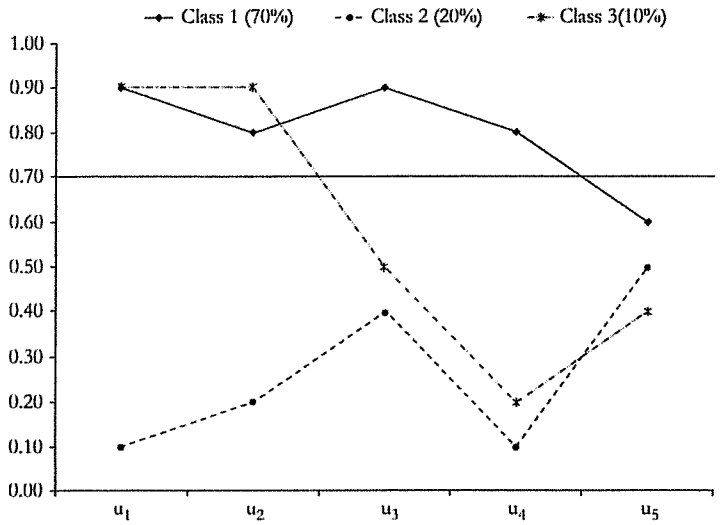

Figure 25.3 Hypothetical example: Class-specific item probability profile plot for a three-class unconditional LCA.

For example, both Classes 2 and 3 are interpreted as groups of individuals with low propensity for endorsing item u_4, but I do not, in the interpretation of Classes 2 and 3, imply that the two classes are somehow distinct with respect to u_4—only that they are both distinct form Class 1 with respect to u_4. I am also careful in my interpretation of the classes with categorical indicators to use explicit language regarding the probability or propensity of item endorsement rather than language that might incorrectly imply continuous indicators. For example, in this setting it would be incorrect to interpret Class 1 as a group of individuals with high levels of u_1 and u_2 with low levels of u_4, on average.

Based on the estimated class proportions, assuming a random and representative sample from the overall population, one might also apply a modifier label of "normal" or "typical" to Class 1 because its members make up an estimated 70% of the population.

The next three subsections present some of the technical details of LCA related to model estimation, model selection, and missing data. For the novice mixture modelers, I suggest that you may want to skip these subsections on your first reading of this chapter and go directly to the real data example that follows.

Model Estimation

As discussed in the mixture modeling historical overview, the most significant turning point for mixture modeling estimation was the development of the EM algorithm by Dempster, Laird, and Rubin (1977) for maximum likelihood (ML) estimation from incomplete data and the realization that if one reconceives of *latent* class membership as *missing* class membership, then the EM algorithm can be used to obtain maximum likelihood estimates (MLEs) of LCA parameters.

The first step in any ML estimation is specifying the likelihood function. The complete data likelihood function, put simply, is the probability density of all the data (the array of all values on all variables, latent and observed, in the model for all individuals in the sample) given a set of parameters values. Maximizing the likelihood function with respect to those parameters yields the maximum likelihood estimates (MLEs) of those parameters—that is, the MLEs are the values of the parameters that maximize the likelihood of the data. For a traditional LCA model, the complete data likelihood for a single individual i, with the missing latent class variable, c_i, is given by

$$l_i(\Theta) = \Pr(\mathbf{u}_i, c_i|\Theta) = \Pr(\mathbf{u}_i|c_i, \Theta) \cdot \Pr(c_i|\Theta), \tag{8}$$

where Θ is a vector of all the model parameters to be estimated. Typically, it is assumed that all cases are identically distributed such that the individual likelihood function, as expressed in Equation 8, is applicable for all cases. In the hypothetical LCA example with five binary indicators and three classes, Θ would include 18 separate parameters: all the class-specific item response probabilities along with the class proportions—that is, $\Theta = (\omega_{\cdot|1}, \omega_{\cdot|2}, \omega_{\cdot|3}, \pi_1, \pi_2, \pi_3)$, with $\omega_{\cdot|k} =$

$(\omega_{1|k}, \omega_{2|k}, \omega_{3|k}, \omega_{4|k}, \omega_{5|k})$. The likelihood function, L, for the whole sample is just the product of the individual likelihoods when assuming that all individuals in the sample are independent observations—that is, $L(\Theta) = \prod l_i(\Theta)$. Usually, it is easier mathematically to maximize the natural log of the likelihood function, $\ln(L(\Theta)) = LL(\Theta)$. Because the natural log is a monotonically increasing function, the values for Θ that maximize the log likelihood function are the maximum likelihood estimates, $\hat{\Theta}_{ML}$.

For most mixture models, with all individuals missing values for c, it is not possible obtain the MLEs by just applying the rules of calculus and solving a system of equations based on partial derivatives of the log likelihood function with respect to each parameter—that is, there is not a closed-form solution. Rather, an iterative approach must be taken in which successive sets of parameters estimates are tried using a principled search algorithm with a pair of stopping rules: (1) a maximum number of iterations and (2) a convergence criterion. To understand the concept behind iterative maximum likelihood estimation, consider the following analogy: imagine that the log likelihood function is a mountain range and the estimation algorithm is a fearless mountain climber. The goal of the climber is to reach the highest peak (global maximum) in the range, but the climber can't see where the highest peak is from the base of the mountain range. So the climber chooses an informed starting point (the initial staring values for the parameter estimates), using what he can see (the observed data), and begins to climb. Each foothold is a new set of parameter estimates. After each step the climber stops and assesses which of the footholds within reach (nearby parameter estimates values) will give him the greatest gain in height in a single step and he then leaves his current position to move to this higher point. He repeats this stepping process until he reaches a peak such that a step in any direction either takes him lower or not noticeably higher. The climber then knows he is at the peak (the convergence criterion is met), and it is here that he plants his flag, at the maximum log likelihood function value. But the climber, even as skilled as he is, cannot climb forever. He has limited food and water and so even if he has not reached the peak, there is a point at which he must stop climbing (the maximum number of iterations). If he runs out of supplies before he reaches a peak (exceeds the maximum number of iterations before meeting the convergence criterion), then he does not plant his flag (fails to converge).

As previously noted, the most common estimation algorithm in use for mixture models is the EM algorithm (Dempster, Laird, & Rubin, 1977; Muthén & Shedden, 1999). Each iteration of the EM algorithm involves an expectation step (E-step) in which the estimated expected value for each missing data value is computed based on the current parameters estimates and observed data for the individual. In the case of LCA, the E-step estimates expected class membership for each individual. The E-step is followed by a maximization step (M-step) in which new parameter estimates, $\hat{\Theta}$, are obtained that maximize the log likelihood function using the complete data from the E-step. Those parameter estimates are then used in the E-step of the next iteration, and the algorithm iterates until one of the stopping rules applies.

Although it would seem to go without saying, for the EM algorithm "mountain climber" to have even the slightest possibility of success in reaching the global peak of the log likelihood function, such a peak must exist. In other words, the model for which the parameters are being estimated must be *identified*—that is, there must be a unique solution for the model's parameters. However, this necessary fact may not be as trivial to establish as it would initially appear. When there is not a closed-form solution for the MLEs available, you cannot *prove*, mathematically speaking, that there is a global maximum. In this case, you are also unable to determine, theoretically, whether the solution you obtain from the estimation procedure is a global or local maximum nor can you tell, when faced with multiple local maxima (a mountainous range with many peaks of varying heights), whether the highest local maxima is actually the global maximum (highest peak). If the estimation algorithm fails to converge, then it could be an indication that the model is not theoretically identified, but it is not solid proof. There is also a gray area of empirical underidentification and weak identification in the span between identified models and unidentified models (i.e., models with no proper solution for all the model's parameters—failure of even one parameter to be identified causes the model to be under- or unidentified). This predicament is made more troublesome by the reality that the log likelihood surfaces for most mixture models are notoriously difficult for estimation algorithms to navigate, tending to have multiple local maxima, saddle points, and regions that are virtually flat, confusing even the most expert "climbers." To better understand some of the challenging log likelihood functions that may present

themselves, I discuss some exemplar log likelihood function plots for a unidimensional parameter space while providing some practical strategies to apply during the mixture model estimation process to help ensure the model you specify is well identified and the MLEs you obtain are stable and trustworthy solutions corresponding to a global maximum.

Figure 25.4 has six panels that represent a range of hypothetical log likelihood functions for a single parameter, θ. The unimodal log likelihood function in Figure 25.4.a has only one local maximum that is the global maximum. $\hat{\theta}_{(0)}$ is the MLE for θ because the $LL(\hat{\theta}_{(0)})$ is the maximum value achieved by $LL(\theta)$ across all values of θ. It is clear that no matter what starting position on the x-axis (starting value, $\hat{\theta}_{(s)}$, for the estimate of the parameter, θ) is selected, the mountain climber would easily find that global peak. This LL function reflects a well-identified model. There is one unique global maximum (MLE) that would be readily reached from any starting point. Now examine the multimodal likelihood function in Figure 25.4.b. There is still a single global maximum, $\hat{\theta}_{(0)}$, but there are three other local maxima, $\hat{\theta}_{(1)}$,$\hat{\theta}_{(2)}$, and $\hat{\theta}_{(3)}$. You can imagine that if you started your algorithm mountain climber at a point $\hat{\theta}_{(s)} < \hat{\theta}_{(2)}$, then he might conclude his climb, reaching the convergence criterion and planting his flag, on the peak of the log likelihood above $\hat{\theta}_{(2)}$, never realizing there were higher peaks down range. Similarly, if you started your climber at a point $\hat{\theta}_{(s)} > \hat{\theta}_{(1)}$, then he might conclude his climb on the peak of the log likelihood above $\hat{\theta}_{(1)}$, never reaching the global peak above $\hat{\theta}_{(0)}$.

With a log likelihood function like the one depicted in Figure 25.4.b, one could expect the estimation algorithm to converge on a local rather than global maximum. If you obtained only one solution, $\hat{\theta}$, using one starting value, $\hat{\theta}_{(s)}$, then you have no way of knowing whether $\hat{\theta}$ corresponds to the highest peak in the range or just a peak of the log likelihood in the range of θ. Because it isn't possible to resolve this ambiguity mathematically, it must be resolved empirically. In keeping with the analogy, if you want to find the highest peak in the range, then rather than retaining a single expert mountain climber, you could retain the services of a whole army of expert mountain climbers. You start each climber at a different point in the range. A few will "converge" to the lower local peaks, but most should reach the global peak. The more climbers from different starting points (random sets of starting values) that converge to the same peak (solution replication), the more confident you are in that particular peak being the global maximum. This strategy corresponds to using multiple sets of random starting values for the EM algorithm, iterating each set of starting values to convergence, and demanding a high frequency (in absolute and relative terms) of replication of the best log likelihood value.

I should note here that although replication of the maximum likelihood solution from different

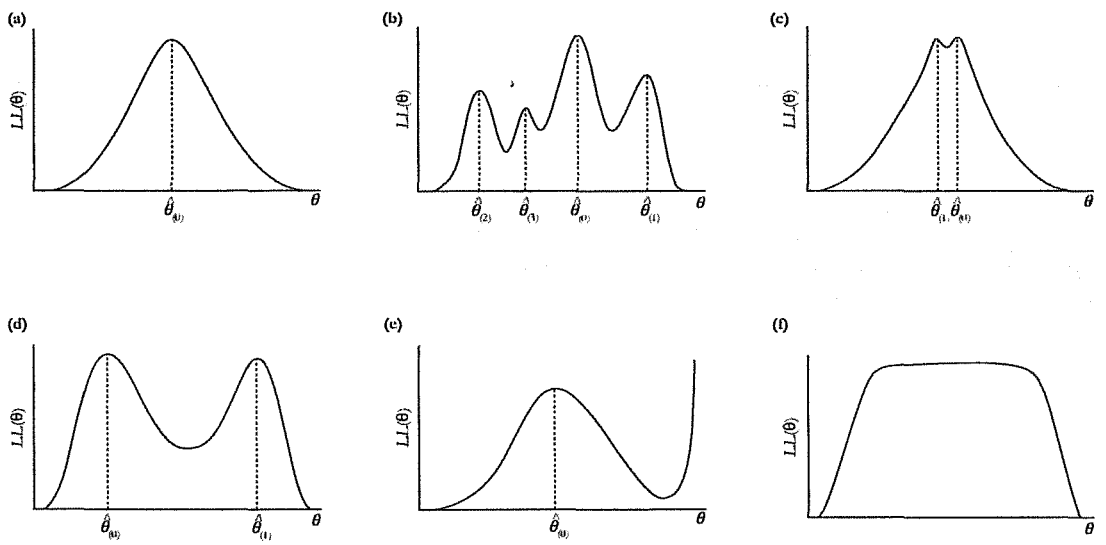

Figure 25.4 Hypothetical log likelihood (LL) functions for a single parameter, θ: (a) unimodal LL; (b) multimodal LL; (c) bimodal LL with proximate local maxima; (d) bimodal LL with distant local maxima; (e) unbounded LL; and (f) LL with flat region.

sets of starting values increases confidence in that solution as the global optimum, replication of the likelihood value is neither a necessary nor a sufficient requirement to ensure that a global (rather than a local) maximum has been reached (Lubke, 2010). Thus, failure to replicate the best log likelihood value does not mean that you must discard the model. However, further exploration should be done to inform your final model selection. Consider the cases depicted in Figures 25.4.c and 25.4.d for which there is a global maximum at $\hat{\theta}_{(0)}$ and a local maximum of nearly the same log likelihood value at $\hat{\theta}_{(1)}$. In cases such as these, the relative frequency of replication for each of the two solutions across a random set of start values may also be comparable. In Figure 25.4.c, not only are the two solution very close in terms of the log likelihood values, they are also close to each other in the range of θ such that $\hat{\theta}_{(0)} \approx \hat{\theta}_{(1)}$. In this case you can feel comforted by the fact that even if you had inadvertently missed the global maximum at $\hat{\theta}_{(0)}$ and incorrectly taken $\hat{\theta}_{(1)}$ as your MLE, your inferences and interpretations would be close to the mark. However, in the case depicted in Figure 25.4.d, $\hat{\theta}_{(1)}$ is quite distant on the scale of θ from $\hat{\theta}_{(0)}$ and you would not want to base conclusions on the $\hat{\theta}_{(1)}$ estimate. To get a sense of whether the highest local peaks in your log likelihood function are proximal or distal solutions in the parameter space, obtain the actual parameter estimates for the best log likelihood value across all the sets of random starting values and make a descriptive comparison to the parameter estimates corresponding to the "second-best" log likelihood value. (For more about comparing local maximum log likelihood solutions to determine model stability, *see*, for example, Hipp & Bauer, 2006.)

I pause here to make the reader aware of a nagging clerical issue that must be tended to whenever different maximum likelihood solutions for mixture models are being compared, whether for models with the same or differing numbers of classes: label switching (Chung, Loken, & Schafer, 2004). The ordering of the latent classes as they are outputted by an estimation algorithm are completely arbitrary—for example, "Class 1" for starting values Set 1 may correspond to "Class 3" for starting values Set 2. Even solutions identical in maximum likelihood values can have class labels switched. This phenomenon is not a problem statistically speaking—it merely poses a bookkeeping challenge. So be cognizant of label switching whenever you are comparing mixture model solutions.

Figures 46.4.e and 46.4.f depict log likelihood functions that would be likely to result in either some or all of the random sets of starting values failing to converge—that is, the estimation algorithm stops because the maximum number of iterations is exceeded before a peak is reached. In Figure 25.4.e, the log likelihood function is unbounded at the boundary of the range of θ (which is not an uncommon feature for the *LL* function of mixture models with more complex within-class variance–covariance structures) but also has a maximum in the interior of the range of $\theta \cdot \hat{\theta}_{(0)}$ represents the proper maximum likelihood solution, and that solution should replicate for the majority of random sets of stating values; however, some in the army of expert mountain climbers are likely to find themselves climbing the endless peak, running out of supplies and stopping before convergence is achieved. The log likelihood function in Figure 25.4.f corresponds to an unidentified model. The highest portion of the log likelihood function is flat and there are not singular peaks or unique solutions. No matter where the estimation algorithm starts, it is unlikely to converge. If it does converge, then that solution is unlikely to replicate because it will be a false optimum.

A model that is *weakly identified* or *empirically underidentified* is a model that, although theoretically identified, has a shape with particular sample data that is nearly flat and/or has many, many local maxima of approximately the same height (think: egg-crate) such that the estimation algorithm fails to converge for all or a considerable number of random sets of starting values. For a model to be identified, there must be enough "known" or observed information in the data to estimate the parameters that are not known. Ensuring positive degrees of freedom for the model is a necessary but not sufficient criterion for model identification. As the ratio of "known" to "unknown" decreases, the model can become weakly identified. One quantification of this ratio of information for MLE is known as the *condition number*. It is computed as the ratio of the smallest to largest eigenvalue of the information matrix estimate based on the maximum likelihood solution. A low condition number, less than 10^{-6}, may indicate singularity (or near singularity) of the information matrix and, hence, model non-identification (or empirical underidentification) (Muthén & Muthén, 1998–2011). A final indication that you may be "spreading" your data "too thin" is class collapsing, which can occur when you are attempting to extract more latent classes than your data will support. This

collapsing usually presents as one or more estimated class proportions nearing zero but can also emerge as a nearly complete lack of separation between two or more of the latent classes.

Strategies to achieve identification all involve reducing the complexity of the model to increase the ratio of "known" to "unknown" information. The number of latent classes could be reduced. Alternatively, the response categories for one or more of the indicator variables in the measurement model could be collapsed. For response categories with low frequencies, this category aggregation will remove very little information about population heterogeneity while reducing the number of class-specific item parameters that must be estimated. Additionally, one or more items might be combined or eliminated from the model. This item-dropping must be done with extreme care, making sure that removal or aggregation does not negatively impact the model estimation (Collins & Lanza, 2010). Conventional univariate, bivariate, and multivariate data screening procedures should result in careful data recoding and reconfiguration that will protect against the most obvious threats to empirical identification.

In summary, MLE for mixture models can present statistical and numeric challenges that must be addressed during the application of mixture modeling. Without a closed-form solution for the maximization of the log likelihood function, an iterative estimation algorithm—typically the EM algorithm—is used. It is usually not possible to prove that the model specified is theoretically identified, and, even if it was, there could still be issues related to weak identification or empirical underidentification that causes problems with convergence in estimation. Furthermore, since the log likelihood surface for mixtures is often multimodal, if the estimation algorithm does converge on a solution, there is no way to know for sure that the point of convergence is at a global rather than local maximum. To address these challenges, it is recommended the following strategy be utilized during mixture model estimation. First and foremost, use multiple random sets of starting values with the estimation algorithm (it is recommended that a minimum of 50–100 sets of extensively, randomly varied starting values be used (Hipp & Bauer, 2006), but more may be necessary to observe satisfactory replication of the best maximum log likelihood value) and keep track of the information below:

1. the number and proportion of sets of random starting values that converge to proper solution (as failure to consistently converge can indicate weak identification);
2. the number and proportion of replicated maximum likelihood values for each local and the apparent global solution (as a high frequency of replication of the apparent global solution across the sets of random starting values increases confidence that the "best" solution found is the true maximum likelihood solution);
3. the condition number for the best model (as a small condition number can indicate weak or nonidentification); and
4. the smallest estimated class proportion and estimated class size among all the latent classes estimated in the model (as a class proportion near zero can be a sign of class collapsing and class overextraction).

This information, when examined collectively, will assist in tagging models that are nonidentified or not well identified and whose maximum likelihoods solutions, if obtained, are not likely to be stable or trustworthy. Any not well-identified model should be discarded from further consideration or mindfully modified in such a way that the empirical issues surrounding the estimation for that particular model are resolved without compromising the theoretical integrity and substantive foundations of the analytic model.

Model Building

Placing LCA in a broader latent variable modeling framework conveniently provides a ready-made general sequence to follow with respect to the model-building process. The first step is always to establish the measurement model for each of the latent variables that appear in the structural equations. For a traditional LCA, this step corresponds to establishing the measurement model for the latent class variable.

Arguably, the most fundamental and critical feature of the measurement model for a latent class variable is the *number of* latent classes. Thus far, in my discussion of model specification, interpretation, and estimation, the number of latent classes, K, has been treated as if it were a known quantity. However, in most all applications of LCA, the number of classes is not known. Even in direct applications, when one assumes *a priori* that the population is heterogeneous, you rarely have specific hypotheses regarding the exact number or nature of the subpopulations. You may have certain hunches about one or

more subpopulations you expect to find, but rarely are these ideas so well formed that they translate into an exact total number of classes and constraints on the class-specific parameters that would inform a measurement model specification similar to the sort associated with CFA. And in indirect applications, as you are only interested in making sure you use enough mixture components (classes) to adequately describe the overall population distribution of the indicator variables, there is no preformed notion of class number. Thus, in either case (direct or indirect), you must begin with the model building with an exploratory class enumeration step.

Deciding on the number of classes is often the most arduous phase of the mixture modeling process. It is labor intensive because it requires consideration (and, therefore, estimation) of a set of models with a varying numbers of classes, and it is complicated in that the selection of a "final" model from the set of models under consideration requires the examination of a host of fit indices along with substantive scrutiny and practical reflection, as there is no single method for comparing models with differing numbers of latent classes that is widely accepted as best (Muthén & Asparouhov, 2006; Nylund, Asparouhov, & Muthén, 2007). This section first reviews the preferred tools available for the statistical evaluation of latent class models and then explains how these tools may be applied in concert with substantive evaluation and the parsimony principle in making the class enumeration determination. The tools are divided into three categories: (1) evaluations of absolute fit; (2) evaluations of relative fit; and (3) evaluations of classification.

Absolute Fit. In evaluating the absolute fit of a model, you are comparing the model's representation of the data to the actual data—that is, the overall model-data consistency. Recall that in traditional LCA, the observed data for individual responses on a set of categorical indicator variables can be summarized by a frequency table where each row represents one of the finite number of possible response patterns and the frequency column contains the number of individuals in the sample manifesting each particular pattern. The entire $n \times M$ data matrix can be identically represented by $R \times (M+1)$ frequency table where R is the number of total observed response patterns. For example, in the hypothetical LCA example with five binary indicator variables, there would be $2^5 = 32$ possible response patterns with $R \leq 32$. Assuming for the moment that $R = 32$, all the observed data on those five binary indicators could be represented in the following format:

u_1	u_2	u_3	u_4	u_5	f_r
1	1	1	1	1	f_1
1	1	1	1	0	f_2
1	1	1	0	1	f_3
⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	f_{32}

where f_r is the number of individuals in the sample with response pattern r corresponding to specific responses to the us displayed in row r of the table and $\sum f_r = n$. Thus, when evaluating absolute fit for a latent class measurement model, comparing the model representation of the data to the actual data will mean comparing the model-estimated frequencies to the observed frequencies across all the response patterns.

The most common test of absolute fit for observed categorical data and the one preferred in the LCA setting is the likelihood ratio (LR) chi-square goodness-of-fit test (Agresti, 2002; Collins & Lanza, 2010; McCutcheon, 1987). The test statistic, X^2_{LR} (sometime denoted by G^2 or L^2), is calculated as follows:

$$X^2_{LR} = 2\sum_{r=1}^{R}\left[f_r \log\left(\frac{f_r}{\hat{f}_r}\right)\right], \tag{9}$$

where R is the total number of observed data response patterns; f_r is the observed frequency count for the response pattern r; and $\hat{f}_r$ is the model-estimated frequency count for the response pattern r. Under the null hypothesis that the data are governed by the assumed distribution of the specified model, the test statistic given in Equation 9 is distributed chi-square with degrees of freedom given by

$$df_{X^2_{LR}} = R - d - 1, \tag{10}$$

where d is the number of parameters estimated in the model. When the model fits the sample data perfectly (i.e., $f_r = \hat{f}_r$, $\forall r$), the test statistic, X^2_{LR}, is equal to zero and the p-value is equal to 1. Failure to reject the null hypothesis implies adequate model-data consistency; rejection of the null implies the model does not adequately fit the data—the larger the test statistic, the larger the discrepancy and the poorer the fit between the model representation and the actual observed data.

Although it is very useful to have a way to statistical evaluate overall goodness-of-fit of a model to the data, the X^2_{LR} test statistic relies on large sample theory and may not work as intended (i.e., X^2_{LR} may not be well approximated by a chi-square distribution under the null hypothesis, marking the p-values based on that distribution of questionable validity) when the data set is small or the data are sparse, meaning there is a non-negligible number of response patterns with small frequencies (Agresti, 2002). There are some solutions, including parametric bootstrapping and posterior predictive checking, that are available to address this shortcoming (Collins & Lanza, 2010) but they are not widely implemented for this particular goodness-of-fit test in most mixture modeling software and are beyond the scope of this chapter.

Chi-square goodness-of-fit tests, in general, are also known to be sensitive to what would be considered negligible or inconsequential misfit in very large samples. In these cases, the null hypothesis may be rejected and the model determined to be statistically inadequate but, upon closer practical inspection, may be ruled to have a "close enough" fit. In factor analysis models, there is a wide array of closeness-of-fit indices for one to reference in addition to the exact-fit chi-square test, but this is not the case for mixture models. However, you can still inspect the closeness-of-fit for latent class models by examining the standardized residuals. Unlike residual diagnostics in the regression model, which compare each individual's predicted outcome to the observed values, or residual diagnostics in factor analysis, which compare the model-estimated means, variances, covariances, and correlations to the observed values, the LCA residuals are constructed using the same information that goes into the overall goodness-of-fit test statistic: the model-estimated response pattern frequencies and the observed frequencies. The raw residual for each response pattern is simply the difference between the observed and model-estimated frequency, $\hat{res}_r = f_r - \hat{f}_r$, and the standardized residual is calculated by

$$stdr\hat{e}s_r = \frac{f_r - \hat{f}_r}{\sqrt{\hat{f}_r\left(1 - \frac{\hat{f}_r}{n}\right)}}. \qquad (11)$$

The values of the standardized residuals can be compared to a standard normal distribution (Haberman, 1973), with large values (e.g., $|stdr\hat{e}s_r| > 3$) indicating response patterns that were more poorly fit, contributing the most to the X^2_{LR} and the rejection of the model. Because the number of possible response patterns can become large very quickly with increasing numbers of indicators and/or response categories per indicator, it is common to have an overwhelmingly large number of response patterns, many with observed and expected frequencies that are very small—that is, approaching or equal to zero. However, there is usually a much smaller subset of response patterns with relatively high frequencies, and it can be helpful to focus your attention on the residuals of these patterns where the bulk of data reside (Muthén & Asparouhov, 2006). In addition to examining the particular response patterns with large standardized residuals, it is also relevant to examine the overall proportion of response patterns with large standardized residuals. For a well-fitting model, one would still expect, by chance, to have some small percentage of the response patterns to have significant residual values, so you would likely only take proportions in notable excess of, say, 1% to 5%, to be an indication of a poor-fitting model.

Relative Fit. In evaluating the relative fit of a model, you are comparing the model's representation of the data to another model's representation. Evaluations of relative fit do not tell you anything about the *absolute fit* so keep in mind even if one model is a far better fit to the data than another, *both* could be poor in overall goodness of fit.

There are two categories of relative fit comparisons: (1) inferential and (2) information-heuristic. The most common ML-based inferential comparison is the likelihood ratio test (LRT) for nested models. For a Model 0 (null model) to be nested within a Model 1 (alternative model), Model 0 must be a "special case" of Model 1—that is, Model 0 is Model 1 with certain parameter restrictions in place. The likelihood ratio test statistic (LRTS) is computed as

$$X^2_{diff} = -2\left(LL_0 - LL_1\right), \qquad (12)$$

where LL_0 and LL_1 are the maximized log likelihood values to which the EM algorithm converges during the model estimation for Model 0 and Model 1, respectively. Under the null hypothesis that there is no difference between the two models (i.e., that the parameter restrictions placed on Model 1 to obtain Model 0 are restrictions that match the true population model) and with certain regularity conditions in place (e.g., the parameter restrictions do not fall on the boundary of the parameter space), X^2_{diff} has a chi-square distribution with degrees of freedom given by

$$df_{diff} = d_1 - d_0, \qquad (13)$$

where d_1 and d_0 are the numbers of parameters estimated in Model 1 and Model 0, respectively. Failure to reject the null hypothesis implies there is not statistically significant difference in fit to the data between Model 0 and Model 1. Thus, Model 0 would be favored over Model 1 since it is a simpler model with comparable fit. Rejection of the null hypothesis would imply that the parameter restrictions placed on Model 1 to obtain Model 0 resulted in a statistically significant decrement of fit. In general, this result would lead you to favor Model 1, unless the absolute fit of Model 0 was already deemed adequately. Following the principle of parsimony, if Model 0 had adequate absolute fit, then it would likely be favored over any more complicated and parameter-laden model, even if the more complicated model fit significantly better, relatively speaking.

There are two primary limitations of the likelihood ratio test comparison of relative model fit: (1) it only allows the comparison of two models at a time, and (2) those two models must be nested under certain regularity conditions. Information-heuristic tools overcome those two limitations by allowing the comparison of relative fit across a set of models that may or may not be nested. The downside is the comparisons are descriptive—that is, you can use these tools to say one model is "better" than another according to a particular criterion but you can't test in a statistical sense, as you can with the X^2_{diff}, *how much* better. Most information-heuristic comparisons of relative fit are based on *information criteria* that weigh the fit of the model (as captured by the maximum log likelihood value) in consideration of the model complexity. These criteria recognize that although one can always improve the fit of a model by adding parameters, there is a cost for that improvement in fit to model parsimony. These information criteria can be expressed in the form

$$-2LL + penalty, \tag{14}$$

where LL is the maximized log likelihood function value to which the EM algorithm converges during the model estimation. The *penalty* term is some measure of the complexity of the model involving sample size and the number of parameters being estimated in the model. For model comparisons, a particular information criterion value is computed for each of the models under consideration, and the model with the minimum value for that criterion is judged as the (relative) best among that set of models. What follows is a cataloging of the three most common information criteria used in mixture model relative fit comparisons. These criteria differ only in the computation of the penalty term.

- Bayesian Information Criterion (BIC; Schwarz, 1978)

$$BIC = -2LL + d\log(n), \tag{15}$$

where d is the number of parameters estimated in the model; n is the number of subjects or cases, in the analysis sample.

- Consistent Akaike's Information Criterion (CAIC; Bozdogan, 1987)

$$CAIC = -2LL + d[\log(n) + 1]. \tag{16}$$

- Approximate Weight of Evidence Criterion (AWE; Banfield & Raftery, 1993)

$$AWE = -2LL + 2d[\log(n) + 1.5]. \tag{17}$$

Although the information-heuristic descriptive comparisons of model are usually ordinal in nature, there are a few descriptive quantifications of relative fit based on information criteria that, although still noninferential, do allow you to get a sense of "how much" better one model is relative to another model or relative to a whole set of models. The two quantifications presented here are based on rough approximations to comparisons available in a Bayesian estimation framework and have been popularized by Nagin (1999) in the latent class growth modeling literature.

The first, the approximate Bayes Factor (BF), is a pairwise comparison of relative fit between two models, Model A and Model B. It is calculated as

$$\hat{BF}_{A,B} = \exp[SIC_A - SIC_B], \tag{18}$$

where SIC is the Schwarz Information Criterion (Schwarz, 1978), given by

$$SIC = -0.5BIC. \tag{19}$$

$BF_{A,B}$ represents the ratio of the probability of Model A being the correct model to Model B being the correct model when Models A and B are considered the competing models. According to Jeffrey's Scale of Evidence (Wasserman, 1997), $1 < BF_{A,B} < 3$ is weak evidence for Model A, $3 < BF_{A,B} < 10$ is moderate evidence for Model A, and $BF_{A,B} > 10$ is considered strong evidence for Model A. Schwarz (1978) and Kass and Wasserman (1995) showed that $\hat{BF}_{A,B}$ as defined in Equation 18 is a reasonable approximation of $BF_{A,B}$ when equal weight is placed on the prior probabilities of Models A and B (Nagin, 1999).

The second, the approximate correct model probability (cmP), allows relative comparisons of each of

J models to an entire set of J models under consideration. There is a cmP value for each of the models, Model A (A = 1, ..., J) computed as

$$\hat{cmP}_A = \frac{\exp\left(SIC_A - SIC_{\max}\right)}{\sum_{j=1}^{J} \exp\left(SIC_j - SIC_{\max}\right)}, \qquad (20)$$

where $SIC_{\max}$ is the maximum SIC score of the J models under consideration. In comparison to the $\hat{BF}_{A,B}$, which compares only two models, the cmP is a metric for comparing a set of more than two models. The sum of the cmP values across the set of models under consideration is equal to 1.00—that is, the true model is assumed to be one of the models in the set. Schwarz (1978) and Kass and Wasserman (1995) showed that $\hat{cmP}_A$ as defined in Equation 20 is a reasonable approximation of the actual probability of Model A being the correct model relative to the other J models under consideration when equal weight is placed on the prior probabilities of all the model (Nagin, 1999). The ratio of $\hat{cmP}_A$ to $\hat{cmP}_B$ when the set of models under consideration is limited to only Models A and B reduces to $\hat{BF}_{A,B}$.

Classification Diagnostics. Evaluating the precision of the latent class assignment for individuals by a candidate model is another way of assessing the degree of class separation and is most useful in direct applications wherein one of the primary objectives is to extract from the full sample empirically well-separated, highly-differentiated groups whose members have a high degree of homogeneity in their responses on the class indicators. Indeed, if there is a plan to conduct latent class assignment for use in a subsequent analysis—that is, in a multistage classify–analyze approach, the within-class homogeneity and across-class separation and differentiation is of primary importance for assessing the quality of the model (Collins & Lanza, 2010). Quality of classification could, however, be completely irrelevant for indirect applications. Further, it is important to keep in mind that it is possible for a mixture model to have a good fit to the data but still have poor latent class assignment accuracy. In other words, model classification diagnostics can be used to evaluate the utility of the latent class analysis as a model-based clustering tool for a given set of indicators observed on a particular sample but should *not* be used to evaluate the model-data consistency in either absolute or relative terms.

All of the classification diagnostics presented here are based on estimated *posterior class probabilities.* Posterior class probabilities are the model-estimated values for each individual's probabilities of being in each of the latent classes based on the maximum likelihood parameter estimates and the individual's observed responses on the indicator variables. The posterior class probability for individual i corresponding to latent Class k, $\hat{p}_{ik}$, is given by

$$\hat{p}_{ik} = \hat{\Pr}(c_i = k|\mathbf{u}_i, \hat{\theta}) = \frac{\hat{\Pr}(\mathbf{u}_i|c_i = k, \hat{\theta}) \cdot \hat{\Pr}(c_i = k)}{\hat{\Pr}(\mathbf{u}_i)}, \qquad (21)$$

where $\hat{\theta}$ is the set of parameter estimates for the class-specific item response probabilities and the class proportions. Standard *post hoc* model-based individual classification is done using *modal* class assignment such that each individual in the sample is assigned to the latent class for which he or she has the largest posterior class probability. In more formal terms, model-based modally assigned class membership for individual i, $\hat{c}_{\text{modal},i}$, is given by

$$\hat{c}_{\text{modal},i} = k : \max(\hat{p}_{i1}, \ldots, \hat{p}_{iK}) = \hat{p}_{ik}. \qquad (22)$$

Table 25.2 provides examples of four individual sets of estimated posterior class probabilities and the corresponding modal class assignment for the hypothetical three-class LCA example. Although individuals 1 and 2 are both modally assigned to Class 1, individual 1 has a very high estimated posterior class probability for Class 1, whereas individual 2 is not well classified. If there were many cases like individual 2, then the overall classification accuracy would be low as the model would do almost no better than random guessing at predicting latent class membership. If there were many cases like individual 1, then the overall classification accuracy would be high. The first classification diagnostic, relative entropy, offers a systematic summary of the levels of posterior class probabilities across classes and individuals in the sample.

Relative entropy, E_K, is an index that summarizes the overall precision of classification for the whole sample across all the latent classes (Ramasway, DeSarbo, Reibstein, & Robinson, 1993). It is computed by

$$E_K = 1 - \frac{\sum_{i=1}^{n} \sum_{k=1}^{K} \left[-\hat{p}_{ik} \ln(\hat{p}_{ik})\right]}{n \log(K)}. \qquad (23)$$

E_K measures the posterior classification uncertainty for a K-class model and is bounded between 0 and 1; $E_K = 0$ when posterior classification is no better than random guessing and $E_K = 1$ when there is perfect posterior classification for all individuals in the sample—that

Table 25.2. Hypothetical Example: Estimated Posterior Class Probabilities and Modal Class Assignment Based on Three-Class Unconditional Latent Class Analysis for Four Sample Participants

i	$\hat{p}_{ik1}$			$\hat{c}_{\text{modal}(i)}$
	$\hat{p}_{i1}$	$\hat{p}_{i2}$	$\hat{p}_{i3}$	
1	0.95	0.05	0.00	1
2	0.40	0.30	0.30	1
3	0.20	0.70	0.10	2
4	0.00	0.00	1.00	3

is, $\max(\hat{p}_{i1}, \hat{p}_{i2}, \ldots, \hat{p}_{iK}) = 1.00$, $\forall i$. Because even when E_K is close to 1.00 there can be a high degree of latent class assignment error for particular individuals, and because posterior classification uncertainty may increase simply by chance for models with more latent classes, E_K was never intended for, nor should it be used for, model selection during the class enumeration process. However, E_K values near 0 may indicate that the latent classes are not sufficiently well separated for the K classes that have been estimated (Ramaswamy et al., 1993). Thus, E_K may be used to identity problematic overextraction of latent classes and may also be used to judge the utility of the LCA directly applied to a particular set of indicators to produce highly-differentiated groups in the sample.

The next classification diagnostic, the average posterior class probability (AvePP), enables evaluation of the specific classification uncertainty for each of the latent classes. The AvePP for Class k, $AvePP_k$, is given by

$$AvePP_k = Mean\{\hat{p}_{ik},\ \forall i : \hat{c}_{\text{modal},i} = k\}. \quad (24)$$

That is, $AvePP_k$ is the mean of the Class k posterior class probabilities across all individuals whose maximum posterior class probability is for Class k. In contrast to E_K which provides an overall summary of latent class assignment error, the set of $AvePP_k$ quantities provide class-specific measures of how well the set of indicators predict class membership in the sample. Similarly to E_K, $AvePP_k$ is bounded between 0 and 1; $AvePP_k = 1$ when the Class k posteriori probability for every individual in the sample modally assigned to Class k is equal to 1. Nagin (2005) suggests a rule-of-thumb that all AvePP values be above 0.70 (i.e., $AvePP_k > .70$, $\forall k$) to consider the classes well separated and the latent class assignment accuracy adequate.

The odds of correct classification ratio (OCC; Nagin, 2005) is based on the $AvePP_k$ and provides a similar class-specific summary of classification accuracy. The odds of correction classification ratio for Class k, OCC_k, is given by

$$OCC_k = \frac{AvePP_k/(1 - AvePP_k)}{\hat{\pi}_k/(1 - \hat{\pi}_k)}, \quad (25)$$

where $\hat{\pi}_k$ is the model-estimated proportion for Class k. The denominator is the odds of correct classification based on random assignment using the model-estimated marginal class proportions, $\hat{\pi}_k$. The numerator is the odds of correct classification based on the maximum posterior class probability assignment rule (i.e., modal class assignment). When the modal class assignment for Class k is no better than chance, then $OCC_k = 1.00$. As $AvePP_k$ gets close to the ideal value of 1.00, OCC_k gets larger. Thus, large values of OCC_k (i.e., values 5.00 or larger; Nagin, 2005) for all K classes indicate a latent class model with good latent class separation and high assignment accuracy.

The final classification diagnostic presented here is the modal class assignment proportion (mcaP). This diagnostic is also a class-specific index of classification certainly. The modal class assignment proportion for Class k, $mcaP_k$, is given by

$$mcaP_k = \frac{\sum_{i=1}^{n} I\{\hat{c}_{\text{modal},i} = k\}}{n}, \quad (26)$$

Put simply, $mcaP_k$ is the proportion of individuals in the sample modally assigned to Class k. If individuals were assigned to Class k with perfect certainty, then $mcaP_k = \hat{\pi}_k$. Larger discrepancies between $mcaP_k$ and $\hat{\pi}_k$ are indicative of larger latent class assignment errors. To gage the discrepancy, each $mcaP_k$ can be compared to the to 95% confidence interval for the corresponding $\hat{\pi}_k$.

Class Enumeration. Now that you have a full set of tools for evaluating models in terms of absolute fit; relative fit, and classification accuracy, I can discuss how to apply them to the first critical step in the latent class modeling process: deciding on the number of latent classes. This process usually begins by specifying a one-class LCA model and then fitting additional models, incrementing the number of classes by one, until the models are no longer well identified (as defined in the subsection

"Model Estimation"). The fit of each of the models is evaluated in the absolute and relative terms. The parsimony principle is also applied such that the model with the fewest number of classes that is statistically and substantively adequate and useful is favored.

In terms of the relative fit comparisons, the standard likelihood ratio chi-square difference test presented earlier cannot be used in this setting, because the necessary regularity conditions of the test are violated when comparing a K-class model to a $(K-g)$-class model (McLachlan & Peel, 2000); in other words, although X^2_{diff} can be calculated, it does not have a chi-square sampling distribution under the null hypothesis. However, two alternatives, currently implemented in mainstream mixture modeling software, are available: (1) the adjusted Lo-Mendell-Rubin likelihood ratio test (adjusted LMR-LRT; Lo, Mendell, & Rubin, 2001), which analytically approximates the X^2_{diff} sampling distribution when comparing a K-class to a $(K-g)$-class finite mixture model for which the classes differ only in the mean structure; and (2) the parametric bootstrapped likelihood ratio test (BLRT), recommended by McLachlan and Peel (2000), which uses bootstrap samples (generated using parameter estimates from a $[K-g]$-class model) to empirically derive the sampling distribution of X^2_{diff} under the null model. Both of these tests and their performance across a range of finite mixture models has been explored in detail in the simulation study by Nylund, Asparouhov, and Muthén (2007). As executed in Mplus V6.1 (Muthén & Muthén, 1998–2011), these tests compare a $(K-1)$-class model (the null model) with a K-class model (the alternative, less restrictive model), and a statistically significant p-value suggests the K-class model fits the data significantly better than a model with one less class.

As mentioned before, there is no single method for comparing models with differing numbers of latent classes that is widely accepted as best (Muthén & Asparouhov, 2006; Nylund et al., 2007). However, by careful and systematic consideration of a set of plausible models, and utilizing a combination of statistical and substantive model checking (Muthén, 2003), researchers can improve their confidence in the tenability of their decision regarding the number of latent classes. I recommend the follow sequence for class enumeration, which is illustrated in detail with the empirical example that follows after the next subsection.

1. Fit a one-class model, recording the log likelihood value (LL); number of parameters estimated ($npar$); the likelihood ratio chi-square goodness-of-fit statistic (X^2_{LR} with df and corresponding p-value); and the model BIC, CAIC, and AWE values.

2. Fit a two-class model, recording the same quantities as listed in Step 1, along with: the adjusted LMR-LRT p-value, testing the two-class model against the null one-class model; the BLRT p-value, testing the two-class model against the null one-class model; and the approximate Bayes factor ($B\hat{F}_{1,2}$), estimating the ratio of the probability of the one-class model being the correct model to the probability of the two-class being the correct model.

3. Repeat the following for $K \geq 3$, increasing K by 1 at each repetition until the K-class model is not well identified:

Fit a K-class model, recording the same quantities as listed in Step 1, along with the adjusted LMR-LRT p-value, testing the K-class model against the null $(K-1)$-class model; the BLRT p-value, testing the K-class model against the null $(K-1)$-class model; and the approximate Bayes factor ($B\hat{F}_{K-1,K}$), estimating the ratio of the probability of the $(K-1)$-class model being the correct model to the probability of the K-class being the correct model.

4. Let $K_{\max}$ be the largest number of classes that could be extracted in a single model from Step 3. Compute the approximate correct model probability (cmP) across the one-class through $K_{\max}$-class models fit in Steps 1–3.

5. From the $K_{\max}$ models fit in Steps 1 through 3, select a smaller subset of two to three candidate models based on the absolute and relative fit indices using the guidelines (a)-(e) that follow. I assume here, since it is almost always the case in practice, that there will be more than one "best" model identified across the different indices. Typically, the candidate models are adjacent to each other with respect to the number of classes (e.g., three-class and four-class candidate models).

a. For absolute fit, the "best" model should be the model with the fewest number of classes that has an adequate overall goodness of fit—that is, the most parsimonious model that is not rejected by the exact fit test.

b. For the BIC, CAIC, and AWE, the "best" model is the model with the smallest value. However, because none of the information criteria are guaranteed to arrive at a single lowest value corresponding to a K-class model with $K < K_{max}$, these indices may have their smallest value at the K_{max}-class model. In such cases, you can explore the diminishing gains in model fit according to these indices with the use of "elbow" plots, similar to the use of scree plots of Eigen values used in exploratory factor analysis (EFA). For example, if you graph the BIC values versus the number of classes, then the addition of the second and third class may add much more information, but as the number of classes increases, the marginal gain may drop, resulting in a (hopefully) pronounced angle in the plot. The number of classes at this point meets the "elbow criterion" for that index.

c. For the adjusted LMR-LRT and BLRT, the "best" model is the model with the smallest number of classes that is *not* significantly improved by the addition of another class—that is, the most parsimonious K-class model that is not rejected in favor of a $(K + 1)$-class model. Note that the adjusted LMR-LRT and BLRT may never yield a non-significant p-value, favoring a K-class model over a $(K + 1)$-class model, before the number of classes reaches K_{max}. In these cases, you can examine a plot of the log likelihood values for an "elbow" as explained in Substep b.

d. For the approximate BF, the "best" model is the model with the smallest number of classes for which there is moderate to strong evidence compared to the next largest model—that is, the most parsimonious K-class model with a $B\hat{F} > 3$ when compared to a $(K + 1)$-class model.

e. For the approximate correct model probabilities, the "best" model is the model with the highest probability of being correct. Any model with $cm\hat{P}_K > .10$ could be considered a candidate model.

6. Examine the standardized residuals and the classification diagnostics (if germane for your application of mixture modeling) for the subset of candidate models selected in Step 5. Render an interpretation of each latent class in each of the candidate models and consider the collective substantive meaning of the resultant classes for each of the models. Ask yourself whether the resultant latent classes of one model help you to understand the phenomenon of interest (Magnusson, 1998) better than those of another. Weigh the simplicity and clarity of each of the candidate models (Bergman & Trost, 2006) and evaluate the utility of the additional classes for the less parsimonious of the candidate models. Compare the modal class assignments of individuals across the candidate models. Don't forget about label switching when you are making your model comparisons. And, beyond label switching, remember that if you estimate a K-class model and then a $(K + 1)$-class model, then there is no guarantee that any of the K classes from the K-class model match up in substance or in label to any of the classes in the $(K + 1)$-class model.

7. On the basis of all the comparisons made in Steps 5 and 6, select the final model in the class enumeration process.

If you have the good fortune of a very large sample, then the class enumeration process can be expanded and strengthened using a split-sample cross-validation procedure. In evaluating the "largeness" of your sample, keep in mind that sample size plays a critical role in the detection of what may be less prevalent classes in the population and in the selection between competing models with differing class structures (Lubke, 2010) and you don't want to split your sample for cross-validation if such a split compromises the quality and validity of the analyses within each of the subsamples because they are not of adequate size. For a split-sample cross-validation approach:

i. Randomly partition the full sample into two (approximately) equally sized subsamples: Subsample A (the "calibration" data set) and Subsample B (the "validation" data set).

ii. Conduct latent class enumeration Steps 1–7 on Subsample A.

iii. Retain all the model parameters estimates from the final K-class model selected in Step 7.

iv. Fit the K-class model to Subsample B, fixing all parameters to the estimated values retained in Step iii.

v. Evaluate the overall fit of the model. If the parameter estimates obtained from the K-class model fit to Subsample A, then provide an acceptable fit when used as fixed parameter values for a K-class model applied to Subsample B, then the model validates well and the selection of the

K-class model is supported (Collins, Graham, Long, & Hansen, 1994).

vi. Next fit a K-class model to Subsample B, allowing all parameters to be freely estimated.

vii. Using a nested-model likelihood ratio test, compare the fit of the K-class model applied to Subsample B using fixed parameter values based on the estimates from the Subsample A K-class model estimation to the fit of the K-class model applied to Subsample B with freely estimated parameters. If there is not a significant decrement in fit for the Subsample B K-class model when fixing parameter values to the Subsample A K-class model parameters estimates, then the model validates well, the nature and distribution of the K latent classes can be considered stable across the two subsamples, and the selection of the K-class model is supported.

There are variations on this cross-validation process that can be made. One variation is to carry out Steps iii through vii for all of the candidate models selected in Step 5 rather than just the final model selected in Step 7 and then integrate in Step 6 the additional information regarding which of the candidate models validated in Subsample B according to the results from both Steps v and vii. Another variation is to do a double (or twofold) cross-validation (Collins et al., 1994; Cudek & Browne, 1983) whereby Steps ii through vii are applied using Subsample A as the calibration data set and Subsample B as the validation data set and then are repeated using Subsample B as the calibration data set and Subsample A as the validation data set. Ideally, the same "best" model will emerge in both cross-validation iterations, although it is not guaranteed (Collins & Lanza, 2010). I illustrate the double cross-validation procedure in the empirical example that follows after the next subsection.

Missing Data

Because most mixture modeling software already utilizes a maximum likelihood estimation algorithm designed for ignorable missing data (primarily the EM algorithm), it is possible to accommodate missingness on the manifest indicators as well, as long as the missing data are either *missing completely at random* (MCAR) or *missing at random* (MAR). Assuming the data on the indicator variables are missing at random means the probability of a missing response for an individual on a given indicator is unrelated to the response that would have been observed, conditional on the individual's actual observed data for the other response variables. Estimation with the EM algorithm is a *full information maximum likelihood* (FIML) method in which individuals with complete data and partially complete data all contribute to the observed data likelihood function. The details of missing data analysis, including the multiple imputation alternative to FIML, is beyond the scope of this chapter. Interested readers are referred to Little and Rubin (2002), Schaefer (1997), and Enders (2010) for more information.

Of all the evaluations of model fit presented prior, the only one that is different in the presence of missing data is the likelihood ratio goodness-of-fit test. With partially complete data, the number of observed response patterns is increased to include observed response patterns with missingness. Returning to the five binary indicator hypothetical example, you might have some of the following incomplete response patterns:

u_1	u_2	u_3	u_4	u_5	f_r
1	1	1	1	1	f_1
1	1	1	1	0	f_2
1	1	1	1	•	f_3
⋮	⋮	⋮	⋮	⋮	⋮
1	0	0	0	0	f_{R^*-2}
0	0	0	0	0	f_{R^*-1}
•	0	0	0	0	f_{R^*}

where "•" indicates a missing response and $R^*$is the number of observed response patterns, *including* partially complete response patterns. The LR chi-square goodness-of-fit test statistic is now calculated as

$$X^2_{LR} = 2\sum_{r^*=1}^{R^*}\left[f_{r^*}\log\left(\frac{f_{r^*}}{\hat{f}_{r^*}}\right)\right], \qquad (27)$$

where $f_{r^*}$is the observed frequency count for the response pattern r^* and $\hat{f}_{r^*}$ is the model-estimated frequency count for the response pattern r^*. The degrees of freedom for the test is given by

$$df = R^* - d - 1. \qquad (28)$$

This test statistic, because it includes contributions from both complete and partially complete response patterns using model-estimated frequencies from a model estimated under the MAR

assumption, is actually a test of *both* the exact fit and the degree to which the data depart from MCAR against the MAR alternative (Collins & Lanza, 2010; Little & Rubin, 2002). Thus, the X^2_{LR} with missing data is inflated version of a simple test of only model goodness-of-fit. However, the X^2_{LR} is easily adjusted by subtracting the contribution to the chi-square from the MCAR component, and this adjusted X^2_{LR} can then be compared to the reference chi-square distribution (Collins & Lanza, 2010; Shafer, 1997). Note that the standardized residuals for partially complete response patterns are similarly inflated, and this should be considered when examining residuals for specific complete and partially complete response patterns.

The next subsection should be the most illuminating of all the subsections under Latent Class Analysis, as it is here that I fully illustrate the unconditional LCA modeling process with a real data example, show the use of all the fit indices, classification diagnostics, the double cross-validation procedures, and demonstrate the graphical presentation and substantive interpretation of a selected model.

Longitudinal Study of American Youth Example for Latent Class Analysis

The data used for the LCA example come from Cohort 2 of the Longitudinal Study of American Youth (LSAY), a national longitudinal study, funded by the National Science Foundation (NSF) (Miller, Kimmel, Hoffer, & Nelson, 2000). The LSAY was designed to investigate the development of student learning and achievement—particularly related to mathematics, science, and technology—and to examine the relationship of those student outcomes across middle and high school to postsecondary educational and early career choices. The students of Cohort 2 were first measured in the fall of 1988 when they were in eighth grade. Study participants were recruited through their schools, which were selected from a probability sample of U.S. public school districts (Kimmel & Miller, 2008). For simplicity's sake, I do not incorporate information related to the complex sampling design or the clustering of schools within districts and students within school for the modeling illustrations in this chapter; however, the analytic framework presented does extend to accommodate sampling weights and multilevel data. There were a total of $n = 3116$ students in the original LSAY Cohort 2 (48% female; 52% male).

For this example, nine items were selected from the eighth grade (Fall, 1998) student survey related to math attitudes for use as observed response indicators for an unconditional latent class variable that was intended to represent profiles of latent math dispositions. The nine self-report items were measured on a five-point, Likert-type scale (1 = strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 = strongly disagree). For the analysis, I dichotomized the items to a 0/1 scale after reverse coding certain items so that all item endorsements (indicated by a value of 1) represented pro-mathematics responses. Table 25.3 presents the original language of the survey prompt for the set of math attitude items along with the full text of the each item statement and the response categories from the original scale that were recoded as a pro-math item endorsements. In examining the items, I determined that the items could be tentatively grouped into three separate aspects of math disposition: items 1–3 are indicators of positive math affect and efficacy; items 4–5 are indicators of math anxiety; and items 6–9 are indicators of the student assessment of the utility of mathematics knowledge. I anticipated that this conceptual three-part formation of the items might assist in the interpretation of the resultant latent classes from the LCA modeling.

Table 25.3 also displays the frequencies and relative frequencies of pro-math item endorsements for the full analysis sample of $n = 2,675$ (excluding 441 of the total participant sample who had missing responses on *all* nine of selected items). Note that all nine items have a reasonable degree of variability in responses and therefore contain information about individual differences in math dispositions. If there were items with relative frequencies very near 0 or 1, there would be very little information about individual differences to inform the formation of the latent classes.

With nine binary response items, there are $2^9 = 512$ possible response pattern, but only 362 of those were observed in the sample data. Of the total sample, 2,464 participants (92%) have complete data on all the items. There are 166 observed response patterns in the data with at least one missing response. Of the total sample, 211 participants (8%) have missing data on one of more of the items, with 135 (64%) of those participants missing on only one item. Upon closer inspection, there is not any single item that stands out with a high frequency of missingness that might indicate a systematic skip pattern of responding that would make one reconsider that item's

Table 25.3. LSAY Example: Pro-math Item Endorsement Frequencies (f) and Relative Frequencies (rf) for the Total Sample and the Two Random Cross-Validation Subsamples, A and B

Survey prompt: "Now we would like you to tell us how you feel about math and science. Please indicate for you feel about each of the following statements."	Pro-math response categories*	Total sample (n_T= 2675)		Subsample A (n_A = 1338)		Subsample B (n_B = 1337)	
		f	*rf*	*f*	*rf*	*f*	*rf*
1) I enjoy math.	sa/a	1784	0.67	894	0.67	890	0.67
2) I am good at math.	sa/a	1850	0.69	912	0.68	938	0.70
3) I usually understand what we are doing in math.	sa/a	2020	0.76	1011	0.76	1009	0.76
4) Doing math often makes me nervous or upset.	d/sd	1546	0.59	765	0.59	781	0.59
5) I often get scared when I open my math book see a page of problems.	d/sd	1821	0.69	917	0.69	904	0.68
6) Math is useful in everyday problems.	sa/a	1835	0.70	908	0.69	927	0.70
7) Math helps a person think logically.	sa/a	1686	0.64	854	0.65	832	0.63
8) It is important to know math to get a good job.	sa/a	1947	0.74	975	0.74	972	0.74
9) I will use math in many ways as an adult.	sa/a	1858	0.70	932	0.70	926	0.70

*Original rating scale: 1 = strongly agree (sa); 2 = agree (a); 3 = not sure (ns); 4 = disagree (d); 5 = strongly disagree (sd).
Recoded to 0/1 with 1 indicating a pro-math response.

inclusion in the analysis. The three most frequent complete data response patterns with observed frequency counts are: $(1,1,1,1,1,1,1,1,1,1)$, $f = 502$; $(1,1,1,0,0,1,1,1,1)$, $f = 111$; and $(1,1,1,0,1,1,1,1,1)$, $f = 94$. More than 70% (258 of 362) of the complete data response patterns have $f < 5$. The three most frequent incomplete data response patterns with observed frequency counts are: $(1,1,1,?,1,1,1,1,1)$, $f = 9$; $(1,1,1,1,1,1,1,?,1)$, $f = 7$; and $(1,1,1,1,1,1,?,1,1)$, $f = 6$ (where "?" indicates a missing value).

Because this is a large sample, it is possible to utilize a double cross-validation procedure for establishing the unconditional latent class model for math dispositions. Beginning with Step i, the sample is randomly split into halves, Subsample A and Subsample B. Table 25.3 provides the item response frequencies and relative frequencies for both subsamples.

The class enumeration process begins by fitting 10 unconditional latent class models with $K = 1$ to $K = 10$ classes. After $K = 8$, the models ceased to be well identified (e.g., there was a high level of nonconvergence across the random sets of starting values; a low level of maximum log likelihood solution replication; a small condition number; and/or the smallest class proportion corresponded to less than 20 individuals). For $K = 1$ to $K = 8$, the models appeared well identified. For example, Figure 25.5 illustrates a high degree of replication of the "best" maximum likelihood value, −6250.94, for the five-class model, depicting the relative frequencies of the final stage log likelihood values at the local maxima across 1000 random sets of start values.

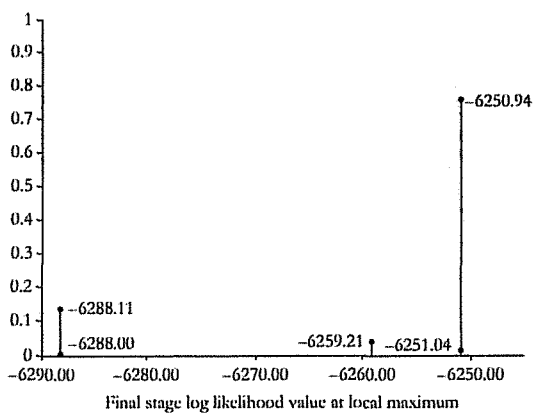

Figure 25.5 LSAY example: Relative frequency plot of final stage log likelihood values at local maxima across 1000 random sets of start values for the five-class unconditional LCA.

Table 25.4 summarizes the results from class enumeration Steps 1 through 5 for Subsample A. Bolded values indicate the value corresponding to the "best" model according to each fit index and the boxes indicate the candidate models based on each index (which include the "best" and the "second best" models). For the adjusted LR chi-square test of exact fit, the four-class model is marginally adequate and the five-class model has a high level of model-data consistency. Although the six-, seven-, and eight-class models also have a good fit to the data, the five-class model is the most parsimonious. The BIC has the smallest value for the five-class model but the six-class BIC value is very close. The same is true for the CAIC. The AWE has the smallest value for the four-class model, with the five-class value a close second. The four-class model is rejected in favor of the five-class model by the adjusted LMR-LRT, but the five-class model is not rejected in favor of the six-class model. All K-class model were rejected in favor of a $(K + 1)$-class model by the BLRT for all values of K considered so there was no "best" or even candidate models to be selected based on the BLRT, and those results are not presented in the summary table. According to the approximate BF, there was strong evidence for the five-class model over the four-class model, and there was strong evidence for the five-class model over the six-class model. Finally, based on the approximate correct model probabilities, of the eight models, the five-class model has the highest probability of being correct followed by the six-class model. Based on all these indices, I select the four-, five-, and six-class models for attempted cross-validation in Subsample B, noting that the five-class model is the one of the three candidate models most favored across all of the indices.

The first three rows of Table 25.5 summarize the cross-validation results for the Subsample A candidate models. For the first row, I took the four-class parameters estimates obtained by fitting a four-class model to Subsample A, used those estimates as fixed parameter values in Subsample B, and evaluated the overall fit of the model, following cross-validation Steps iv through v. The overall fit of the model, as determined by the LR chi-square goodness-of-fit test, was not adequate, and by this criterion, the estimated four-class model from Subsample A did not replicate in Subsample B. I next estimated a four-class model in Subsample B, allowing all parameters to be freely estimated, and compared the fit to the model with all the parameters fixed to the estimated values from Subsample A, following cross-validation Steps vi through vii. The likelihood ratio test of

Table 25.4. LSAY Example: Model Fit Indices for Exploratory Latent Class Analysis Using Calibration Subsample A (n_A=1338)

Model	*LL*	*npar*[*]	Adj. X^2_{LR} (df), *p*-value	BIC	CAIC	AWE	Adj. LMR-LRT *p*-value (H_0:K classes; H_1:K+1 classes)	$\hat{BF}_{K,K+1}$	$cm\hat{P}_K$
one-class	-7328.10	9	1289.21 (368), <.01	14721.00	14730.00	14812.79	<0.01	<0.10	<0.01
two-class	-6612.88	19	909.29 (358), <.01	13362.55	13381.55	13556.33	<0.01	<0.10	<0.01
three-class	-6432.53	29	551.76 (348), <.01	13073.83	13102.83	13369.60	<0.01	<0.10	<0.01
four-class	-6331.81	39	**347.24 (338), .35**	12944.38	12983.38	**13342.13**	<0.01	<0.10	<0.01
five-class	-6250.94	49	199.81 (328), >.99	**12854.63**	**12903.63**	13354.37	**0.15**	**6.26**	**0.87**
six-class	-6216.81	59	157.25 (318), >.99	12858.35	12917.35	13460.09	0.13	>10	0.13
seven-class	-6192.32	69	105.70 (308), >.99	12881.37	12950.37	13585.09	0.23	>10	<0.01
eight-class	-6171.11	79	69.55 (298), >.99	12910.93	12989.93	13716.64	-	-	<0.01
nine-class	Not well identified								
ten-class	Not well identified								

[*]number of parameters estimate

Table 25.5. LSAY Example: Double Cross-Validation Summary of Model Fit Using the Two Random Subsamples, A and B ($n_A = 1338$; $n_B = 1337$)

Model	Calibration	Validation	Adj. X^{2*}_{LR}	*df*	*p*-value	LRTS**	*df***	*p*-value
four-class	Subsample A	Subsample B	501.975	363	<0.001	38.50	39	0.49
five-class		353.036	363	0.64	59.71	49	0.14	
six-class		365.876	363	0.45	136.66	59	<0.001	
four-class	Subsample B	Subsample A	425.04	377	0.04	43.67	39	0.28
five-class		282.63	377	1.00	64.21	49	0.07	
six-class		260.37	377	1.00	101.85	59	<0.001	

*Goodness-of-fit of the model to validation subsample with all parameter values fixed at the estimates obtained from the calibration subsample.

** $LRTS = -2(LL_0 - LL_1)$ where LL_0 is the maximized log likelihood value, −6250.94, for the K-class model fit to the validation subsample with all parameter values fixed at the estimates obtained from the calibration subsample and LL_1 is the maximized log likelihood value for the K-class model fit to the validation subsample with all parameters freely estimated.

*** df = number of parameters in the K-class model

these nested models was not significant, indicating that the parameter estimates for the four-class model using Subsample B data were not significantly different from the parameter estimates from Subsample A. Thus, by this criterion, the estimated four-class model from Subsample A did replicate in Subsample B (indicated by bolded text in the table). The five-class model from Subsample A was the only one of the three candidate models that validated by both criteria (indicated by the boxed text).

For a double cross-validation, the full process above is repeated for Subsample B. I estimated $K = 1$ to $K = 10$ class models; selected a subset of candidate models, which were the same four-, five-, and six-class models as I selected for Subsample A; favoring the five-class model; and then cross-validated using Subsample A. As shown in Table 25.5, the five-class model from Subsample B was the only one of the three candidate models that cross-validated by both criteria in Subsample A.

Before the five-class model is anointed as the "final" unconditional model, there are a few more evaluations necessary. Although the five-class model is not rejected in the LR chi-square exact fit test, it is still advisable to examine the standardized residuals. Only six of the response patterns with model-estimated frequencies above 1.0 have standardized residuals greater than 3.0, only slightly more than the 1% one would expect by chance, and only one of those standardized residuals is greater than 5.0. Thus, closer examination of the model residuals does not raise concern about the fit of the five-class model to the data. Table 25.6 provides a summary of the observed and model-estimated frequencies for all observed response patterns with frequencies greater than 10 along with the standardized residual values.

Because I have approached this analysis as a direct application of mixture modeling, in that I am assuming *a priori* that the population is heterogeneous with regards to math dispositions and that the items selected for the analysis are indicators of membership in one of an unknown number of subgroups with characteristically different math disposition profiles, it is also necessary to examine the classification diagnostics for the five-class model as well as evaluate the substantive meaning and utility of the resultant classes. Table 25.7 summarizes the classification diagnostic measures for the five-class model with relative entropy of $E_5 = .77$. The modal class assignment proportions (mcaP) are all very near the estimated class proportions and well within the corresponding 95% (bias-corrected bootstrap) confidence intervals for $\hat{\pi}_k$, the AvePP are all greater than 0.70, and the odds of correct classification ratios are all well above 5.0, collectively indicating that the five classes are well separated and there is high accuracy in the latent class assignment. This result further endorses the choice of the five-class model.

The interpretation of the resultant five classes is based primarily on the model-estimated, class-specific item response probabilities provided in Table 25.8 and depicted graphically in the profile plot

Table 25.6. LSAY Example: Observed Response Patterns ($f > 10$), Observed and Estimated Frequencies, and Standardized Residuals for Subsample A with Estimated Posterior Class Probabilities and Modal Class Assignments Based on the Five-Class Unconditional LCA

Item+ response patterns (r^*)													$\hat{p}_{ik}$					
r^*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	f_{r^*}	$\hat{f}_{r^*}$	$stdr\hat{e}s_{r^*}$	$\hat{p}_{i1}$	$\hat{p}_{i2}$	$\hat{p}_{i3}$	$\hat{p}_{i4}$	$\hat{p}_{i5}$	$\hat{c}_{\text{modal (i)}}$
1	1	1	1	1	1	1	1	1	1	254.00	234.24	1.44	0.99	0.01	0.01	0.00	0.00	1
2	1	1	1	0	0	1	1	1	1	53.00	47.91	0.75	0.00	0.99	0.00	0.01	0.00	2
3	1	1	1	0	1	1	1	1	1	46.00	44.80	0.18	0.86	0.12	0.01	0.01	0.00	1
4	0	0	0	0	0	0	0	0	0	36.00	23.90	2.50	0.00	0.00	0.00	0.00	1.00	5
5	1	1	1	1	1	1	0	1	1	31.00	39.62	−1.39	0.93	0.01	0.06	0.00	0.00	1
6	0	1	1	1	1	1	1	1	1	26.00	29.00	−0.56	0.95	0.00	0.02	0.03	0.00	1
7	1	1	1	1	1	0	1	1	1	22.00	22.24	−0.05	0.85	0.02	0.13	0.00	0.00	1
8	1	1	1	1	0	1	1	1	1	19.00	16.91	0.51	0.00	0.97	0.02	0.01	0.00	2
9	1	1	1	1	1	0	0	0	0	18.00	10.54	2.31	0.00	0.00	0.99	0.00	0.01	3
10	1	1	1	1	1	1	1	1	0	17.00	18.12	−0.27	0.84	0.01	0.15	0.00	0.00	1
11	0	0	0	0	0	1	1	1	1	17.00	9.51	2.44	0.00	0.00	0.00	1.00	0.00	4
12	1	1	1	1	1	0	0	1	1	15.00	8.07	2.45	0.37	0.01	0.61	0.00	0.00	3
13	0	0	0	1	1	0	0	0	0	15.00	4.75	4.72	0.00	0.00	0.02	0.01	0.98	5
14	1	0	1	1	1	1	1	1	1	14.00	19.63	−1.28	0.93	0.01	0.01	0.05	0.00	1
15	0	0	1	1	1	1	1	1	1	14.00	5.87	3.36	0.37	0.00	0.02	0.61	0.00	4
16	1	1	1	1	1	1	1	0	1	13.00	14.87	−0.49	0.88	0.02	0.11	0.00	0.00	1
17	1	1	1	1	1	0	1	1	0	11.00	6.74	1.65	0.19	0.01	0.81	0.00	0.00	3

+(1) I enjoy math; (2) I am good at math; (3) I usually understand what we are doing in math; (4) Doing math often makes me nervous or upset; (5) I often get scared when I open my math book see a page of problems; (6) Math is useful in everyday problems; (7) Math helps a person think logically; (8) It is important to know math to get a good job; (9) I will use math in many ways as an adult. (~Reverse coded.)

Table 25.7. LSAY Example: Model Classification Diagnostics for the Five-Class Unconditional Latent Class Analysis ($E_5 = .77$) for Subsample A ($n_A = 1338$)

Class k	$\hat{\pi}_k$	95% C.I.*	$mcaP_k$	$AvePP_k$	OCC_k
Class 1	0.392	(0.326, 0.470)	0.400	0.905	14.78
Class 2	0.130	(0.082, 0.194)	0.125	0.874	46.42
Class 3	0.182	(0.098, 0.255)	0.176	0.791	17.01
Class 4	0.190	(0.139, 0.248)	0.189	0.833	21.26
Class 5	0.105	(0.080, 0.136)	0.109	0.874	59.13

*Bias-corrected bootstrap 95% confidence intervals

Table 25.8. LSAY Example: Model-Estimated, Class-Specific Item Response Probabilities Based on the Five-Class Unconditional Latent Class Analysis Using Subsample A ($n_A = 1338$)

		$\hat{\omega}_{m\|k}$				
Item aspects	Item statements	Class 1 (39%)	Class 2 (13%)	Class 3 (18%)	Class 4 (19%)	Class 5 (10%)
Math affect and math efficacy	I enjoy math.	0.89	0.99	0.72	0.21	0.18
	I am good at math.	0.93	0.91	0.84	0.17	0.14
	I usually understand what we are doing in math.	0.96	0.89	0.91	0.43	0.23
Math anxiety	~Doing math often makes me nervous or upset.	0.86	0.26	0.71	0.32	0.25
	~I often get scared when I open my math book see a page of problems.	1.00	0.10	0.82	0.52	0.37
Math utility	Math is useful in everyday problems.	0.92	0.85	0.33	0.77	0.09
	Math helps a person think logically.	0.86	0.83	0.37	0.67	0.06
	It is important to know math to get a good job.	0.95	0.89	0.47	0.83	0.11
	I will use math in many ways as an adult.	0.94	0.89	0.35	0.79	0.05

~ Reverse coded.

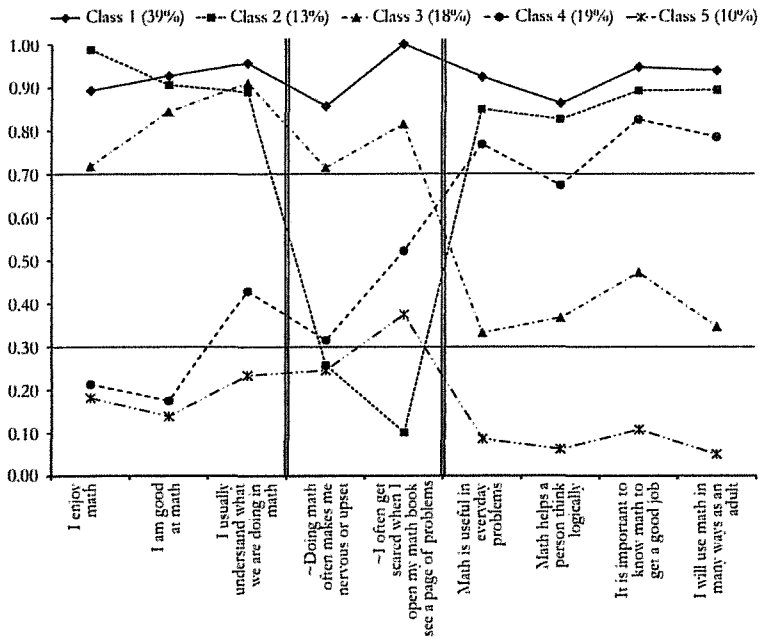

Figure 25.6 LSAY example: Model-estimated, class-specific item probability profile plot for the five-class unconditional LCA.

in Figure 25.6. Item response probabilities with a high degree of class homogeneity (i.e., estimated values greater than 0.7 or less than 0.3) are bolded in Table 25.8. All the items have high class homogeneity for at least three of the five classes, indicating that all nine items are useful for characterizing the latent classes. In Figure 25.6, the horizontal lines at the 0.7 and 0.3 endorsement probability levels help provide a visual guide for high levels of class homogeneity. These lines also help with the visual inspection of class separation with respect to each item—for example, two classes with item response probabilities above the 0.7 line for a given item are likely not well separated with respect to that item. Table 25.9 provides all the model-estimated item response odds ratios for each pairwise latent class comparison. Bolded values indicate the two classes being compared are well separated with respect to that set of items. The numbers in Table 25.9 correspond to visual impressions based on Figure 25.6; for example, Class 1 and Class 5 both have high homogeneity with respect to items 1 through 3 and appear to be well separated as confirmed with very large item response odds ratios (all in great excess of 5.0).

Tables 25.8 and 25.9 along with Figure 25.6 also distinguish the observed items by their affiliation with one of three substantive aspects of math disposition previously discussed. As can be seen in both the tables and figure, the class-specific item probabilities are similar in level of class homogeneity within each of these three aspects as are the pattern of class separation—that is, most pairs of classes are either well separated with respect to all or none of the items within an aspect group. Thus, as anticipated earlier, these three aspects can be used to refine the substantive interpretation of the five classes rather than characterizing the classes item by item. In attaching substantive meaning to the classes, I take into account both class homogeneity and class separation with respect to all the items. It is also useful to return to the actual observed response patterns in the data to identify *prototypical* response patterns for each the classes. Prototypical patterns should have reasonably sized observed frequencies, non-significant standardized residuals, and an estimated posterior probability near 1.0 for the class to which an individual with that response pattern would be modally assigned. I identify prototypical patterns for each of the five classes using the information provided in Table 25.6; some prototypical responses are boxed by solid lines in the table.

Class 1, with an estimated proportion of 39%, is characterized by an overall positive math disposition, with high probabilities of endorsing positive math affect and efficacy items, positive math anxiety items (indicating a low propensity for math anxiety), and positive math utility items. Class 1 has a high

Table 25.9. LSAY Example: Model-Estimated Item Response Odds Ratios for All Pairwise Latent Class Comparisons Based on the Five-Class Unconditional Latent Class Analysis Using Subsample A ($n_A = 1338$)

		$\hat{OR}_{m\|jk}$									
Item aspects	Item statements	Class 1 vs. 2	Class 1 vs. 3	Class 1 vs. 4	Class 1 vs. 5	Class 2 vs. 3	Class 2 vs. 4	Class 2 vs. 5	Class 3 vs. 4	Class 3 vs. 5	Class 4 vs. 5
Math affect and math efficacy	I enjoy math.	0.11	3.28	30.91	37.83	30.72	>100	>100	9.42	11.53	1.22
	I am good at math.	1.31	2.34	59.92	78.96	1.78	45.60	60.10	25.61	33.75	1.32
	I usually understand what we are doing in math.	2.70	2.15	28.99	71.52	0.80	10.75	26.52	13.49	33.28	2.47
Math anxiety	~Doing math often makes me nervous or upset.	17.32	2.39	13.03	18.47	0.14	0.75	1.07	5.45	7.72	1.42
	~I often get scared when I open my math book see a page of problems.	>100	>100	>100	>100	0.03	0.10	0.19	4.03	7.36	1.82
Math utility	Math is useful in everyday problems.	2.16	24.48	3.67	>100	11.36	1.70	60.04	0.15	5.29	35.30
	Math helps a person think logically.	1.32	10.85	3.05	>100	8.19	2.30	71.66	0.28	8.75	31.16
	It is important to know math to get a good job.	2.13	19.83	3.74	>100	9.29	1.75	68.99	0.19	7.43	39.33
	I will use math in many ways as an adult.	1.81	28.79	4.17	>100	15.91	2.30	>100	0.14	9.99	69.06

~Reverse coded.

level of homogeneity with respect to all the items. This class might be labeled the "Pro-math without anxiety" class, where "pro-math" implies both liking and valuing the utility of mathematics. Response pattern 1 in Table 25.6 is a prototypical response pattern for Class 1, with individuals endorsing all nine items.

Class 5, with an estimated proportion of 10%, is characterized by an overall negative math disposition, with low probabilities of endorsing positive math affect and efficacy items, positive math anxiety items (indicating a high propensity for math anxiety), and positive math utility items. Class 5 has a high level of homogeneity with respect to all the items and is extremely well separated from Class 1 with respect to all the items. This class might be labeled the "Anti-math with anxiety" class, where "anti-math" implies both disliking and undervaluing the utility of mathematics. Response pattern 4 in Table 25.6 is a prototypical response pattern for Class 5, with individuals endorsing none of the nine items.

Because Classes 1 and 5 represent clear profiles of positive and negative math dispositions across the entire set of items with high levels of class homogeneity across all the items (with the exception of item 5 in Class 5) and are well separated from each other with respect to all items (with item response odds ratios all well in excess of 5.0), the class separation of the remaining three classes will be evaluated primarily with respect to Classes 1 and 5.

Class 2, with an estimated proportion of 13%, is characterized by an overall positive math disposition like Class 1, with the exception that this class has very low probabilities of endorsing positive math anxiety items (indicating a high propensity for math anxiety). Class 2 has a high level of homogeneity with respect to all the items, is well separated from Class 1 with respect to the math anxiety items but not the math affect and efficacy or the math utility items (with the exception of item 1), and is well separated from Class 5 with respect to the math affect and efficacy and the math utility items. This class might be labeled the "Pro-math with anxiety" class. Response pattern 2 in Table 25.6 is a prototypical response pattern for Class 2, with individuals endorsing all but the two math anxiety items.

Class 3, with an estimated proportion of 18%, is characterized by high probabilities of endorsing positive math affect and efficacy items and positive math anxiety items (indicating a low propensity for math anxiety). Class 3 does not have a high level of homogeneity with respect to the math utility items which means that this class is *not* characterized by either high or low response propensities. However, Class 3 is well separated from Class 1 and Class 5 with respect to those items. Generally speaking, Class 3 is not well separated from Class 1 with respect to the math affect and efficacy and the math anxiety items but is well separated from Class 5 with respect to those same items. This class might be labeled the "Math lover" class, where "love" implies both a positive math affect and a low propensity for math anxiety. Response pattern 9 in Table 25.6 is a prototypical response pattern for Class 3, with individuals endorsing all but the math utility items.

Class 4, with an estimated proportion of 19%, is mostly characterized by low probabilities of endorsing positive math affect and efficacy items and high probabilities of endorsing positive math utility items. Class 4 does not have a high level of homogeneity with respect to the math anxiety items, which means that this class is not characterized by either high or low response probabilities. It is well separated from Class 1 with respect to the math anxiety items as well as the math affect and efficacy item but not well separated from Class 5 for those same items. Class 4 is well separated from Class 5 with respect to the math utility item but not well separated from Class 1. This class might be labeled the "I don't like math but I know it's good for me" class. Response pattern 11 in Table 25.6 is a prototypical response pattern for Class 4, with individuals endorsing only the math utility items.

None of the five resultant classes have an estimated class proportion corresponding to a majority share of the overall population nor are any of the classes distinguished from the rest by a relatively small proportion. Thus, although it is quite interesting that the "Pro-math without anxiety" class is the largest at 40%, and the "Anti-math with anxiety" class is the smallest at 10%, the estimated class proportions themselves, in this case, did not contribute directly to the interpretation of the classes.

As a final piece of the interpretation process, I also examine response patterns that are not well fit and/or not well classified by the selected model. These patterns could suggest additional population heterogeneity that does not have a strong "signal" in the present data and is not captured by the resultant latent classes. Noticing patterns that are not well fit or well classified by the model can deepen understanding of the latent classes that do emerge and

may also suggest directions for future research, particularly regarding enhancing the item set. Enclosed by a dashed box in Table 25.6, response pattern 13 has a large standardized residual and is not well fit by the model. Although individuals with this response pattern have a high posterior probability for Class 5, their pattern of response, only endorsing the math anxiety items, is not prototypical of any of the classes. These cases are individuals who have a low propensity toward math anxiety but are inclined to dislike and undervalue mathematics. They don't like math but are "fearless." These individuals could represent just a few random outliers or they could be indicative of a smaller class that is not detected in this model but is one that might emerge in a future study with a larger sample and with an expanded item set. Individuals with response pattern 15 in Table 25.6 are also not well classified. Although individuals with this response pattern would be modally assigned to Class 4, the estimated posterior probability for Class 4 is only 0.61 while the estimated posterior probability for Class 1 is 0.37. These individuals, endorsing all but the first two math affect and efficacy items, although more consistent with the Class 4 profile, are very similar to individuals with response patterns such as pattern 14 in Table 25.6, that endorse all but one of the math affect and efficacy items and have a high estimated posterior probability for Class 1. Not surprisingly, it is harder to classify response patterns to classes without a high degree of homogeneity on the full set of items, such as Classes 3 and 4, as is evident from the relative lower AvePPs found in Table 25.7 for Classes 3 and 4 compared to Classes 1, 2, and 5.

Concluding now the full empirical illustration of latent class analysis, I switch gears to introduce traditional finite mixture modeling, also known as LPA (the moniker used herein) and LCCA.

Latent Profile Analysis

Essentially, a latent profile model is simply a latent class model with continuous—rather than categorical—indicators of the latent class variable. Almost everything learned in the previous section on LCA can be applied to LPA, but there are a few differences—conceptual, analytic, and practical—that must be remarked on before proceeding to the real data example of LPA. This section follows the same order of topics as the section on LCA, beginning with LPA model formulation.

Model Formulation

I begin the formal LPA model specification with an unconditional model in which the only observed variables are the continuous manifest variables of the latent class variable. This model is the unconditional measurement model for the latent class variable.

Suppose there are M continuous (interval scale) latent class indicators, $y_1, y_2, \ldots, y_M$, observed on n study participants, where y_{mi} is the observed response to item m for participant i. It is assumed for the unconditional LPA that there is an underlying unordered categorical latent class variable, denoted by c, with K classes, where $c_i = k$ if individual i belongs to Class k. As before, the proportion of individuals in Class k, $\Pr(c = k)$, is denoted by π_k. The K classes are exhaustive and mutually exclusive such that each individual in the population has membership in exactly one of the K latent classes and $\Sigma\pi_k = 1$. The relationship between the observed responses on the M items and the latent class variable, c, is expressed as

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} [\pi_k \cdot f_k(\mathbf{y}_i)], \tag{29}$$

where $\mathbf{y}_i = (y_{1i}, y_{1i}, \ldots, y_{Mi})$, $f(\mathbf{y}_i)$ is the multivariate probability density function for the overall population, and $f_k(\mathbf{y}_i) = f(\mathbf{y}_i | c_i = k)$ is the class-specific density function for Class k. Thus, the LPA measurement model specifies that the overall joint distribution of the M continuous indicators is the result of a mixing of K component distributions of the M indicators, with $f_k(\mathbf{y}_i)$ representing the component-specific joint distribution for $\mathbf{y}_i$.

As with the LCA model, the structural parameters are those related to the distribution of the latent class variable, which for the unconditional LPA model are simply the class proportions, π_k. The measurement parameters are all those related to the class-specific probability distributions. Usually, as was done in the very first finite mixture model applications, the within-class distribution of the continuous indicator variables is assumed to be multivariate normal. That is,

$$[\mathbf{y}_i | c_i = k] \sim \text{MVN}(\alpha_k, \Sigma_k), \tag{30}$$

where α_k is the vector of the Class k means for the ys (i.e., $\text{E}(\mathbf{y}_{i|k}) = \alpha_k$) and Σ_k is the Class k variance–covariance matrix for the ys (i.e., $\text{Var}(\mathbf{y}_{i|k}) = \Sigma_k$). Alternatively, the expression in Equation 30 can be written as

$$\begin{aligned} \mathbf{y}_{i|k} &= \alpha_k + \varepsilon_{ik}, \\ \varepsilon_{ik} &\sim \text{MVN}(\mathbf{0}, \Sigma_k). \end{aligned} \tag{31}$$

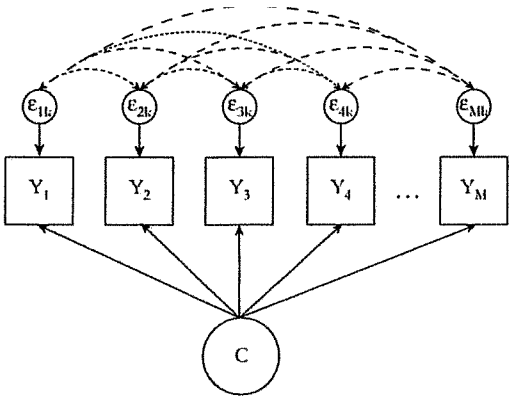

Figure 25.7 Generic path diagram for an unconditional latent profile model.

The measurement parameters *are* then the class-specific means, variances, and covariances of the indicator variables. Notice that although one necessarily assumes a particular parametric distribution *within* each class that is appropriate for the measurement scales of the variables, there are not any assumptions made about the joint distribution of the indicators in the overall population.

The model expressed in Equations 29 and 30 can be represented by a path diagram as shown in Figure 25.7. If you compare Figure 25.7 to Figure 25.2, along with replacing the u with y to represent continuous rather than categorical manifest variables, "residuals" terms have been added, represented by the ε indexed by k, to indicate that there is within-class variability on the continuous indicators that may differ across the classes in addition to the mean structure of the y that may vary across the classes as indicated by the arrows from c directly to the y. Unlike with categorical indicators, the class-specific estimated means and variances/covariances (assuming normality within class) and can be uniquely identified for each class.

Traditionally, the means of the y are automatically allowed to vary across the classes as part of the measurement model—that is, the mean structure is always class-varying. The within-class variances may be class-varying or constrained to be class-invariant (i.e., within-class variances held equal across the classes). And, as implied by Figure 25.7, the conditional independence assumption is not necessary for the within-class covariance structure. Unlike LCA, latent profile models do not require partial conditional independence for model identification—all indicators can covary with all other indicators within class. Hence, the latent class variable does not have to be specified to explain all of the covariation between the indicators in the overall population.

With increased flexibility in the within-class model specification comes additional complexity in the model-building process. But before getting into the details of model building for latent profiles models, let me formally summarize the main within-class variance–covariance structures that may be specified for Σ_k (presuming here that α_k will be left unconstrained within and across the classes in all cases). Starting from the least restrictive of variance–covariance structures, there is *class-varying, unrestricted* Σ_k of the form

$$\Sigma_k = \begin{bmatrix} \theta_{11k} & & & \\ \theta_{21k} & \theta_{22k} & & \\ \vdots & \vdots & \ddots & \\ \theta_{M1k} & \theta_{M2k} & \cdots & \theta_{MMk} \end{bmatrix}, \quad (32)$$

where θ_{mmk} is the variance of item m in Class k and θ_{mjk} is the covariance between items m and j in Class k. In this structure for Σ_k, all the indicator variables are allowed to covary within class, and the variances and covariances are allowed to be different across the latent classes. The *class-invariant, unrestricted* Σ_k has the form

$$\Sigma_k = \Sigma = \begin{bmatrix} \theta_{11} & & & \\ \theta_{21} & \theta_{22} & & \\ \vdots & \vdots & \ddots & \\ \theta_{M1} & \theta_{M2} & \cdots & \theta_{MM} \end{bmatrix}, \quad \forall k \in (1, \ldots, K), \quad (33)$$

such that all the indicator variable are allowed to covary within class, and the variances and covariances are constrained to be equal across the latent classes (class-invariant). The *class-varying, diagonal* Σ_k has the form

$$\Sigma_k = \begin{bmatrix} \theta_{11k} & & & \\ 0 & \theta_{22k} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \theta_{MMk} \end{bmatrix}, \quad (34)$$

such that conditional independence is imposed and the covariances between the indicators are fixed at zero within class while the variances are freely estimated and allowed to be different across the latent classes. The most constrained within-class variance–covariance structure is the *class-invariant, diagonal*

Σ_k with the form

$$\Sigma_k = \Sigma = \begin{bmatrix} \theta_{11} & & & \\ 0 & \theta_{22} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \theta_{MM} \end{bmatrix},$$
$$\forall k \in (1, \ldots, K), \qquad (35)$$

such that conditional independence is imposed and the covariances between the indicators are fixed at zero within class while the variances are constrained to be equal across the latent classes.

The determination of the number of latent classes as well as the estimates of the structural parameters (class proportions) and the measurement parameters (class-specific means, variances, and covariances) and interpretation of the resultant classes will very much depend on the specification of the within-class joint distribution of the latent class indicators. This dependence is analogous to the dependence of clustering on the selection of the attribute space and the resemblance coefficient in a cluster analysis. As it happens, specifying a class-invariant, diagonal Σ_k in a K-class LPA model will yield a solution that is the model-based equivalent to applying a *K-means* clustering algorithm to the latent profile indicators (Vermunt & Magidson, 2002).

To better understand how the number and nature of the latent classes can be influenced by the specification of Σ_k, let's consider a hypothetical data sample drawn from an unknown but distinctly non-normal bivariate population distribution. The scatter plot for the sample observations is displayed in Figure 25.8.a. Figure 25.8.b shows a path diagram for a three-class latent profile model with a class-invariant, diagonal Σ_k along with the empirical results of applying the three-class LPA model to the sample data depicted as a scatter plot with: individual observations marked with symbols corresponding to modal assignment into one of the three latent classes (circles, x, and triangles); diamonds representing the class centroids, $(\alpha_{1k}, \alpha_{2k})$—that is, the model-estimated, class-specific means for y_1 and y_2; trend lines representing the class-specific linear associations for y_2 versus y_1; and ellipses to provide a visual impression of the model-estimated, class-specific variances for y_1 and y_2, where the width of each ellipse is equal to three model-estimated, class-specific standard deviations for y_1 and the height is equal to three model-estimated, class-specific standard deviations for y_2. The model in Figure 25.8.b imposes the conditional independence assumption, and thus, y_1 and y_2 are uncorrelated within class, shown by the flat trend lines for each of the three classes. The model also constrains the within-class

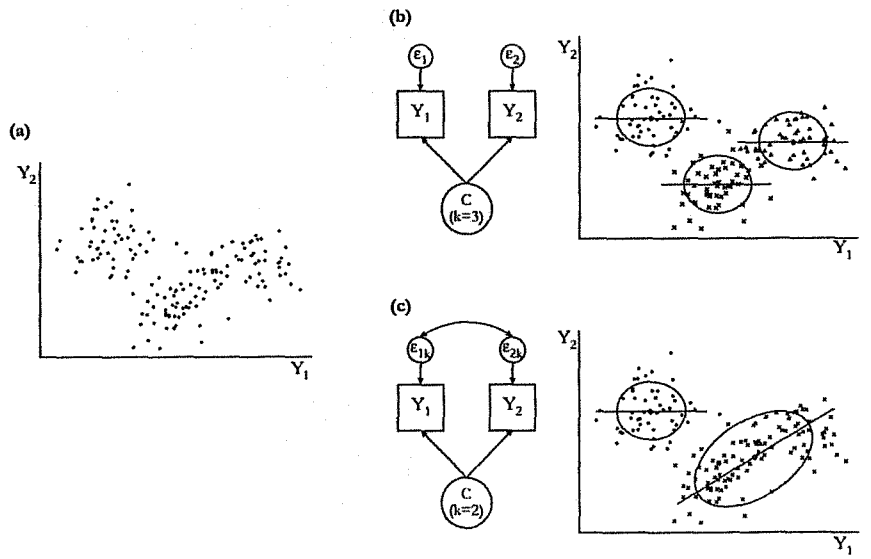

Figure 25.8 (a) Bivariate scatterplot based on a hypothetical sample from an overall bivariate non-normal population distribution; (b) Path diagram for a three-class model with class-invariant, diagonal Σ_k and the scatter plot of sample values marked by modal latent class assignment based on the three-class model; and (c) Path diagram for a two-class model with class-varying, unrestricted Σ_k and the scatter plot of sample values marked by modal latent class assignment based on the two-class model. In (b) and (c), diamonds represent the model-estimated class-specific bivariate mean values, trend lines depict the model-estimated within-class bivariate associations, and the ellipse heights and widths correspond to 3.0 model-estimated within-class standard deviations on y_2 and y_1, respectively.

variance–covariance structure to be the same across the class, shown by the same size ellipses for each of the three classes.

Figure 25.8.c displays a path diagram for a two-class latent profile model with a class-varying, unconstrained Σ_k along with the empirical results of applying the two-class LPA model to the sample data depicted as a scatter plot using the same conventions as Figure 25.8.b. The results of these two models, shown in Figures 25.8.b and 25.8.c, applied to the same sample data shown in Figure 25.8.a are different both in the number and nature of the latent classes. They provide alternative representations of the population heterogeneity with respect to the latent class continuous indicators, y_1 and y_2. And they would lead to quite different substantive interpretations. You could make comparisons of fit between the two models to determine whether one is more consistent with the observed data, but if they both provide adequate fit and/or are comparable in fit to each other, then you must rely on theoretical and practical considerations to choose one representation over the other. Because you don't ever know the "true" within-class variance–covariance structure just as you don't ever know the "correct" number of latent classes when you embark on a latent profile analysis, and now understanding how profoundly the specification of Σ_k could influence the formation of the latent classes, the LPA model-building process must compare models, statistically and substantively, across a full range of Σ_k specifications.

Model Interpretation

If you were engaged in an indirect application of finite mixture modeling to obtain a semi-parametric approximation for an overall non-normal homogeneous population, then you would focus on the "remixed" results for the overall population and would not be concerned with the distinctiveness or separation of the latent classes and would not interpret the separate mixture components. However, if you are using a latent profile analysis in a direct application, assuming *a priori* that the population is made up of two or more normal homogeneous subpopulations, then you would place high value on results that yield classes that are disparate enough from each other that it is reasonable to interpret each class as representative of a distinct subpopulation.

In some sense, the direct application of finite mixture modeling is a kind of stochastic model-based clustering method in which one endeavors to arrive at a latent class solution with the number and nature of latent classes (clusters) such that the individual variability with respect to the indicator variables within the classes is minimized and/or the between-class variability is maximized. (For more on mixture modeling as a clustering method and comparison to other clustering techniques, *see* Vermunt & Magidson, 2002, and the chapter on clustering within this handbook.) These clustering objectives can be restated in the terms used when presenting the interpretation of latent class models: For distinct and optimally interpretable latent classes, it is desirable to have a latent profile model with a high degree of class homogeneity (low within-class variability) along with a high degree of class separation (high between-class variability).

Just as was done with LCA, the concepts of latent class homogeneity and latent class separation and how they both relate to the parameters of the unconditional measurement model will be discussed as well as how they inform the interpretation of the latent classes resulting from a LPA. To assist this discussion, consider a hypothetical example with two continuous indicators ($M = 2$) measuring a two-class categorical latent variable ($K = 2$). And suppose that you decide to use a class-varying, unrestricted Σ_k specification for the LPA. The unconditional model is given by

$$f(y_{1i}, y_{2i}) = \sum_{k=1}^{2} \left[\pi_k \cdot f_k(y_{1i}, y_{2i})\right], \qquad (36)$$

where

$$[y_{1i}, y_{2i} | c_i = k] \sim \text{MVN}\left(\alpha_k = \begin{bmatrix} \alpha_{1k} & \alpha_{2k} \end{bmatrix}, \Sigma_k = \begin{bmatrix} \theta_{11k} & \\ \theta_{21} & \theta_{22k} \end{bmatrix}\right). \qquad (37)$$

Class Homogeneity. The first and primary way that you can evaluate the degree of class homogeneity is by examining the model-estimated within-class variances, $\hat{\theta}_{mmk}$, for each indicator m across the K classes and comparing them to the total overall sample variance, $\hat{\theta}_{mm}$, for the continuous indicator. It is expected that all of the within-class variances will be notably smaller than the overall variance. Classes with smaller values of $\hat{\theta}_{mmk}$ are more homogeneous with respect to item m than classes with larger values of $\hat{\theta}_{mmk}$. You can equivalently compare within-class standard deviations, $\sqrt{\hat{\theta}_{mmk}}$, for each item m across the K classes, that approximate for each class the average distance of class members' individual values on item m to the corresponding model-estimated

class mean, $\hat{\alpha}_{mk}$. You want classes for which class members are close, on average, to the class-specific mean because you want to be able to use the class mean values in your interpretation of the latent classes as values that "typify" the observed responses on the indicator variables for members of that class.

You cannot, of course, directly compare values of $\hat{\theta}_{mmk}$ across items because different items may have very different scales and the magnitude of the variance (and, hence, the standard deviation) is scale-dependent. Even for items with the same measurement scales, you cannot compare within-class variances across items unless the overall variances of those items are comparable. However, it is possible to summarize class homogeneity across items and classes by calculating the percent of the overall total variance in the indicator set explained by the latent class variable, similarly to the calculations done in a principal component analysis (Thorndike, 1953).

The phrase "class homogeneity" refers here to an expectation that the individuals belonging to the same class will be more similar to each other with respect to their values on the indicator variables than they are to individuals in other classes. However, you should still keep in mind that a LPA assumes *a priori* that the classes *are* homogeneous in the sense that all members of a given class are assumed to draw from a single, usually multivariate normal, population distribution. And, as such, any within-class correlation between continuous indicators, if estimated, is assumed to be an association between those variables that holds for all members of that class. Evaluating the statistical and practical significance of an estimated within-class indicator correlation, if not fixed at zero in the model specification, can assist in judging whether that correlation could be used in the characterization of the subpopulation represented by that particular latent class. Significant within-class correlations, when present, may be as much a part of what distinguishes the classes as the class-specific means and variances.

Class Separation. The first and primary way you can evaluate the degree of class separation is by assessing the actual distance between the class-specific means. It is not enough to simply calculate the raw differences in estimated means (i.e., $\hat{\alpha}_{mj} - \hat{\alpha}_{mk}$) because what is most relevant is the degree of overlap between the class-specific distributions. And the degree of overlap between two normal distributions depends not only on the distance between the means but the variances of the distributions as well. Consider, for example, the two scenarios shown in

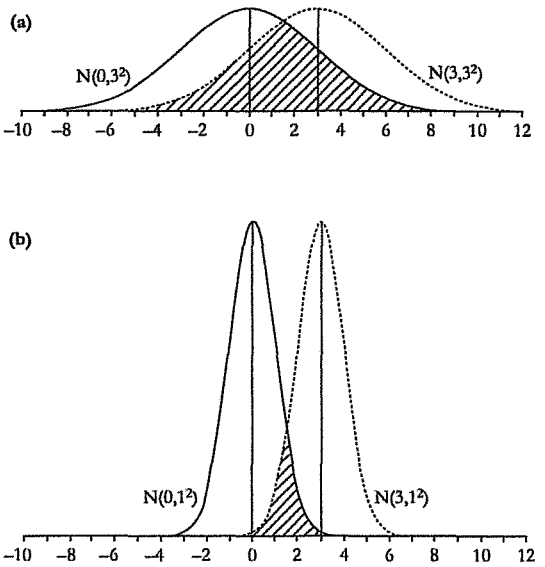

Figure 25.9 Hypothetical finite mixture distributions for a single continuous indicator variable with K = 2 underlying latent classes with class-specific means of 0 and 3, respectively, and with-class standard deviations of (a) 3, and (b) 1.

Figure 25.9. Figure 25.9.a depicts two hypothetical class-specific indicator distributions with means 3.0 units apart and class-specific standard deviations of 3, and Figure 25.9.b depicts two hypothetical class-specific indicator distributions with the same mean separation as in Figure 25.9.a with class-specific standard deviations of 1. There is considerable overlap of the distributions in Figure 25.9.a and very little overlap in Figure 25.9.b. The two classes in Figure 25.9.b are far better separated than the two classes in Figure 25.9.a with respect to the indicator, although the difference in means is the same. To quantify class separation between Class j and Class k with respect to a particular item m, compute a standardized mean difference, adapting the formula for Cohen's d (Cohen, 1988), as given below,

$$\hat{d}_{mjk} = \frac{\hat{\alpha}_{mj} - \hat{\alpha}_{mk}}{\hat{\sigma}_{mjk}}, \tag{38}$$

where $\hat{\sigma}_{mjk}$ is a pooled standard deviation given by

$$\hat{\sigma}_{mjk} = \sqrt{\frac{\left(\hat{\pi}_j\right)\left(\hat{\theta}_{mmj}\right) + \left(\hat{\pi}_k\right)\left(\hat{\theta}_{mmk}\right)}{\left(\hat{\pi}_j + \hat{\pi}_k\right)}}. \tag{39}$$

A large $\left|\hat{d}_{mjk}\right| > 2.0$ corresponds to less than 20% overlap in the distributions, meaning that less than 20% of individuals belonging to either Class j or Class k have values on item m that fall in

the range of y_m corresponding to the area of overlap between the two class-specific distributions of y_m. A large $|\hat{d}_{mjk}|$ indicates a high degree of separation between the Classes j and k with respect to item m. A small $|\hat{d}_{mjk}| < 0.85$ corresponds to more than 50% overlap and a low degree of separation between the Classes j and k with respect to item m.

If you are using a latent profile model specification that allows a class-varying variance–covariance structure for the classes, then you can also evaluate whether the classes are distinct from each other with respect to the item variances or covariances. To make a descriptive assessment of the separation of the classes in this regard, you can examine whether there is any overlap in the 95% confidence intervals for the estimates of the class-specific variances and covariances with non-overlap indicating good separation. An equivalent assessment can be made using the model-estimated class-specific item standard deviations and correlations.

Class Proportions. The guidelines and cautions provided in the section on LCA for the use of the estimated class proportions in the interpretation of the latent classes are all applicable for LPA as well.

Hypothetical Example. Continuing with the hypothetical example of a two-class LPA with two continuous indicators initially presented in Figure 25.8.c, Table 25.10 provides the overall sample means and standard deviations along with the model-estimated class-specific means, standard deviations, and correlations (with the standard deviations and correlation estimates calculated using the measurement parameter estimates for the class-specific item variances and covariances). The class-specific standard deviations for the items y_1 and y_2 are all noticeably smaller than the corresponding overall sample standard deviations, but Class 1 is more homogenous than Class 2 with respect to both indicators—particularly y_1. There is a small, non-significant correlation between y_1 and y_2 in Class 1 but a large and significant positive correlation between y_1 and y_2 in Class 2 that should therefore be considered in the interpretation of Class 2.

Applying Equations 38 and 39 to the class-specific mean and standard deviation estimates given in Table 25.10, the standardized differences in indicator means between Class 1 and Class 2 was calculated as $\hat{d}_1 = -2.67$, indicating a high degree of separation with respect to y_1, and $\hat{d}_2 = 1.70$, indicating a moderate degree of separation with respect to y_2. In terms of the class-specific variance–covariance structures, evaluate the separation between the classes

Table 25.10. Hypothetical Example: Overall Sample Means and Standard Deviations (SD); Model-Estimated, Class-Specific Means, Standard Deviations, and Correlations With Corresponding Bias-Corrected Bootstrap 95% Confidence Intervals Based on a Two-Class Latent Profile Analysis with Class-Varying, Unrestricted Σ_k

				Correlations	
	Variable	Mean	SD	(1)	(2)
Overall sample	y_1	0.06	2.71	1.00	
	y_2	1.47	1.70	–.21	1.00
Class	Variable	Mean ($\hat{\alpha}_{mk}$)	SD ($\sqrt{\hat{\theta}_{mmk}}$)	Correlations	
				(1)	(2)
Class 1 (33%)	y_1	–2.93	1.00	1.00	
		(–3.19, –2.58)	(0.80, 1.26)		
	y_2	2.97	1.12	0.04	1.00
		(2.65, 3.35)	(0.95, 1.38)	(–0.25, 0.26)	
Class 2 (77%)	y_1	1.55	1.93	1.00	
		(1.18, 1.93)	(1.77, 2.11)		
	y_2	0.73	1.41	0.68	1.00
		(0.45, 0.99))	(1.26, 1.59)	(0.54, 0.76)	

with respect to the within-class item standard deviations and correlations by examining the differences in the point estimates and also observing the presence of overlap in the 95% confidence intervals for the point estimates. Note that the 95% confidence intervals provided in Table 25.10 are estimated using a bias-corrected bootstrap technique because the sampling distributions for standard deviations are not symmetric and estimated correlations are nonlinear functions of three different maximum likelihood parameter estimates. The variability in Class 2 for y_1 is notably larger than Class 1, whereas the classes are not well separated with respect to the standard deviations for y_2. The correlation between y_1 and y_2 is very different for Classes 1 and 2 where there is virtually no correlation at all in Class 1 but there is a large and significant correlation within Class 2. Thus, there is a high degree of separation between Classes 1 and 2 with respect to the relationship between y_1 and y_2.

The class homogeneity and separation information contained in Table 25.10 is not always, but can be, depicted graphically in a series of bivariate scatter plots, particularly when the total number of latent class indicator variables is small. In this example, with only two items, a single bivariate plot is all that is needed. The estimated class-specific means are plotted and specially identified with data point markers different from the observed data points. All the observed data points are included in the plot and are marked according to their modal class assignment. A trend line is drawn through each class centroid derived from the model-estimated class-specific correlations between the two items. Ellipses are drawn, one centered around each class centroid, with the axis lengths of the ellipse corresponding to three standard deviations on the corresponding indicator variable. All of these plot features are displayed in Figure 25.8.c and help to provide a visual impression of all the aspects of the class-specific distributions that distinguish the classes (along with those aspects that don't) and the overall degree of class separation.

You can see visually in Figure 25.8.c what I have already remarked on using the information in Table 25.10: Class 1 (individual cases in the sample with modal class assignment to Class 1 have data points marked by circles) is more homogenous with respect to both y_1 and y_2—particularly y_1—than Class 2 (individual cases with modal class assignment to Class 2 have data points marked by x); there is a high degree of separation between Classes 1 and 2 with respect to values on y_1 and only moderate separation with respect to values on y_2; there is a strong positive association between y_1 and y_2 in Class 2 that is not present in Class 1. You could interpret Class 1 as a homogenous group of individuals with a low average level on y_1 ($\hat{\alpha}_{11} = -2.93$), relative to the overall sample mean, and a high average level of y_2 ($\hat{\alpha}_{21} = 2.97$). You could characterize Class 2 as a less homogeneous (relative to Class 1) group of individuals with a high average level on y_1 ($\hat{\alpha}_{12} = 1.55$), low average level of y_2 ($\hat{\alpha}_{22} = 0.73$), and a strong positive association between individual levels on y_1 and y_2 ($r_2 = .68$).

Based on the estimated class proportions, assuming a random and representative sample from the overall population, you might also apply a modifier label of "normal" or "typical" to Class 2 because its members make up an estimated 67% of the population.

Model Estimation

As with LCA, the most common approach for latent profile model estimation is FIML estimation using the EM algorithm under the MAR assumption. And, as with latent class models, the log likelihood surfaces for finite mixture models can be challenging for the estimation algorithms to navigate. Additionally, although the log likelihood function of a identified latent profile model with class-invariant Σ_k usually has a global maximum in the interior of the parameter space, the log likelihood functions for LPA models with class-varying Σ_k are unbounded (like Fig. 25.4.e), which means that the maximum likelihood estimate (MLE) as a global maximizer does not exist. But you may still proceed as the MLE may still exist as a local maximizer possessing the necessary properties of consistency, efficiency, and asymptotic normality (McLachlan & Peel, 2000). When estimating latent profile models, I recommend following the same strategy of using multiple random sets of starting values and keeping track of all the convergence, maximum likelihood replication, condition number, and class size information as with LCA model estimation, to single out models that are not well identified.

Model Building

Principled model building for LPA proceeds in the same manner described in the section on LCA, beginning with the establishment of the (unconditional) measurement model for the latent class variable, with the chief focus during that stage of model building on latent class enumeration. The

following subsections highlight any differences in the evaluations of absolute fit, relative fit, classification accuracy, and the class enumeration process for LPA compared to what has already been advanced in this chapter for LCA.

Absolute Fit. At present, there are not widely accepted or implemented measures of absolute fit for latent profile models. Although it would be theoretically possible to modify exact tests of fit or closeness-of-fit indices available for factor analysis, most of these indices are limited to assessing the model-data consistency with respect to only the mean and variance–covariance structure, which would not be appropriate for evaluation of overall fit for finite mixture models. With finite mixture modeling, you are using an approach that requires individual level data because the formation of the latent class variable depends on all the high-order moments in the data (e.g., the skewness and kurtosis)—not just the first- and second-order moments. You would choose finite mixture modeling over a robust method for estimating just the mean and variance–covariance structure (robust to non-normality in the overall population), even for indirect applications, if you believed that those higher order moments in the observed data provide substantively important information about the overall population heterogeneity with respect to the item set. Because the separate individual observations are necessary for the model estimation, any overall goodness-of-fit index for LPA models would need to compare each observed and model-estimated individual value across all the indicator variables, similarly to techniques used in linear regression diagnostics.

Although you are without measures of absolute model fit, you are not without some absolute fit diagnostic tools. It is possible to compute the overall model-estimated means, variances, covariances, univariate skewness, and univariate kurtosis of the latent class indicator variables and compare them to the sample values, providing residuals for the first- and second-order multivariate moments and the univariate third- and fourth-order moments for the observed items. These limited residuals allow at least some determination to be made about how well the model is fitting the observed data beyond the first- and second-order moments and also allow some comparisons of relative overall fit across models.

In addition to these residuals, you can provide yourself with an absolute fit benchmark by estimating a fully-saturated mean and variance–covariance model that *is* an exact fit to the data with respect to the first- and second-order moments but assumes all higher-order moments have values of zero. This corresponds to fitting a one-class LPA with an unrestricted Σ specification. In the model-building process, you would want to arrive at a measurement model that fit the individual data *better* (as ascertained by various relative fit indices) than a model only informed by the sample means and covariances.

Relative Fit. All of the measures of relative fit presented and demonstrated for latent class models are calculated and applied in the same way for latent profile models.

Classification Diagnostics. It is possible to obtain estimated posterior class probabilities for all individuals in the sample using the maximum likelihood parameters estimates from the LPA and the individuals' observed values on the continuous indicator variables. Thus, all of the classification diagnostics previously described and illustrated for latent class models are calculable and may be used in the same manner for evaluating latent class separation and latent class assignment accuracy for latent profile models.

Class Enumeration. The class enumeration process for LPA is similar to the one for LCA but with the added complication that because the specification of Σ_k can influence the formation of the latent classes, you should consider a full range of Σ_k specifications. I recommend the following approach:

Stage I: Conduct a separate class enumeration sequence following Steps 1through 7 as outlined in the LCA section of this chapter for each type of Σ_k specification: class-invariant, diagonal Σ_k; class-varying, diagonal Σ_k; class-invariant, unrestricted Σ_k; and class-varying, unrestricted Σ_k. Note that the one-class models for the class-invariant, diagonal Σ_k and class-varying, diagonal Σ_k specifications will be the same, as will the one-class models for class-invariant, unrestricted Σ_k and class-varying, unrestricted Σ_k specifications. The "benchmark" model mentioned in the subsection on absolute fit *is* the initial one-class model for class-invariant, unrestricted Σ_k specification.

Stage II: Take the four candidate models yielded by (I) and recalculate the approximate correct model probabilities using just those four models as the full set under consideration. Repeat Steps 5 through 7 with the four candidate models to arrive at your final model selection.

The only two modifications of class enumeration Steps 5 through 7 necessary for applying Stages I and II in LPA are in Steps 5a and 6. In regards to Step 5a:

Rather than relying on the exact test of fit for absolute fit, the "best" model should be the model with the fewest number of classes that has a better relative fit (in terms of the log likelihood value) than the "benchmark" model. Regarding Step 6: Rather than examining the standardized residuals and the classification diagnostics, you should examine the residuals for the means, variances, covariances, univariate skewness, and univariate kurtosis of the indicator variables along with the classification diagnostics. Cross-validation of the final measurement model can be done in the same fashion as described for latent class analysis.

I should note that in Step 7, for both Stages I and II, there may be occasions in the LPA setting for which the model favored by the parsimony principle it not the same the model favored by the interests of conceptual simplicity and clarity. Take the hypothetical example in Figure 25.8. Let's suppose that the two models depicted in Figures 25.8.b and 25.8.c are comparable on all relative fit measures as well as residuals and classification diagnostics. One might perceive the two-class model as more parsimonious than the three-class model (although the two-class model has one more freely estimated parameter than the three-class model), but to interpret and assign substantive labels to the latent classes, you have to account for not only the different means (locations) of the latent classes but also the differences between the classes with respect to the within-class variability and the within-class correlation, which could get decidedly unsimple and unclear in its presentation. However, for the three-class model, you only need to consider the different class-specific means (and the corresponding class separation) to interpret and assign substantive labels to the latent classes because the model imposes constraints such that the classes are identical with respect to within-class variability, and the class indicators are assumed to be unrelated within class for all the classes. There is not an obvious model choice in this scenario. In such a situation, and in cases where the models are agonizingly similar with respect to their fit indices, it is essential to apply substantive and theoretical reflections in the further scrutiny of the model usefulness, especially keeping in mind the intended conditional models to be specified once the measurement model is established.

In the next subsection, I fully illustrate the unconditional LPA modeling process with a real data example, with special attention to elements of the process that are distinct for LPA in comparison to what was previously demonstrated for LCA.

Diabetes Example for Latent Profile Analysis

The data used for the LPA example come from a study of the etiology of diabetes conducted by Reaven and Miller (1979). The data were first made publically available by Herzberg and Andrews (1985) and have become a "classic" example for illustrating multivariate clustering-type techniques (*see*, for example, Fraley & Raftery, 1998, and Vermunt & Magidson, 2002). The original study of 145 nonobese subjects measured participants' ages, relative weights, and collected experimental data on a set of four metabolic variables commonly used for diabetes diagnosis: fasting plasma glucose, area under the plasma glucose curve for the 3-hour oral glucose tolerance test (a measure of glucose intolerance), area under the plasma insulin curve for the oral glucose tolerance test (a measure of insulin response to oral glucose), and the steady state plasma glucose response (a measure of insulin resistance) (Reaven & Miller, 1979). The correlation between the fasting plasma glucose and area under the plasma glucose curve was 0.97 and so the original authors excluded the fasting plasma glucose measure in their analyses of the data. For this illustration, Table 25.11 lists the same three remaining metabolic measures utilized, by name and label, along with descriptive statistics for the study sample. Also included in the example data are the conventional clinical classifications of the subjects into one of the three diagnostic groups (non-diabetics, chemical diabetics, and overt diabetics) made by Reaven and Miller (1979) applying standard clinical criteria that each take into account only one aspect of a participant's carbohydrate metabolism. In their 1979 paper, Reaven and Miller were interested using their data to exploring the viability of a multivariate analytic technique that could classify subjects on the basis of *multiple* metabolic characteristics, independent of prior clinical assessments, as an alternative to the rigid clinical classification with arbitrary cut-off value criteria (e.g., individuals with fating plasma glucose levels in excess of 110 mg/mL are classified as *overt diabetics*). In this example, the original research aim is furthered by investigating the classification of subjects using LPA and comparing the results to the conventional clinical classifications.

In conducting the class enumeration process, knowledge of the existing clinical classification scheme is ignored so that is does not influence decisions with respect to either the number of classes or their interpretation. I begin Stage I of the class enumeration by fitting six models with $K = 1$

Table 25.11. Diabetes Example: Descriptive Statistics for Indicator Measures ($n = 145$)

Measure	Variable name	Mean	SD	Skewness	Kurtosis	[Min, Max]	Correlations (1)	Correlations (2)
(1) Glucose area (mg/10mL/hr)	*Glucose*	54.36	31.70	1.78	2.16	[26.90, 156.80]	1.00	
(2) Insulin area (μU/0.10mL/hr)	*Insulin*	18.61	12.09	1.80	4.45	[1.00, 74.80]	−0.34**	1.00
(3) Steady state plasma glucose (mg/10mL)	*SSPG*	18.42	10.60	0.69	−0.23	[2.90, 48.00]	0.77**	0.01

**$p < 0.01$

to $K = 6$ classes for each of four within-class variance–covariance specifications: class-invariant, diagonal Σ_k; class-varying, diagonal Σ_k; class-invariant, unrestricted Σ_k; and class-varying, unrestricted Σ_k. After $K = 5$, the models for the diagonal unrestricted Σ_k specifications ceased to be well identified, as was the case after $K = 4$ for the class-invariant, unrestricted Σ_k specification and after $K = 3$ for the class-varying, unrestricted Σ_k.

Table 25.12 summarizes the results from the set of class enumerations for each of the Σ_k specifications. Only the results from the well-identified models are presented. Recall that the one-class models for the class-invariant, diagonal Σ_k and class-varying, diagonal Σ_k specifications are the same, as are the $K = 1$ models for the class-invariant, unrestricted Σ_k and class-varying, unrestricted Σ_k specifications. Recall also that the one-class model for the unrestricted Σ_k specification is the minimum-goodness-of-fit benchmark model, and results from this model are enclosed by a bold dashed box for visual recognition. Bolded values in Columns 5 through 10 indicate the value corresponding to the "best" model within each set of enumerations according to each fit index. As was the case for the LCA example, all K-class model were rejected in favor of a $(K+1)$-class model by the BLRT for all values of K considered so there was no "best" or even candidate models to be selected based on the BLRT and those results are not presented in the summary table.

Figure 25.10 displays four panels with plots of the: (a) LL; (b) BIC; (c) CAIC and (d) AWE model values, all plotted on the y-axis versus the number of classes. Each panel has four plot lines, one for each of the Σ_k specifications. The double horizontal line corresponds to the index value of the minimum-goodness-of-fit benchmark of the one-class, unrestricted Σ_k specification. These plots clearly show that all of the models with $K \geq 2$ are improvements over the benchmark model. These plots also illustrate the concept of the "elbow" criteria mentioned in the initial description of the class enumeration process in the LCA section. Observe the BIC plot for the class-varying, diagonal Σ_k specification. Although the smallest BIC value out of the $K = 1$ to $K = 5$ class models corresponds to the four-class model, the BIC values for the three-, four-, and five-class models are nearly the same compared to the values for the one- and two-class models. There is evidence of an "elbow" in the BIC plot at $K = 3$. The bolded values in Column 2 of Table 25.12 indicate the pair of candidate models selected within each of the class enumeration for further scrutiny (following class enumeration Steps 5 and 6 in Stage I) and the boxed values indicates the "best" model selected within each of the class enumeration sets (Step 7 of Stage I). The selection of the four "best" models concluded Step I of the LPA class enumeration process.

For Stage II, I compared the four candidate models, one from each of the Σ_k specifications. Column 11 in Table 25.12 displays the results of recalculating the correct model probabilities using only those four models. This index strongly favors the three-class model with class-varying, unrestricted Σ_k, enclosed by a solid box in Table 25.12. The single horizontal line in all the panel plots of Figure 25.10 corresponds to the best indice values across all the models considered. It is clear from Figure 25.10 that the models with class-varying Σ_k specifications (either diagonal or unrestricted) offer consistently better fit over the models with class-invariant specifications, although the five-class models with class-invariant, diagonal Σ_k approaches the fit of the three- and four-class

Table 25.12. Diabetes Example: Model Fit Indices for Exploratory Latent Profile Analysis Using Four Different Within-Class Variance–Covariance Structure Specifications ($n = 145$)

1	2	3	4	5	6	7	8	9	10	11
$\boldsymbol{\Sigma}_k$	# of classes ($K \leftarrow$	LL	$npar$*	BIC	CAIC	AWE	Adj. LMR-LRT p-value (H_0:K classes; H_1:K$\Downarrow$1 classes)	$\hat{BF}_{K,K\Downarrow 1}$	$cm\hat{P}_K$	$cm\hat{P}_.$
Class-invariant, diagonal $\boldsymbol{\Sigma}_k \Im \boldsymbol{\Sigma}$	1	−1820.68	6	3671.22	3677.22	3719.08	<0.01	<0.10	<0.01	–
	2	−1702.55	10	3454.88	3464.88	3534.64	<0.01	<0.10	<0.01	–
	3	−1653.24	14	3376.15	3390.15	3487.82	<0.01	<0.10	<0.01	–
	4	−1606.30	18	3302.18	3320.18	3445.76	0.29	<0.10	<0.01	–
	5	−1578.21	22	3265.90	3287.90	3441.39	–	–	>0.99	<0.01
Class-varying, diagonal $\boldsymbol{\Sigma}_k$	1	−1820.68	6	3671.22	3677.22	3719.08	<0.01	<0.10	<0.01	–
	2	−1641.95	13	3348.60	3361.60	3452.30	<0.01	<0.10	<0.01	–
	3	−1562.48	20	3224.49	3244.49	3384.03	<0.01	0.38	0.25	–
	4	−1544.10	27	3222.57	3249.57	3437.95	0.15	7.76	0.66	0.08
	5	−1528.73	34	3226.67	3260.67	3497.88	–	–	0.09	–
Class-invariant, unrestricted $\boldsymbol{\Sigma}_k \Im \boldsymbol{\Sigma}$	1	−1730.40	9	3505.60	3514.60	3577.39	<0.01	<0.10	<0.01	–
	2	−1666.63	13	3397.95	3410.95	3501.65	<0.01	<0.10	<0.01	–
	3	−1628.86	17	3342.33	3359.33	3477.93	0.19	<0.10	<0.01	–
	4	−1591.84	21	3288.19	3309.19	3455.70	–	–	>0.99	<0.01
Class-varying, unrestricted $\boldsymbol{\Sigma}_k$	1	−1730.40	9	3505.60	3514.60	3577.39	<0.01	<0.10	<0.01	–
	2	−1590.57	19	3275.69	3294.69	3427.25	<0.01	<0.10	<0.01	–
	3	−1536.64	29	3217.61	3246.61	3448.93	–	–	>0.99	0.92

*number of parameters estimated

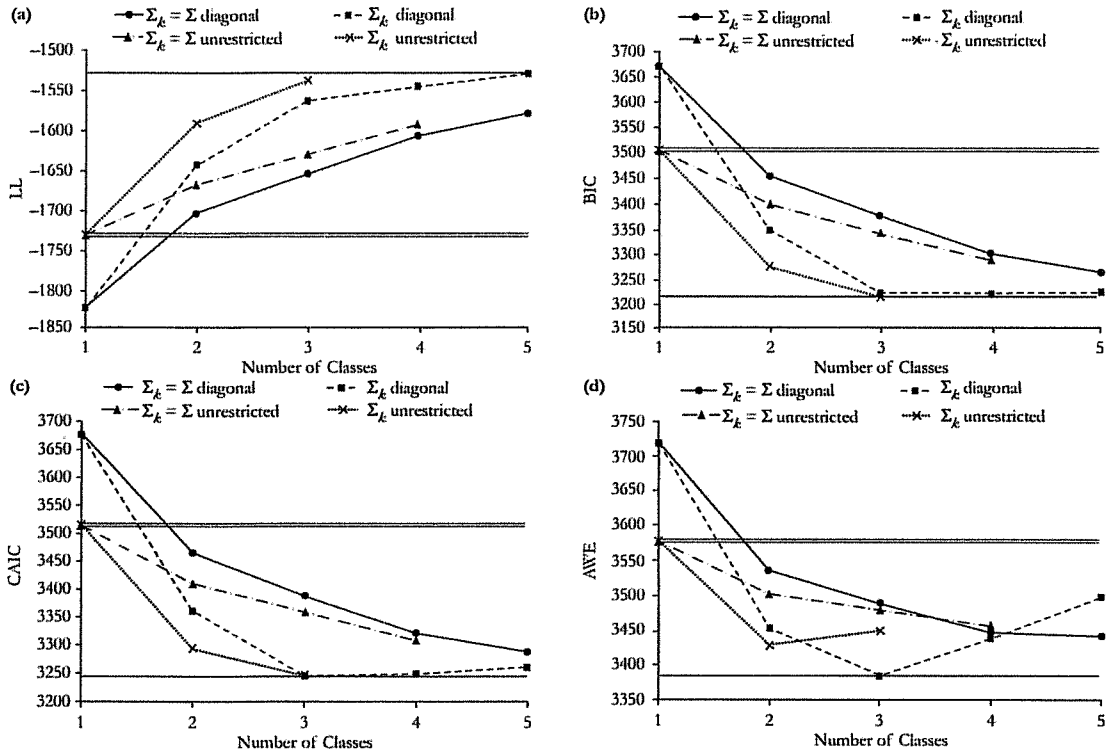

Figure 25.10 Diabetes example: Plots of model (a) LL, (b) BIC, (c) CAIC, and (d) AWE values versus the latent class enumeration ($K = 1, 2, 3, 4, 5$) across four different within-class variance–covariance structure specifications.

models with class-varying Σ_k. The three-class model with class-varying, unrestricted Σ_k and the four-class model with class-varying, diagonal Σ_k were the two candidate models selected for further inspection. Following Step 6 in Stage II, I closely examined the residuals and classification diagnostics of the final two candidate models. Table 25.13 displays the observed, model-estimated, and residuals for the means, variance, covariances, and univariate skewness and kurtosis values of the data for the three-class model with class-varying, unrestricted Σ_k showing a satisfactory fit across all these moments. The four-class model with class-varying, diagonal Σ_k had satisfactory fit in this regard as well, although the fit to the variance–covariance structure of the data was not quite as close. Table 25.14 summarizes the classification diagnostic measures for the three-class model with class-varying, unrestricted Σ_k. All the measures indicate that the three classes are very well separated and there is high accuracy in the latent class assignment. The four-class model with class-varying, diagonal Σ_k had comparably good values on the classification diagnostics. Considering all the information from Stage II, Steps 5 and 6, the three-class model with class-varying, unrestricted Σ_k was selected as the "final" unconditional latent profile model. I should remark here that this model was not in any way conspicuously better fitting than the other candidate model and another researcher examining the same results could ultimately select the other model by giving slightly less weight to model parsimony and giving less consideration that a match between the final class enumeration and the number of diagnostic groups greatly simplifies the planned comparison between subjects' latent class assignments and their conventional clinical classifications.

For the interpretation of the resultant three classes from the final model, it is necessary to examine the model-estimated, class-specific item means, standard deviations, and correlations, provided in Table 25.15 and depicted graphically by the three scatter plots in Figure 25.11. Inspecting the class-specific standard deviation estimates, Class 1 has a high level of homogeneity with respect to all three indicator variables, with notably less variability than in the overall sample and less than either of the other two classes. Class 3 is the least homogeneous with respect to *glucose* and *SSPG*, with variability in both actually greater than the overall sample. Class

Table 25.13. Diabetes Example: Observed, Mixed Model-Estimated, and Residual Values for Means, Variances, Covariances and Correlations, Univariate Skewness, and Univariate Kurtosis Based on the Three-Class Latent Profile Analysis with Class-Varying, Unrestricted Σ_k (n = 145)

Variable	Observed	Model-estimated	Residual
Mean(Glucose)	54.36	54.36	0.00
Mean(Insulin)	18.61	18.61	0.00
Mean(SSPG)	18.42	18.42	0.00
Var(Glucose)	1004.58	997.65	6.93
Var(Insulin)	146.25	145.24	1.01
Var(SSPG)	112.42	111.65	0.78
Cov(Glucose,Insulin) (Correlation)	−129.18 (−0.34)	−128.29 (−0.34)	−0.89 (0.00)
Cov(Glucose,SSPG) (Correlation)	259.09 (0.77)	257.30 (0.77)	1.79 (0.00)
Cov(Insulin,SSPG) (Correlation)	1.02 (0.01)	1.01 (0.01)	0.01 (0.00)
Skewness(Glucose)	1.78	1.75	0.03
Skewness(Insulin)	1.80	1.49	0.31
Skewness(SSPG)	0.69	0.72	−0.03
Kurtosis(Glucose)	2.16	2.49	−0.32
Kurtosis(Insulin)	4.45	2.96	1.49
Kurtosis(SSPG)	−0.23	0.19	−0.42

Table 25.14. Diabetes Example: Model Classification Diagnostics for the Three-Class Latent Profile Analysis With Class-Varying, Unrestricted Σ_k ($E_3 = .88$; $n = 145$)

Class k	$\hat{\pi}_k$	95% C.I.*	$mcaP_k$	$AvePP_k$	OCC_k
Class 1	0.512	(0.400, 0.620)	0.524	0.958	21.74
Class 2	0.211	(0.119, 0.307)	0.221	0.918	41.86
Class 3	0.277	(0.191, 0.386)	0.255	0.973	94.06

*Bias-corrected bootstrap 95% confidence intervals

2 is the least homogenous with respect to *insulin*, also having greater variability than the overall sample. The similarities and differences in the level of class homogeneity with respect to each of the three items can be judged visually in Figure 25.11 by length and width of the overlaid ellipses in the three plots.

In judging class separation for the two classes that do not have a high degree of homogeneity for at least one of the indicator variables, the distances

Table 25.15. Diabetes Example: Model-Estimated, Class-Specific Means, Standard Deviations (SDs), and Correlations with Corresponding Bias-Corrected Bootstrap 95% Confidence Intervals Based on the Three-Class Latent Profile Analysis With Class-Varying, Unrestricted Σ_k ($n = 145$)

				Correlations	
Class	Variable	Mean ($\hat{\alpha}_{mk}$)	SD ($\sqrt{\hat{\theta}_{mmk}}$)	(1)	(2)
Class 1 (52%)	(1) *Glucose*	35.69 (34.09, 37.18)	4.39 (3.11, 5.50)	1.00	
	(2) *Insulin*	16.58 (15.31, 17.96)	5.17 (4.24, 6.11)	0.15 (-0.14, 0.42)	1.00
	(3) *SSPG*	10.50 (8.90, 12.43)	4.33 (3.48, 5.97)	0.29 (–0.05, 0.57)	0.36** (0.08, 0.57)
Class 2 (22%)	(1) *Glucose*	47.66 (43.93, 52.72)	7.29 (4.95, 10.73)	1.00	
	(2) *Insulin*	34.35 (27.38, 44.06)	15.12 (11.43, 19.40)	0.36 (–0.33, 0.77)	1.00
	(3) *SSPG*	24.41 (22.52, 25.99)	3.71 (2.15, 5.49)	0.03 (–0.40, 0.50)	–0.10 (–0.73, 0.54)
Class 3 (26%)	(1) *Glucose*	93.92 (78.13, 112.48)	35.76 (30.30, 41.51)	1.00	
	(2) *Insulin*	10.38 (7.97, 13.31)	6.03 (4.74, 8.34)	–0.76** (–0.87, –0.58)	1.00
	(3) *SSPG*	28.48 (24.42, 33.93)	10.65 (8.22, 12.80)	0.73** (0.41, 0.85)	–0.40** (–0.61, –0.08)

** $p < 0.01$

between the class means for those variables must be large for the overlap between the classes to still be small. Table 25.16 presents the distance estimates (i.e., standardized differences in means) for each pairwise class comparison on each of the three indicators variables. Large estimated absolute distance values greater than 2.0, corresponding to less than 20% overlap, are bolded for visual clarity. All classes are well separated with moderate to large estimated distances on at least two of the three items, and every item distinguishes between at least two of the three classes. The classes are all well separated with respect to their means on *glucose*, with the greatest distances between Class 1 and the other two classes. There is a similar pattern for the separation on *SSPG* with large distances between Class 1 and Classes 2 and 3. However, in the case of *SSPG*, there is a very small separation between Classes 2 and 3—meaning that there is a high degree of overlap in the distribution of individual values on *SSPG* across those two classes, rendering those two classes difficult to distinguish with respect to *SSPG*. In contrast, Classes 2 and 3 have a large distance between their means for *insulin*, whereas there is only a modest separation between Classes 1 and 3. Figure 25.11 provides a visual impression of these varying degrees of separation across the classes with respect to each of the three measures.

Because the final model selected had a class-varying, unrestricted Σ_k, the distinctness of the classes must also be evaluated with regards to the class-specific variance–covariance structure before a full substantive interpretation of the classes is rendered. I have already remarked, when assessing class homogeneity, that Class 3 was much more variable than the other two classes with respect to *glucose* and *SSPG* and that Class 2 was much more variable with respect to *insulin*. In terms of the covariance structure, presented as correlations in Table 25.15, Class 3 has a large and significant negative

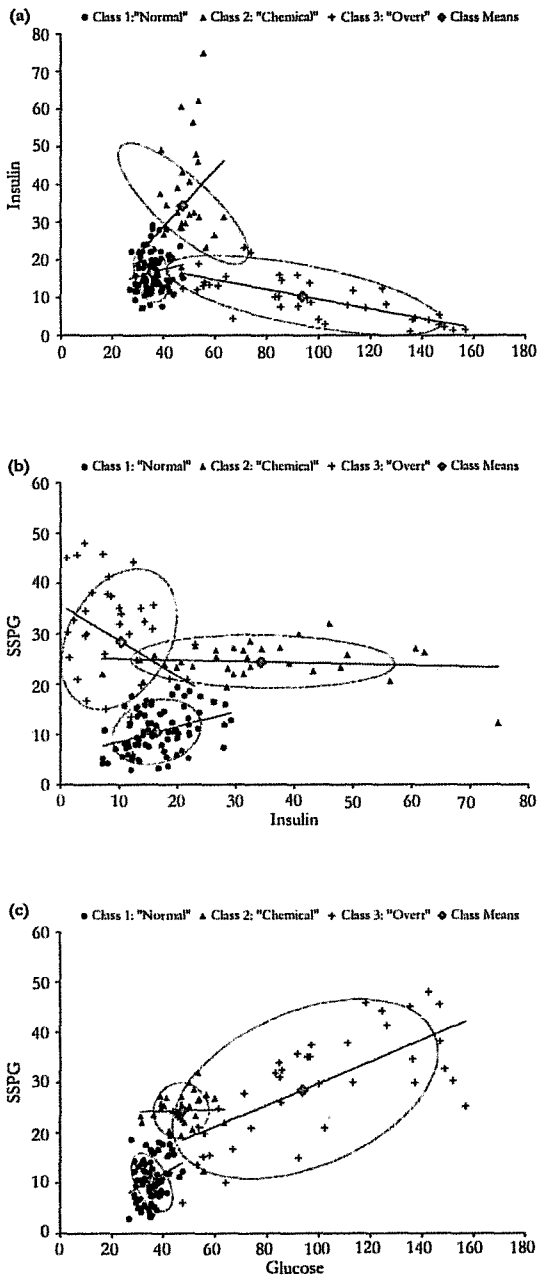

Figure 25.11 Diabetes example: Scatter plots of observed sample values marked by modal latent class assignment based on the unconditional three-class LPA for (a) insulin versus glucose, (b) SSPG versus insulin, and (c) SSPG versus glucose. For (a)-(c), diamonds represent the model-estimated class-specific bivariate mean values, trend lines depict the model-estimated within-class bivariate associations, and the ellipse heights and widths correspond to 3.0 model-estimated within-class standard deviations for the indictors on the y- and x-axis, respectively.

correlation between *glucose* and *insulin*, whereas that correlation is positive and non-significant for both Classes 1 and 2. Class 3 has a large and significant positive correlation between *glucose* and *SSPG*, whereas that correlation, although positive, is non-significant for both Classes 1 and 2. Class 3 has a moderate and significant negative correlation between *insulin* and *SSPG*, whereas Class 1 has a moderate and significant negative correlation and Class 2's correlation is negative and non-significant. Because the correlation between *insulin* and *SSPG* is the only significant correlation for Class 1 and none of the correlations were significant for Class 2, Classes 1 and 2 are not well separated by their covariance structure. Class 3 is the class with two quite large and all significant correlations, and these features are an important part of what distinguishes Class 3, and Class 3 is well separated from both Class 1 and Class 2 with respect to all covariance elements. However, because Class 3 only represents 26% of the population, it is not surprising that the results of the three-class model with a class-varying, diagonal Σ_k were so close to the results of this model, allowing the within-class correlations.

Table 25.16. Diabetes Example: Estimated Standardized Differences in Class-Specific Indicator Means, $\hat{d}_{m\cdot\cdot}$, Based on Model-Estimated, Class-Specific Indicator Means and Variances from the Three-Class Latent Profile Analysis With Class-Varying, Unrestricted Σ_k ($n = 145$)

Variable	Class 1 vs. Class 2	Class 1 vs. Class 3	Class 2 vs. Class 3
(1) Glucose	−2.21	−2.78	−1.73
(2) Insulin	−1.91	1.13	−2.15
(3) SSPG	−3.34	−2.53	−0.49

For the substantive class interpretation, I begin with the class most distinct in means and variance–covariance structure from the other classes, Class 3, with an estimated proportion of 26%. Class 3 consists of individuals with high values on *glucose*, on average, compared to the overall sample and Classes 1 and 2. Within this class, there is a strong negative association between *glucose* and *insulin* and strong positive association between *glucose* and *SSPG* such that the individuals in Class 3 with higher values on *glucose* have lower values on *insulin* and higher values on *SSPG*, on average. The high average value on *glucose* and *SSPG* with the lower average value on *insulin* along with the very strong associations across the three indicators, leads this class to be labeled the "overt" diabetic class.

Class 2, with an estimated proportion of 22%, consists of individuals with *insulin* and *SSPG* values, higher, on average, than the overall population means and notably higher than Class 1. Class 2 is not much different than Class 3 with respect to *SSPG* but has a much higher mean on *insulin*. Individuals in Class 2 have higher-than-average values for *glucose*, and Class 2 is nearly as different from Class 1 as Class 3 is in terms of average *glucose* values even though the mean in Class 2 is lower than Class 3. There are no significant associations between the three indicators in Class 2. The higher-than-average values on *glucose*, *insulin*, and *SSPG*, with a notably higher *insulin* mean but lower *glucose* mean than the "overt" diabetic class, suggests the label of the "chemical" diabetic for this class.

Class 1, with an estimated proportion of 52%, consists of individuals with *glucose* and *SSPG* values lower, on average, than the overall population mean and notably lower than either Class 2 or Class 3. The individuals in Class 1 have *insulin* values, on average, near the overall population mean, higher than the "chemical" diabetic class and lower than the "overt" diabetic class. For the Class 1 subpopulation, there is a moderate positive association between *glucose* and *SSPG* such that individuals with higher values on *glucose* in Class 1 have higher value, on average, for *SSPG*. This association is quite different from the moderate negative association in the "overt" diabetic group such that among those in Class 3, individuals with higher values on *glucose* have lower values on *SSPG*, on average. The lower *glucose* and *SSPG* average levels, the average *insulin* levels, the positive association between *glucose* and *SSPG*, and the estimated class proportion greater than 50% suggest the label of the "normal" (non-diabetic) for this class.

With the results from the unconditional LPA in-hand, I can compare individual model-estimated class membership for individuals in the sample to their clinical classifications. As it happens, a three-class model for the LPA was selected and the latent classes were interpreted in a way that matched, at the conceptual level, the three clinical classification categories. To make the descriptive, *post hoc* comparisons, I use the modal class assignment for each participant to compare to the clinical classification. Because the comparison is descriptive (rather than inferential) and there is a very high level of classification accuracy for all three classes (*see* Table 25.14), it is reasonable to use the modal class assignment to get a sense of the correspondence between "true" class membership and the clinical classifications. Table 25.17 displays a cross-tabulation comparison between latent class (modal) membership and the clinical classifications. Cells corresponding to "matches" between the modal class assignments and the clinical classifications are boxed in bold. In general, there is a strong concordance across all three latent classes, with only 22 (15%) of the participants having a mismatch between modal latent classification and clinical status. The highest correspondence is between the "normal" latent class and the non-diabetic clinical classification, with 91% of those modally assigned to the "normal" class also having a non-diabetic clinical status. The lowest correspondence is between the "chemical" latent class and the chemical clinical classification but was still reasonably high, with 72% of those modally assigned to the "chemical" diabetes class also having a chemical diabetic clinical status. It is also informative to examine the nature of the noncorrespondence. Of those individuals modally assigned to the "normal" class, none had an overt diabetic clinical status. Similarly, of those individuals modally assigned to the "overt" class, none had a non-diabetic clinical status. In both cases, the mismatch involved individuals with a chemical diabetic clinical status. Of the individuals modally assigned to the "chemical" diabetes class that did not have a chemical diabetic clinical status, most had a non-diabetic clinical status, but two did have an overt diabetic clinical status.

Because it was originally of interest whether a multivariate model-based classification could offer improvements over the univariate cut-off criteria used in the clinical classifications, I closely examined the 22 cases for which there is a mismatch. Table 25.18 summarizes the average posterior class probabilities stratified by both modal class assignments and clinical classifications. What can be seen in this table is that the average posterior class probabilities for the modally assigned classes, bolded and boxed in Table 25.18, are all reasonably high. In other words, even those groups of individuals with a mismatch between the modal latent class membership and clinical status are relatively well classified, on average, by the model. If one examines the raw data for these individuals, it can be seen that these individuals were *not* well classified by the clinical criteria. For example, most of the patients with a chemical diabetic clinical status and a "normal" modal class assignment were all borderline on clinical diagnosis criteria. Some of the patients with a non-diabetic clinical status that were hyperinsulinemic and insulin-resistant, but managed to maintain

Table 25.17. Diabetes Example: Modal Latent Class Assignment vs. Clinical Classification Frequencies and Row Percentages

	Clinical classification			
Modal class assignment	Non-diabetic	Chemical diabetic	Overt diabetic	Total
"Normal"	69 (91%)	7 (9%)	0 (0%)	76 (100%)
"Chemical"	7 (22%)	23 (72%)	2 (6%)	32 (100%)
"Overt"	0 (0%)	6 (16%)	31 (84%)	37 (100%)
Total	76	36	33	145

Table 25.18. Diabetes Example: Average Posterior Class Probabilities by Modal Latent Class Assignment and Clinical Classification

Modal class assignment	Clinical classification	f	Mean($\hat{p}_{normal}$)	Mean($\hat{p}_{chemical}$)	Mean($\hat{p}_{overt}$)
"Normal"	Non-diabetic	69	0.97	<0.01	0.02
	Chemical diabetic	7	0.79	0.05	0.15
	Overt diabetic	0	–	–	–
"Chemical"	Non-diabetic	7	0.06	0.85	0.09
	Chemical diabetic	23	0.02	0.93	0.04
	Overt diabetic	2	<0.01	>0.99	<0.01
"Overt"	Non-diabetic	0	–	–	–
	Chemical diabetic	6	0.07	0.08	0.85
	Overt diabetic	31	<0.01	<0.01	>0.99

normal glucose tolerance, were modally assigned by the model to the "chemical" diabetes class. These differences suggest that using a model that takes into account multiple metabolic characteristics may offer improved and more medically comprehensive classification over the rigid and arbitrary univariate clinical cut-off criteria.

Latent Class Regression

The primary focus, thus far, has been on the process for establishing the measurement model for latent class variables with either categorical indicators (LCA) or continuous indicators (LPA). However, that process is usually just the first step in a structural equation mixture analysis in which the latent class variable (with its measurement model) is placed in a larger system of variables that may include hypothesized predictors and outcomes of latent class membership. To provide readers with a sense of how these structural relationships can be specified, I present in this section a latent class regression (LCR) model for incorporating predictors of latent class membership. This presentation is applicable for both LCA and LPA models.

Covariates of latent class membership may serve different purposes depending on the particular aims of the study analysis. If attention is on developing and validating the measurement model for a given construct using a latent class variable, covariates can be used to assess criterion-related validity of the latent class measurement model. It may be possible, based on the conceptual framework for the latent class variable, to generate hypotheses about how the latent classes should relate to a select set of covariates. These hypotheses can then be evaluated using a LCR model (Dayton & Macready, 2002); support for the hypotheses equates to increased validation of the latent class variable. You may also gain a richer characterization and interpretation of

the latent classes through their relationships with covariates.

Beyond construct validation, covariates can be used to test hypotheses related to a theoretical variable system in which the latent class variable operates. In such a variable system, you may have one or more theory-driven covariates that are hypothesized to explain individual variability in an outcome where the individual variability is captured by the latent class variable.

In the remainder of this section I describe the formulation of the LCR model and illustrate its use in the LSAY example.

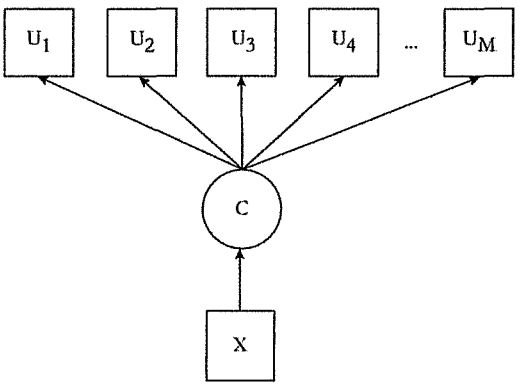

Figure 25.12 Generic path diagram for a latent class regression model.

Model Formulation

For a LCR model, the measurement model parameterization, describing the relationships between the latent class variable and its indicators, remains the same as for the unconditional models but the structural model changes in that the latent class proportions are now conditional on one or more covariates. For example, in the LCA specification, the conditional version of Equation 4 is given by

$$\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}, x_i) = \sum_{k=1}^{K} \left[\Pr(c_i = k|x_i) \cdot \Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}|c_i = k)\right]. \quad (40)$$

A multinomial regression is used to parameterize the relationship between the probability of latent class membership and a single covariate, x, such that

$$\Pr(c_i = k|x_i) = \frac{\exp(\gamma_{0k} + \gamma_{1k}x_i)}{\sum_{j=1}^{K} \exp(\gamma_{0j} + \gamma_{1j}x_i)}, \quad (41)$$

where Class K is the reference class and $\gamma_{0K} = \gamma_{1K} = 0$ and for identification. γ_{0k} is the log odds of membership in Class k versus the reference class, Class K, when $x = 0$ and γ_{1k} is the log odds ratio for membership in Class k (versus Class K) corresponding to a one unit difference on x. Equations 40 and 41 are represented in path diagram format as shown in Figure 25.12. Equation 41 can easily be expanded to include multiple covariates. (For more general information about multinomial regression, *see*, for example, Agresti, 2002.) It is also possible to examine latent class difference with respect to a grouping or concomitant variable using a multiple-group approach similar to multiple-group factor analysis (Collins & Lanza, 2010; Dayton & Macready, 2002), but such models are beyond the scope of this chapter.

Model Building

As previously explained in the earlier model-building subsections, the first step in the model-building process—even if the ultimate aims of the analysis include testing hypotheses regarding the relationships between predicting covariates and latent class membership—is to establish the measurement model for the latent class variable. Based on simulation work (Nylund & Masyn, 2008), showing misspecification of covariate effects in a LCA can lead to bias in the class enumeration, it is strongly recommended that the building of the measurement model—particularly the class enumeration stage—is conducted with unconditional models, only adding covariates once the final measurement model has been selected. The selection and order of covariate inclusion should be theory-driven and follow the same process as with any regular regression model with respect to risk factors or predictors of interest, control of potential confounders, and so forth.

The specification provided in Equations 40 and 41 assumes that there is no direct effect of x on the latent class indicator variables (which would be represented in Figure 25.12 by a path from x to one or more the us). However, omission of direct covariate effects (if actually present) can lead to biased results (similarly to the omission of direct effects in a latent factor model). If direct effects are incorrectly omitted, then the measurement parameters for the latent class variable can be distorted, shifting from their unconditional model estimates and potentially misrepresenting the nature of the latent classes; in

addition, the estimated latent class proportions and the effects of the covariate on latent class membership can be biased. In fact, if no direct effects are included and the latent classes in the LCR model change substantively in size or meaning relative to the final unconditional latent class model, then this can signal a misspecification of the covariate associations with the latent class indicators, recommending a more explicit test of direct effects. The presence of direct effects is analogous to the presence of measurement non-invariance in a factor model or differential item functioning in an item response theory model—a direct effect on an indicator variable means that individuals belonging to the *same* latent class but with *different* values of x have different expected outcomes for that observed indicator. Although elaborating on the process of testing for direct effects and measurement non-invariance with respect to covariates being incorporated into a LCR model is beyond the scope of this chapter, I do recommend that, in the absence of prior knowledge or strong theoretical justification, direct effect should be tested as part of the conditional model-building process and the addition of latent class covariates. (For more on covariate direct effects, measurement non-invariance, and violations of the conditional independence assumption resulting from direct covariate effects in LCRs, *see*, for example, Bandeen-Roche, Miglioretti, Zeger, Rathouz, & Paul, 1997; Hagenaars, 1988; and Reboussin, Ip, & Wolfson, 2008.)

I should remark here that a LCR analysis following the building of a latent class measurement model using a full latent class enumeration process without any *a priori* restrictions on the number and nature of the latent classes is a blend of confirmatory (LCR) and exploratory (latent class enumeration) elements. Although the establishment of the measurement model proceeds in a more exploratory way, the model that you carry forward to inferential structural models is not constrained in the same way it would be when conducting an EFA and then subsequent CFA in the same sample, and thus you do not face the same dangers of inflating Type I error rates and capitalizing on chance. However, it is preferable, if possible, to validate the measurement model with new data so that you can feel more confident that the measurement model might generalize to other samples and that your latent classes are not being driven by sampling variability and are not overfit to the particular sample data at hand. Otherwise, it is important to acknowledge in the interpretation of the results the exploratory and confirmatory character of the analysis (Lubke, 2010).

Longitudinal Study of American Youth Example for Latent Class Regression

To illustrate LCR, I return to the LSAY example used in the Latent Class Analysis section. In addition to the nine math disposition items, the example data set also included the variable of student sex (coded here as *female* = 1 for females students and *female* = 0 for male students). Beginning with the five-class unconditional model, I fit two models: Model 0, a five-class model with the latent class variable regressed on *female* but with all multinomial regression coefficients for *female* fixed at zero; Model 1, a five-class model with the latent class variable regressed on *female* with all multinomial regression coefficients for *female* freely estimated. I conducted parallel analyses for both Subsamples A and B and found similar results; only the results for Subsample A are presented here.

There is a significant overall association between student sex and math disposition class membership (Model 0 vs. Model 1: $X^2_{diff} = 27.76$, $df = 4$, $p < .001$). There was no significant shift in the measurement parameters between Model 0 and Model 1 that would have suggested the presence of one or more direct effects of *female* on the items themselves. This descriptive comparison of parameter estimates is not a concrete test of direct effects (that should normally be done), but because explicit testing for differential item functioning in latent class models is beyond the scope of this chapter, I will cautiously treat this model comparison as a satisfactory heuristic evaluation of measurement invariance that allows me to proceed with an interpretation of the LCR results without including direct effects.

Examining the results of the LCR presented in Table 25.19, the multinomial regression parameters represent the effects of student sex on class membership in each class relative to the reference class (selected here as Class 1: "Pro-math without anxiety"). Given membership in either Class 1 ("Pro-math without anxiety") or Class 2 ("Pro-math with anxiety"), females are significantly less likely to be in Class 2 than Class 1 ($\hat{OR} = 0.52$), whereas females are significantly more likely to be in Class 4 ("I don't like math but I know it's good for me") than Class 1 ($\hat{OR} = 1.72$). There is no significant difference in the likelihood of membership in Class 5 ("Anti-math with anxiety") among males

Table 25.19. LSAY Example: Five-Class Latent Class Regression Results for the Effects of Student Sex (female = 1 for female; female = 0 for male) on Latent Class Membership for Subsample A ($n_A = 1338$)

C regressed on *female*		$\hat{\gamma}_{1k}$	s.e.	*p*-value	$\hat{OR}$
Class 1 (ref)	"Pro-math without anxiety"	0.00	–	–	1.00
Class 2	"Pro-math with anxiety"	–0.66	0.21	<0.01	0.52
Class 3	"Math lover"	0.17	0.21	0.43	1.18
Class 4	"I don't like math but I know it's good for me"	0.55	0.19	<0.01	1.72
Class 5	"Anti-math with anxiety"	–0.32	0.22	0.14	0.73

and females in either Class 1 or 5. Rather than making all pairwise class comparisons for student sex by changing the reference class, a better impression of the sex differences in class membership can be given through a graphical presentation such as the one depicted in Figure 25.13, which shows the model-estimated class proportions for the total population and for the two values of the covariate—that is, for males and females. You can see in this figure that the sex differences are primarily in the distribution of individuals across Classes 2 ("Pro-Math With Anxiety") and 4 ("I Don't Like Math but I Know It's Good for Me") with females more likely than males, overall, to be in Class 4 and less likely to be in Class 2.

Post Hoc Class Comparisons

This section has presented a LCR model that simultaneously estimates the latent class measurement model and the structural relationships between the latent class variable and one or more covariates. The simultaneous estimation of the measurement and structural models is recommended whenever possible. However, there is a not-so-unusual practice in the applied literature of doing *post hoc* class comparisons, taking the modal class assignments based on the unconditional latent class measurement model and treating those values as observed values on a manifest multinomial variable in subsequent analyses. This is what I did for the diabetes example, comparing the modal class assignments to the clinical classifications, and such a *post hoc* comparison can be a very useful descriptive technique for further understanding and validation of the latent classes. The problem of this *post hoc* classification approach comes when modal class assignments are used in formal hypothesis testing, moving beyond the descriptive to inferential analyses.

Such a "classify-analyze" approach is problematic because it ignores the error rates in assigning subjects to classes. Because the error rates can vary from class to class, with smaller classes having higher prior probabilities of incorrect assignment, even with well-separated classes, there can be bias in the point estimates as well as the standard errors for parameters related to latent class membership. In addition, there is error introduced from the posterior class probabilities that are used for the modal class assignment because they are computed using parameter estimates and contain the uncertainty from those estimates. Studies have shown that assignment error rates can be considerable (Tueller, Drotar, & Lubke, 2011), posing serious threats to the validity of *post hoc* testing.

Advanced Mixture Modeling

Although a substantial amount of information has been covered in this chapter, I have only scratched the surface in terms of the many types of population heterogeneity that can be modeled using finite mixtures. However, what is provided here is the foundational understanding that will enable you to explore these more advanced models. Just as with factor analysis and traditional structural equation modeling, the basic principles of model specification, estimation, evaluation, and

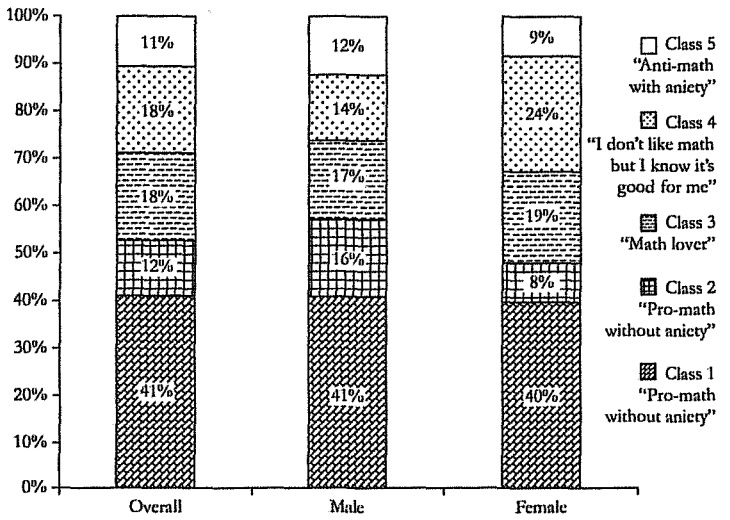

Figure 25.13 LSAY example: Model-estimated overall and sex-specific latent class proportions for the five-class LCR.

interpretation extend quite naturally into more complicated modeling scenarios.

This section provides a very brief overview of some of modeling extensions currently possible in a mixture modeling framework. The first extension relates to the latent class indicators and their within-class distributions. I presented two models—LCA and LPA—that had exclusively categorical or exclusively continuous indicator variables. However, recent advances in maximum likelihood estimation using complex algorithms in a general latent variable modeling framework (*see*, for example, Asparouhov & Muthén, 2004, and Skrondal & Rabe-Hesketh, 2004) have rendered the necessity of uniformity of measurement scales among the indicators obsolete, allowing indicators for a single latent class variable to be of mixed measurement modalities, while also expanding the permissible scales of measures and error distributions for the manifest variables. It is now possible to specify a latent class variable with indicators of mixed modalities or measurement scales including interval and ratio scales of measures, censored interval scales, count scales, ordinal or Likert scales, binary or multinomial responses, and so forth. It is also possible to specify a range of within-class distributions for those indicators—for example, Poisson, zero-inflated Poisson, or negative binomial for count scales; normal, censored normal, censored-inflated normal for interval scales, and so forth. Additionally, the class-specific distribution functions can be from different parametric families across the classes.

Another extension involves the scale of the latent class variable. In this presentation, I used the traditional formulation of the latent class variable as a latent multinomial variable. However, there are latent class models that bridge the gap between the latent multinomial variable models and the latent factor models, such as discretized latent trait models, located latent class models, and latent class scaling models (Croon, 1990, 2002; Dayton, 1998; Heinen, 1996)—all forms of *ordered* latent class models. In addition, recent advances have further blurred the lines of conventional classification schemes for latent variable models (Heinen, 1996) by allowing both latent factors *and* latent class variables to be included in the same analytic model. These so-called hybrid models, also termed *factor mixture models*, include both continuous and categorical latent variables as part of the same measurement model (Arminger, Stein, & Wittenburg, 1999; Dolan & van der Maas, 1998; Draney, Wilson, Gluck, & Spiel, 2008; Jedidi, Jagpal, & DeSarbo, 1997; Masyn, Henderson, & Greenbaum, 2010; Muthén, 2008; Vermunt & Magidson, 2002; von Davier & Yamamoto, 2006; Yung, 1997). These models combine features from both conventional factor analysis and LCA. Special cases of these hybrid models include mixture item response theory models and growth mixture models.

Extensions in mixture model specification and estimation include the accommodation of complex sampling weights (Patterson, Dayton, & Graubard, 2002); the use of Bayesian estimation

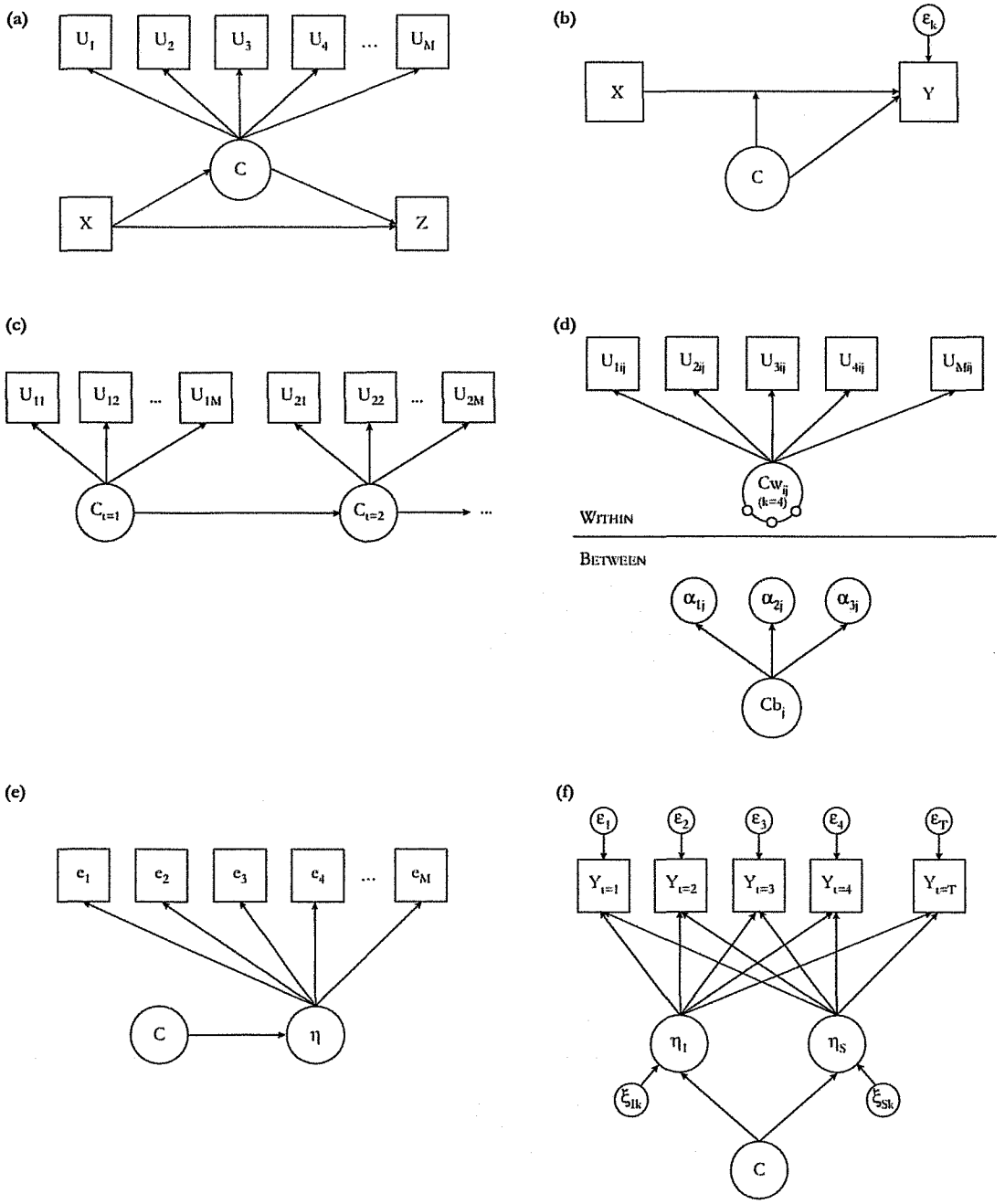

Figure 25.14 Generic path diagrams for a (a) latent class mediation model, (b) regression mixture model, (c) latent transition model, (d) multilevel latent class model, (e) discrete-time survival factor mixture model, and (f) growth mixture model.

techniques (Asparouhov & Muthén, 2010; Garrett & Zeger, 2000; Gelfand & Smith, 1990; Lanza, Collins, Schafer, & Flaherty, 2005) in place of full-information maximum likelihood; the adaptation of fuzzy clustering algorithms and allowing graded latent class membership (Asparouhov & Muthén, 2008; Yang & Yu, 2005); and the use of multiple imputation for missing data combined with MLE (Vermunt, Van Ginkel, Van der Ark, & Sijtsma, 2008).

The six panels of Figure 25.14 display path diagram representations of some of the many advanced mixture models available to researchers. Figure 25.14.a depicts is a latent class mediation model

(Petras, Masyn, & Ialongo, 2011), extending the LCR model to include an outcome of latent class membership that may also be influenced by the covariate. Figure 25.14.b depicts a regression mixture model (RMM; Desarbo, Jedidi, & Sinha, 2001; Van Horn, Jaki, Masyn, Ramey, Antaramian, & Lamanski, 2009) in which the latent class variable is measured by the conditional distribution of an outcome variable, y, regressed on x—that is, the latent class is specified to characterize differential effects of x on y present in the overall population. Figure 25.14.c displays the longitudinal extension of latent class analysis: latent transition analysis (LTA). In LTA (Collins & Lanza, 2010; Nylund, 2007), a special case of a broader class of mixture models called Markov chain models (Langeheine & van de Pol, 2002), there is a latent class variable at each time-point or wave, and the relationship between the classes across time describe individual transitions in class membership through time. Figure 25.14.d displays the multilevel extension of LCA. In MLCA (Asparouhov & Muthén, 2008; Henry & Muthén, 2010; Nylund-Gibson, Graham, & Juvonen, 2010), the class proportions within cluster (represented by shaded circles on the boundary of the *within*-cluster latent class variable, *cw*) vary across clusters. And the variability in class proportions across clusters is captured by a *between*-cluster latent class variable, *cb*. The classes of *cb* represent different groups of clusters characterized by their distributions of individuals across classes of *cw*. Figures 46.14.e and 46.14.f depict two special types of factor mixture models. The diagram in Figure 25.14.e represents a discrete-time survival factor mixture model (Masyn, 2009) in which there is an underlying factor that captures individual-level frailty in the discrete-time survival process measured by the event history indicator, e_m, and the latent class variable characterizes variability in the individual frailties. The diagram in Figure 25.14.f represents a growth mixture model (Feldman, Masyn, & Conger, 2009; Muthén & Asparouhov, 2009; Petras & Masyn, 2010) in which there are latent growth factors that capture the intra-individual growth process, defining individual growth trajectories, and a latent class variable that characterizes (part of) the interindividual variability in the growth trajectories. Examples of other advanced mixture models not depicted in Figure 25.14 include pattern-mixture and selection models for non-ignorable missing data (Muthén, Asparouhov, Hunter, & Leuchter, 2011) and complier average causal effect models (Jo, 2002). What I have provided here is by no means a fully comprehensive or exhaustive list of advanced mixture models but is intended to give the reader a flavor of what extensions are possible.

Conclusion

This chapter represents what I believe to be the current, prevailing "best practices" for basic mixture modeling, specifically LCA and LPA, in terms of model specification, estimation, evaluation, selection, and interpretation. I have also provided a very limited introduction to structural equation mixture modeling in the form of LCR. In addition, in the previous section, you have been given a partial survey of the many more advanced mixture models currently in use. It should be evident that mixture models offer a flexible and powerful way of modeling population heterogeneity. However, mixture modeling, like all statistical models, has limitations and is perhaps even more susceptible to misapplication that other more established techniques. Thus, I take the opportunity in closing to remind readers about some of the necessary (and untestable) assumptions of mixture modeling and caution against the most common misuses.

Most of this chapter has focused on *direct* applications of mixture modeling, for which one assumes *a priori* that the overall population consists of two or more homogeneous subpopulations. The direct application is far more common in social science applications than the indirect application. One assumes that there are, in truth, distinct types of groups of individuals that are in the population to be revealed. "This assumption is critical, because it is always possible to organize any set of data into classes, which then can be said to indicate types, but there is no real finding if an analysis merely indicates classifications in a particular sample. To be of scientific value, the classifications must represent lawful phenomena, must be replicable, and must be related to other variables within a network that defines construct validity." (Horn, 2000, p. 927) Because this assumption is an *a priori* assumption of a mixture model, utilizing a direct mixture modeling approach does *not* test a hypothesis about the existence of discrete groups or subtypes. (There are analytic approaches that are designed to explore the underlying latent structure of a given construct, e.g., whether the underlying construct is continuous or categorical in nature, and interested readers are referred to the chapter in this handbook on taxometric methods and also Masyn, Henderson, and

Greenbaum, 2010.) Nor does the fact that a K-class model is estimable with the sample data prove there are K classes in the population from which the sample was drawn.

Furthermore, the (subjective) selection of a final K-class model does not prove the existence of *exactly* K subgroups. Recall how the specification of Σ_k in a latent profile analysis can influence which class enumeration is "best." The number of latent classes that you settle on at the conclusion of the class enumeration process could very well not reflect the actual number of distinct groups in the population. Attention must also be paid, during the interpretation process, to the fact that the latent classes extracted from the data are inextricably linked to the items used to identify those classes because the psychometric properties of the items can influence the formation of the classes. You assume that you have at your disposal the necessary indicators to identify all the distinct subgroups in the population and can only increase confidence in this assumption through validation of the latent class structure.

I did not provide concrete guidelines about sample size requirements for mixture modeling because they depend very much on the model complexity; the number, nature, and separation of the "true" classes in the population; and the properties of the latent class indicators themselves (Lubke, 2010). "Analyses for a very simple latent class models may be carried out probably with as few as 30 subjects, whereas other analyses require thousands of subjects." (Lubke, 2010, p. 215) Thus, what is critical to be mindful of in your interpretation of findings from a mixture model is that mixture models can be sensitive to sampling fluctuation that may limit the generalizability of the class structure found in a given sample and that smaller samples may be underpowered to detect smaller and/or not well-separated classes (Lubke, 2010).

None of these limitations detracts from the usefulness of mixture modeling or the scientific value of the emergent latent class structure for characterizing the population heterogeneity of interest. However, any interpretation must be made with these limitations in mind and care must be taken not to reify the resultant latent classes or to make claims about proof of their existence.

Future Directions

In the historical overview of mixture modeling at the beginning of this chapter, I remarked on the rapid expansion in the statistical theory (model specification and estimation), software implementation, and applications of mixture modeling in the last 30 years. And the evolution of mixture modeling shows no signs of slowing. There are numerous areas of development in mixture modeling, and many investigations are currently underway. Among those areas of development are: measures of overall goodness-of-fit, individual fit indices, graphical residual diagnostics, and assumption-checking *post hoc* analyses—particularly for mixture models with continuous indicators and factor mixture models; Bayesian estimation and mixture model selection; class enumeration processes for multilevel mixture models with latent class variable on two or more levels; missing data analysis—particularly maximum likelihood approaches and multiple imputation approaches for non-ignorable missingness related to latent class membership; detection procedures for differential item functioning in latent class measurement models; multistage and simultaneous approaches for analyzing predictors and distal outcomes of latent class membership including multiple imputation of latent class membership by way of plausible values from Bayesian estimation techniques; integration of causal inference techniques such as propensity scores and principal stratification with mixture models; and informed study design, including sample size determination, power calculations, and item selection. In addition to these more specific areas of methods development, the striking trend of extending other statistical models by integrating or overlaying finite mixtures will surely continue and more hybrid models are likely to emerge. Furthermore, there will be advancing substantive areas, yielding new kinds of data, for which mixture modeling may prove invaluable—for example, genotypic profile analysis of single nucleotide polymorphisms. And although it is difficult to predict which area of development will prove most fruitful in the coming decades, it is certain that mixture modeling will continue to play an increasingly prominent role in ongoing empirical quests to describe and explain general patterns and individual variability in social science phenomena.

List of Abbreviations

ANOVA	Analysis of variance (ANCOVA—Analysis of covariance)
AvePP	Average posterior class probability

AWE	Approximate weight of evidence criterion
BF	Bayes factor
BIC	Bayesian information criterion
CACE	Complier average causal effect
CAIC	Consistent Akaike information criterion
CFA	Confirmatory factor analysis (EFA—Exploratory factor analysis)
cmP	Correct model probability
df	Degree(s) of freedom
DIF	Differential item functioning
E_K	Entropy
EM	Expectation-maximization algorithm
GMM	Growth mixture model
IRT	Item response theory
LCA	Latent class analysis
LCCA	Latent class cluster analysis
LCR	Latent class regression
LL	Log likelihood
LPA	Latent profile analysis
LR	Likelihood ratio (LRT—Likelihood ratio test; LRTS—LRT statistic; LMR-LRT—Lo, Mendell, & Rubin LRT; BLRT—bootstrapped LRT)
LSAY	Longitudinal Study of American Youth
LTA	Latent transition analysis
MAR	Missing at random (MCAR—missing completely at random)
mcaP	Modal class assignment proportion
ML	Maximum likelihood (MLE—Maximum likelihood estimate; FIML—Full information maximum likelihood)
MVN	Multivariate normal distribution
npar	Number of free parameters
OCC	Odds of correct classification ratio
OR	Odds ratio
RMM	Regression mixture model
SIC	Schwarz information criterion
SSPG	Steady state plasma glucose

Appendix

A technical appendix with Mplus syntax and supplementary Excel files for tabulating and constructing graphical summaries of modeling results is available by request from the chapter author.

Author Note

The author would like to thank Dr. Karen Nylund-Gibson and her graduate students, Amber Gonzalez and Christine Victorino, for research and editorial support. Correspondence concerning this chapter should be addressed to Katherine Masyn, Graduate School of Education, Harvard University, Cambridge, MA 02138. E-mail: katherine_masyn@gse.harvard.edu.

References

Agresti, A. (2002). *Categorical data analysis.* Hoboken, NJ: John Wiley & Sons.

Arminger, G., Stein, P., & Wittenburg, J. (1999). Mixtures of conditional and covariance structure models. *Psychometrika, 64*(4), 475–494.

Asparouhov, T., & Muthén, B. (2004). *Maximum-likelihood estimation in general latent variable modeling.* Unpublished manuscript.

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G.R. Hancock & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27–51). Charlotte, NC: Information Age Publishing.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association,* 92(440), 1375–1386.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49,* 803–821.

BeLue, R., Francis, L. A., Rollins, B., & Colaco, B. (2009). One size does not fit all: Identifying risk profiles for overweight in adolescent population subsets. *The Journal of Adolescent Health, 45*(5), 517–524.

Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly,* 52(3), 601–632.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345–370.

Bray, B. C. (2007). *Examining gambling and substance use: Applications of advanced latent class modeling techniques for cross-sectional and longitudinal data.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park.

Christie, C. A., & Masyn, K. E. (2010). Latent profiles of evaluators' self-reported practices. *The Canadian Journal of Program Evaluation,* 23(2), 225–254.

Chung, H., Flaherty, B. P., & Schafer, J. L. (2006). Latent class logistic regression: Application to marijuana use and attitudes among high school seniors. *Journal of the Royal Statistical Society: Series A, 169*(4), 723–743.

Chung, H., Loken, E., & Schafer, J. L. (2004). Difficulties in drawing inferences with finite-mixture models. *The American Statistician, 58*(2), 152–158.

Clogg, C. C. (1977). *Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users.* University Park, PA: Population Issues Research Center, Pennsylvania State University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, L. M., Graham, J. W., Long, J. D., & Hansen, W. B. (1994). Crossvalidation of latent class models of early substance use onset. *Multivariate Behavioral Research, 29*(2), 165–183.

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.

Crissey S. R. (2005). Race/ethnic differences in the marital expectations of adolescents: The role of romantic relationships. *Journal of Marriage and the Family, 67*(3), 697–709.

Croon, M. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology, 43*, 171–192.

Croon, M. (2002). Ordering the classes. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 137–162). Cambridge: Cambridge University Press.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147–167.

Dayton, C.M. (1998). *Latent class scaling analysis.* Thousand Oaks, CA: Sage.

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

DeCarlo, L. T. (2005). A model of rater behavior in essay grading based on signal detection theory. *Journal of Educational Measurement, 42*(1), 53–76.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, 39*, 1–38.

Desarbo, W. S., Jedidi, K., & Sinha, I. (2001). Customer value analysis in a heterogeneous market. *Strategic Management Journal, 22*(9), 845–857.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63*, 227–253.

Draney, K., Wilson, M., Gluck, J., & Spiel, C. (2008). Mixture models in a developmental context. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 199–216). Charlotte, NC: Information Age Publishing.

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211–218.

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford Press.

Feldman, B., Masyn, K., & Conger, R. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical data. *Developmental Psychology, 45*(3), 652–676.

Formann, A. K. (1982). Linear logistic latent class analysis. *Biometrical Journal, 24*, 171–190.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476–486.

Fraley, C., and Raftery, A.E. (1998). *How many clusters? Which clustering method? Answers via model-based cluster analysis. The Computer Journal, 41*(8), 578–588.

Garrett, E. S., & Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics, 56*, 1055–1067.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*,85, 398–409.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 3–55). Cambridge: Cambridge University Press.

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205–220.

Hagenaars, J.A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods & Research, 26*, 436–486.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis.* Cambridge; New York: Cambridge University Press.

Hamil-Luker, J. (2005). Trajectories of public assistance receipt among female high school dropouts. *Population Research and Policy Review, 24*(6), 673–694.

Hancock, G. R., & Samuelson, K. M. (Eds.). (2008). *Advances in latent variable mixture models.* Charlotte, NC: Information Age Publishing.

Heinen, D. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage Publications.

Henry, K. & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling*, 17, 193–215.

Herzberg, A. M., & Andrews, D. F. (1985). *Data: A collection of problems from many fields for the student and research worker.* Brooklyn, NY: Springer.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*,11(1), 36–53.

Horn, J. L. (2000) Comments on integrating person-centered and variable-centered research on problems associated with the use of alcohol. *Alcoholism: Clinical and Experimental Research, 24*(6) 924–930.

Huang, G. H., & Bandeen-Roche, K. (2004). Latent class regression with covariate effects on underlying and measured variables. *Psychometrika, 69*(1), 5–32.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16(1), 39–59.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, 7, 178–193.

Kass, R. E., & and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(434), 928–934.

Kimmel, L. G., & Miller, J. D. (2008). The longitudinal study of American youth: Notes on the first 20 years of tracking and data collection. *Survey Practice*, December, 2008.

Langeheine, R., & van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 304–341). Cambridge, UK: Cambridge University Press.

Lanza, S. T., Collins, L. M., Schafer, J. L., & Flaherty, B. P. (2005). Using data augmentation to obtain standard

errors and conduct hypothesis tests in latent class and latent transition analysis. *Psychological Methods, 10*, 84–100.

Lanza, S. T., Dziak, J. J., Huang, L., Xu, S., & Collins, L. M. (2011). *PROC LCA & PROC LTA User's Guide Version 1.2.6.* University Park: The Methodology Center, Penn State.

Laursen, B. & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly, 52*, 377–389.

Lazarsfeld, P., & Henry, N. (1968). *Latent structure analysis.* New York: Houghton Mifflin.

Little, R. J. & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley and Sons.

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778.

Lubke, G. H. (2010). Latent variable mixture modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences.* New York, NY: Routledge.

Magnusson, D. (1998). The logic and implications of a person-oriented approach. In R. B. Cairns, L. R. Bergman, & J. Kagan (Eds.), *Methods and models for studying the individual* (pp. 33–64). Thousand Oaks, CA: Sage.

Markon, K. E, & Krueger, R. F. (2005). Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry, 62*(12), 1352–1359.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*(2), 191–225.

Masyn, K., Henderson, C., & Greenbaum, P. (2010). Exploring the latent structures of psychological constructs in social development using the Dimensional-Categorical Spectrum. *Social Development, 19*(3), 470–493.

Masyn, K. E. (2009). Discrete-time survival factor mixture analysis for low-frequency recurrent event histories. *Research in Human Development, 6*, 165–194.

McCutcheon, A. C. (1987). *Latent class analysis.* Beverly Hills, CA: Sage.

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

Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (2000). *Longitudinal study of American youth: User's manual.* Chicago, IL: Northwestern University, International Center for the Advancement of Scientific Literacy.

Muthén & Muthén. (September, 2010). *Bayesian analysis of latent variable models using Mplus.* (Technical Report, Version 4). Los Angeles, CA: Asparouhov, T., & Muthén, B. (2010).

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling. *Psychological Methods, 8*, 369–377.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing.

Muthén, B. & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors, 31*, 1050–1066.

Muthén, B. & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton: Chapman & Hall/CRC Press.

Muthén, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2011). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16*, 17–33.

Muthén, B., & Muthén, L. K. (1998–2011). *Mplus* (Version 6.11) [Computer software]. Los Angeles, CA: Muthén & Muthén.

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

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139–157.

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

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

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.

Nylund, K. L., & Masyn, K. E. (May, 2008). *Covariates and latent class analysis: Results of a simulation study.* Paper presented at the meeting of the Society for Prevention Research, San Francisco, CA.

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.

Patterson, B., Dayton, C. M., & Graubard, B. (2002). Latent class analysis of complex survey data: application to dietary data. *Journal of the American Statistical Association*, 97, 721–729.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London (Series A), 185*, 71–110.

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

Petras, H., Masyn, K., & Ialongo, N. (2011). The developmental impact of two first grade preventive interventions on aggressive/disruptive behavior in childhood and adolescence: An application of latent transition growth mixture modeling. *Prevention Science*, 12, 300–313.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika, 69*, 167–190.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103–124.

Reaven. G. M., & Miller, R. G. (1979). An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologia, 16*(1), 17–24.

Reboussin, B. A., Ip, E. H., & Wolfson, M. (2008), Locally dependent latent class models with covariates: an application to under-age drinking in the USA. *Journal of the Royal Statistical Society: Series A, 171*, 877–897.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

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

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models.* London: Chapman Hall.

Soothill, K., Francis, B., Awwad, F., Morris, Thomas, & McIllmurray. (2004). Grouping cancer patients by psychosocial needs. *Journal of Psychosocial Oncology, 22*(2), 89–109.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Tan, W. Y., & Chang, W. C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *Journal of the American Statistical Association, 67*(339), 702–708.

Thorndike, R. L. (1953). Who belong in the family? *Psychometrika, 18*(4), 267–276.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* Chichester, UK: John Wiley & Sons.

Tueller, S. J., Drotar, S. & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling, 18*, 110–131.

Van Horn, M. L., Jaki, T., Masyn, K., Ramey, S. L., Smith, J. A., & Antaramian, S. (2009). Assessing differential effects: Applying regression mixture models to identify variations in the influence of family resources on academic achievement. *Developmental Psychology, 45*(5), 1298–1313.

Vermunt, J. K. (1999). A general non parametric approach to the analysis of ordinal categorical data. *Sociological Methodology*,29, 197–221.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge: Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2005). *Latent GOLD 4.0 users's guide.* Belmont, MA: Statistical Innovations.

Vermunt, J. K., & McCutcheon, A. L. (2012). *An introduction to modern categorical data analysis.* Thousand Oaks, CA: Sage.

Vermunt, J.K., Van Ginkel, J. R., Van der Ark, and L. A., Sijtsma K. (2008). Multiple imputation of categorical data using latent class analysis. *Sociological Methodology, 33*, 369–297.

von Davier, M., & Yamamoto, K. (2006). Mixture distribution Rasch models and hybrid Rasch models. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 99–115). New York: Springer-Verlag.

Weldon, W. F. R. (1893). On certain correlated variations in carcinus moenas. *Proceedings of the Royal Society of London*, 54, 318–329.

Yang, M. & Yu, N. (2005). Estimation of parameters in latent class models using fuzzy clustering algorithms. *European Journal of Operational Research, 160*, 515–531.

Yang, X. D., Shaftel, J., Glasnapp, D., & Poggio, J. (2005). Qualitative or quantitative differences? Latent class analysis of mathematical ability for special education students. *Journal of Special Education, 38*(4), 194–207.

Yung, Y. F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika, 62*, 297–330.