ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Review; 2024; 15(2); 79-93

A Systematic Review of Factor Mixture Model Applications

Sedat ŞEN*   Allan S. COHEN**

Abstract

In this study, a systematic review was conducted on peer-reviewed articles of factor mixture model (FMM) applications. A total of 304 studies were included with 334 applications published from 2003–2022. FMM was mostly used in these studies to detect latent classes and model heterogeneity. Most of the studies were conducted in the U.S. with samples including students, adults, and the general population. The average sample size was 3,562, and the average number of items was 17.34. Measurement tools containing mostly Likert type items and measuring structures in the field of psychology were used in these FMM analyses. Most FMM studies that were reviewed were applied with maximum likelihood estimation methods as implemented in Mplus software. Multiple fit indices were used, the most common of which were AIC, BIC, and entropy. The mean numbers of classes and factors across the 334 applications were 2.96 and 2.17, respectively. Psychological and behavioral disorders, gender, and age variables were mostly the focus of these studies and included use of covariates in these analyses. As a result of this systematic review, the trends in FMM analyses were better understood.

Keywords: mixture models, factor mixture model, systematic review

Introduction

The factor mixture model (FMM; Muthén & Shedden, 1999) is a combination of common factor and latent class models (Lubke & Muthén, 2005). The FMM is a type of mixture model and comprises a family of statistical models useful for evaluating data in which there may be multiple latent variables that underlie the observed variables. FMMs can provide a powerful tool for analyzing data where multiple unobserved variables may be influencing the observed variables. As with other latent variable mixture models, FMM is also a flexible analytical method that enables researchers to explore research problems about data patterns and assess the degree to which observed patterns are related to important variables (Berlin et al., 2013). Typically, FMMs attempt to estimate latent classes within a sample based on the response patterns (i.e., the observed variables) respondents have made to a given set of items. Thus, they are considered "person-centered" statistical methods as the detection of latent classes is based on person characteristics. FMMs are often used to explain unobserved population heterogeneity as well as to detect latent classes by relaxing the assumption that all respondents in the sample are drawn from the same population. The latent classes may differ either qualitatively or quantitatively or both.

Unlike latent class analysis (LCA; Lazarsfeld & Henry, 1968) and exploratory factor analysis (EFA; Spearman, 1904), FMMs have the flexibility to model hybrids of continuous latent variables (factors) and categorical latent variables (latent classes). Thus, FMMs are also sometimes known as hybrid latent variable models. In FMM, the factor analysis part seeks to uncover shared latent content (i.e., factors) among the observable variables, the latent class analysis part is intended to identify latent subgroups or latent classes of a study population. In addition to FMM, several models with different names have been developed in the literature based on combining categorical and continuous latent variables. These fall

* Assoc. Prof. Dr., Harran University, Faculty of Education, Şanlıurfa-Turkey, sedatsen@harran.edu.tr, ORCID ID: 0000-0001-6962-4960
** Professor Emeritus, The University of Georgia, Department of Educational Psychology, Georgia-USA, acohen@uga.edu, ORCID ID: 0000-0002-8776-9378

To cite this article:
Şen, S., & Cohen, A. S. (2024). A systematic review of factor mixture model applications. *Journal of Measurement and Evaluation in Education and Psychology, 15*(2), 79-93. https://doi.org/10.21031/epod.1423427

Received: 21.01.2024
Accepted: 6.06.2024

under the heading of mixture item response theory (mixture IRT; Mislevy & Verhelst, 1990; Rost, 1990), and latent class factor analysis (LCFA; Magidson & Vermunt, 2001).

The combination of both categorical and continuous latent variables enables the structure to be simultaneously categorical and dimensional, making the FMM useful for researchers (see Clark et al., 2013). This is because the FMM allows for the simultaneous modeling of latent class membership and the distribution both within and between latent classes. One of the main advantages of FMMs is their capability to handle multiple types of data within the same model, and to simultaneously model the relationships between the latent variables and the observed variables. This makes FMMs particularly useful for data where the relationships between the variables are complex and multifaceted.

The general FMM equation can be written as follows:

$$\mathbf{Y}_{ik} = \boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}_{ik} + \boldsymbol{\varepsilon}_{ik}, \quad (1)$$

where the *i* indicates persons and *k* indicates latent classes to which individuals are assigned. The $\mathbf{Y}_{ik}$ matrix in Equation (1) includes the response sets of the i-th individual in latent class *k*. Parameters $\boldsymbol{\tau}_k$, $\boldsymbol{\Lambda}_k$, $\boldsymbol{\eta}_{ik}$, and $\boldsymbol{\varepsilon}_{ik}$ represent the intercept vector, the factor loading vector, the vector of an individual's factor scores, and the residual, respectively. In addition, residuals are assumed to be normally distributed with a mean of zero and a variance of $\boldsymbol{\Theta}_\mathbf{k}$.

The FMM assigns each individual to a latent class based on posterior probabilities (a.k.a. class probabilities). Once individuals have been classified, the FMM allows for individual intra-class differences by estimating a factorial model for each class (Clark et al., 2013). Class-specific item thresholds and slopes can be estimated in addition to factor variance(s), covariances, and mean(s). The parameters of FMMs can be estimated using frequentist (e.g., maximum likelihood) or Bayesian (e.g., Markov chain Monte Carlo) methods.

There are different model labels, however, according to the restrictions applied within FMM itself (Clark et al., 2013). Clark et al. identified four different types of FMMs, labeled as FMM-1, FMM-2, FMM-3, and FMM-4. Each of these has different parameter restrictions and measurement invariance assumptions. The most restricted FMM is FMM-1 and is equivalent to the LCFA. The least restrictive is FMM-4. With respect to FMM-1, the factor mean is the only parameter that changes across classes. The item thresholds and factor loadings are constrained to be equal across classes. In addition, the factor covariance matrix is fixed at zero in FMM-1 in order to assign the same factor scores to all individuals within a single latent class. In FMM-2, factor means, factor variances, and covariances are freely estimated. This model is also known as a mixture factor analysis (McDonald, 2003; Yung, 1997). FMM-1 and FMM-2 incorporate strict factorial invariance (Masyn et al., 2010). FMM-3 allows the factor covariance matrix and item thresholds to change across classes, but holds the factor loadings to equality, and fixes factor means to zero for identification purposes. Finally, in FMM-4, the factor means are fixed to zero and all other elements (intercepts, loadings, and factor covariances) can vary across classes.

Although there are examples of FMM used for confirmatory purposes, FMM is an exploratory model. That is, the model is used to estimate different solutions and the numbers of factors and latent classes are determined based on the model that best explains the data. Researchers typically analyze different models by changing the number of factors and increasing the number of classes one by one. The best fitting model is the one among all the candidates that is best fitting, both theoretically and statistically. Reporting all models that fit the data, comparing them, and outlining the decision-making process is an important part of the model selection process. Model selection can be challenging as the statistical results and the content-based theory do not always agree on the same number of latent classes.

There are several fit indices that can be used for model selection, including information criteria (IC) indices and likelihood ratio (LR) tests (McLachlan & Peel, 2000). The traditional LR test, which is appropriate for nested models, cannot be used for FMMs because regularity conditions are not met (see Nylund et al., 2007 and McLachlan & Peel, 2000). Thus, several adjusted versions of LR tests have been developed for use with FMMs. These are the bootstrapped LR test (BLRT; McLachlan & Peel, 2000), Lo-Mendell-Rubin LR test (LMR; Lo et al., 2001), adjusted LR test (aLMR; Lo et al., 2001), and the

Vuong-Lo-Mendell-Rubin LR test (VLMR; Lo et al., 2001). IC indices including Akaike's information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Schwarz, 1978), and extensions of these two indices, consistent AIC (CAIC), corrected AIC (AICc), and sample size adjusted BIC (SABIC), are also used for model selection. Nylund et al. (2007) describe simulation studies on how some of the fit indices perform in FMM selection. Based on the result in Nylund et al., BIC and BLRT are among the best choices for selecting the model with the correct number of latent classes.

An important advantage of FMMs is that covariates, such as gender, age, and education level, can also be added to the model in order to account for the uncertainty in class membership and to validate the FMM (Brown, 2013). Covariate inclusion can be done as either one-step or multiple-step (e.g., two- or three-step) approaches. In the first approach, covariates can be added directly to the FMM model. In the second approach, an unconditional FMM is analyzed first, then latent classes are estimated and the relationship between the latent class memberships and covariates is examined with a regression model. Both approaches have advantages and disadvantages (see Wang, et al., 2022 for a comparison). FMMs have been employed in several disciplines, including psychology, education, sociology, and the health sciences (e.g., Lin & Masse, 2021; Moors et al., 2014; Morin & Marsh, 2015).

Previous reviews on latent variable mixture models (e.g., Killian et al., 2019) have largely focused on other models, including mixture IRT (Sen & Cohen, 2019), latent profile analysis (LPA; Spurk et al., 2020), latent class analysis (LCA; Ulbricht et al., 2018) and growth mixture modeling (GMM; Baron et al., 2017). A review of the literature reveals that there are some review studies on FMMs. A few investigations have reported small-scale studies of the FMM. In this regard, Hofmans et al. (2020) presented an overview of four studies of careers, career counseling, and vocational behavior that used FMM analysis. Krawietz and Pett (2023) have conducted a systematic review of 95 studies including latent variable mixture modeling in communication scholarship. Kim et al. (2023) have recently conducted a systematic review of 76 FMM applications. However, this search was based only on studies in the PsycInfo database and did not select keywords that would include all possible factor mixture models such as LCFA. As a result, there is a lack of review studies that include all FMM applications. This present study fills this gap by systematically reviewing 334 applications of FMMs published in peer-reviewed journals across a variety of databases. This study (i) reviews existing applications of FMMs, (ii) to improve understanding of FMMs and how they are applied. The findings in this review can improve our understanding of how FMMs are applied. Inconsistent and incomplete reporting practices can set a bad example for any new study that plans to use FMMs. Thus, it is necessary to identify and summarize best practices to ensure the quality of FMM applications. Knowledge gained from this review will provide useful information for future researchers who may use FMMs in their studies.

Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) standards were followed for conducting this study. Methods used in the review are presented below.

Inclusion/Exclusion Criteria

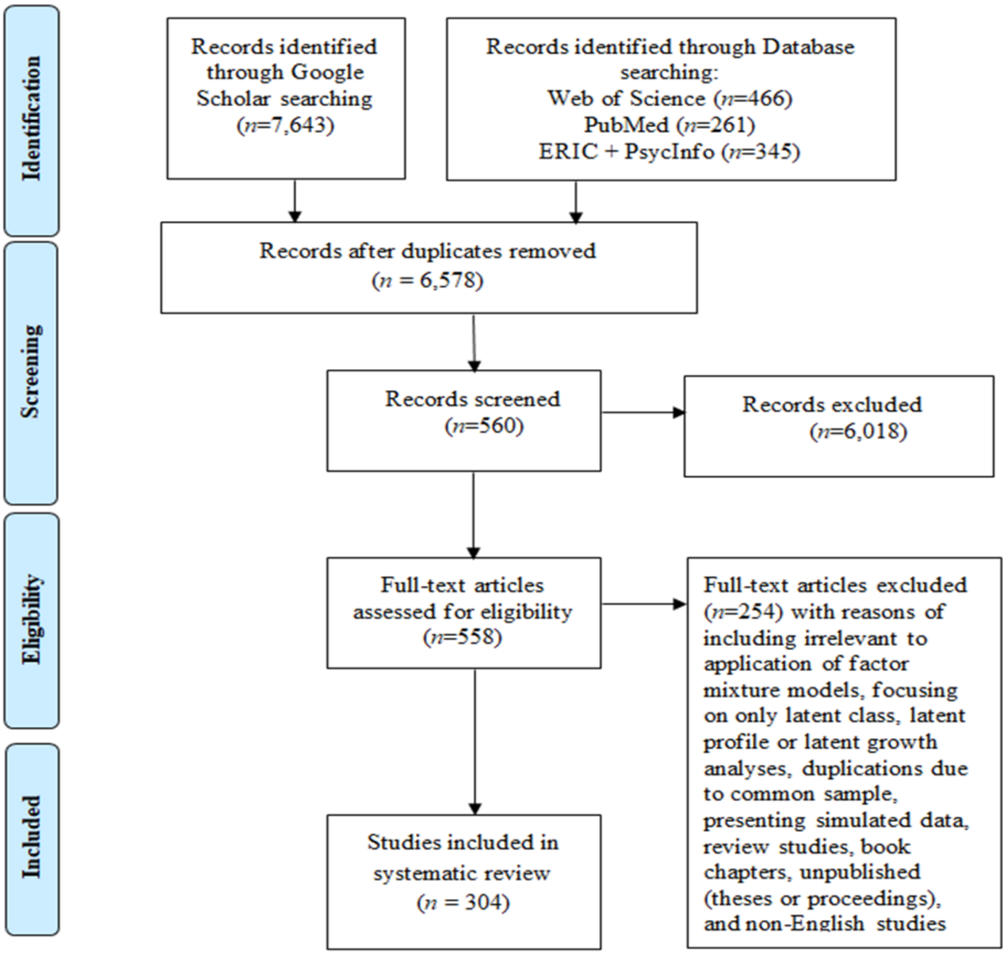
The studies included in this systematic review were screened to include only those published in peer-reviewed journals in English before 2023. Studies needed to apply at least one FMM to be included in this review. Studies which included only simulations, used simulated data, contained use of only latent class, latent profile, or latent growth analysis, unpublished studies (such as theses or proceedings), book chapters, review studies, and studies not written in English also were excluded. However, the application part of the methodology studies that include application is also included in the review. Mixture IRT based studies were not included in this review as this review was written within the framework of factor analysis, and a systematic review has already been reported on mixture IRT studies (see Sen & Cohen, 2019).

Search Strategy

The review strategy used in this study is reported below. The search was conducted on January 1, 2023 using the Google Scholar search engine and included several databases including the Web of Science, PubMed, ERIC, and PscyInfo. Different labels were used for factor mixture model, including factor mixture analysis, mixture factor model, mixture factor analysis, latent class factor model, and latent class factor analysis. To cover all of these terms in the search process, the following search strings were used: "factor mixture*" OR "mixture factor*" OR "latent class factor*". The first search yielded 7,643 studies on Google Scholar, 466 studies on the Web of Science, 266 studies on PubMed, and 345 studies on the ERIC and PsycInfo databases. Duplicate articles (*n*=1,065) were deleted. From the remaining 6,578 unique studies, all articles that did not use any form of factor mixture models (*n*=6,018) were excluded. Among the remaining 558 full studies, 254 were excluded as being irrelevant to the application of factor mixture models, focusing on only latent class, latent profile, latent growth analyses, duplications due to a common sample, presenting only simulated data, review studies, book chapters, unpublished (theses or proceedings), or studies written in a language other than English. The final sample consisted of 304 peer-reviewed articles. Some studies included more than one factor mixture model analysis applied to different samples. This resulted in 334 applications from the 304 studies. A PRISMA (Moher et al., 2009) diagram showing each step of this search process is presented in Figure 1. The references of the studies included in the review can be requested from the first author.

Figure 1

PRISMA flow diagram of study selection



Data coding and analysis

Each study included in the review was coded for study and model characteristics using the following coding scheme: (a) characteristics of the study (author(s), year of publication, and journal title); (b) country and region of application; (c) construct measured; (d) use of FMM; (e) FMM type; (f) other model types applied before FMM analysis; (g) sample size and number of items; (h) population type; (i) model fit statistics; (j) number of classes checked and decided; (k) software package; (l) estimation type; (m) covariate used in FMMs; (n) missing data handling method. While creating this coding scheme, some previous review studies (Killian et al., 2019; Sen & Cohen, 2019; Spurk et al., 2020; Ulbricht et al., 2018) were taken as references. Some of these variables are coded continuously, and some are coded categorically. The percentage and frequencies of the categorical variables are reported, and also the arithmetic mean and standard deviations of the continuous variables.

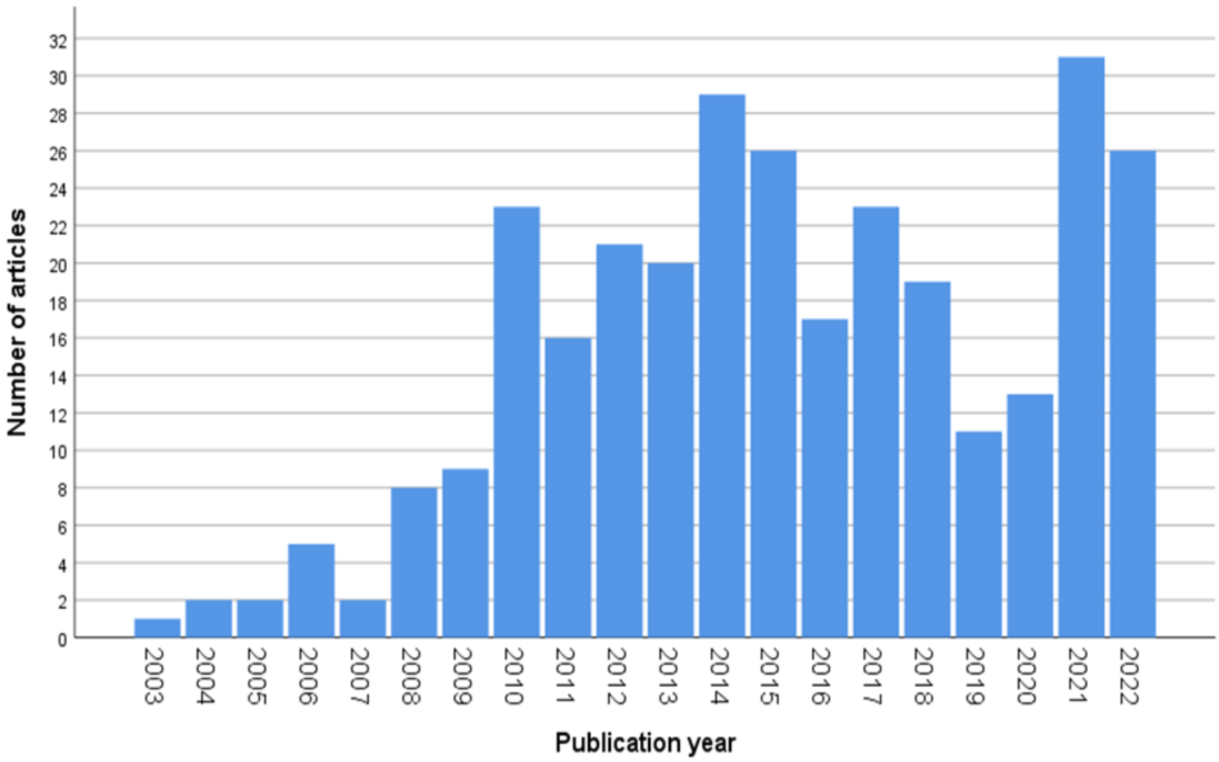
Results

Characteristics of the study

In this section we review information about the publication year of the 304 studies, the country/region in which they were published, and the journals in which they were published. The papers in our review were published in 194 different journals between 2003 and 2022 (see Figure 2). The distribution of published studies by year is shown in Figure 2. The year in which the fewest studies were published was 2003 (*n*=1), and the year in which the most studies were published was 2021 (*n*=31). The average number of studies per year was 15.2. The fact that the number of studies is higher in the last 10 years than in the previous decade is likely evidence that studies using FMM are on the rise.

Figure 2

FMM applications published between 2003 and 2022



Among journals with the most FMM studies published are *Structural Equation Modeling: A Multidisciplinary Journal* (*n*=11), *Psychological Assessment* (*n*=8), *Frontiers in Psychology* (*n*=7), *Psychological Medicine* (*n*=6), *Journal of Personality Assessment* (*n*=5), *Journal of Abnormal Psychology* (*n*=5), *Journal of Anxiety Disorders* (*n*=5), *Psychiatry Research* (*n*=5), *Plos One* (*n*=5), and *Social Indicators Research* (*n*=4). The studies considered in this review were conducted in 36 different countries. The country in which most FMM studies were conducted was the U.S. (*n*=145) followed by Australia (*n*=20) and the Netherlands (*n*=19). Fourteen FMM studies were conducted in Germany, 13 FMM studies were conducted in Canada, and 12 FMM studies were conducted in Finland and Italy. Seventeen studies were conducted with multinational samples. Six percent of the studies did not mention the region of the study. Overall, 48.8% FMM applications were conducted in North America and 32% in Europe. The percentages of studies conducted on other continents are as follows: Asia (7.2%), Oceania (6.3%), South America (1.8%), and Africa (0.6%).

Table 1

Overview of study and FMM characteristics of the reviewed studies (k=304 with 334 applications)

Study Characteristics	***N* (%), or *M* (SD), Median**
Average number of studies per year	*M*=15.2
Region	
North America	163 (48.8%)
Europe	107 (32.0%)
Asia	24 (7.2%)
Oceania	21 (6.3%)
South America	6 (1.8%)
Not reported or unclear	6 (1.8%)
Africa	2 (0.6%)
Multi-continent	5 (1.5%)
Sample description	
Students	63 (18.9%)
Adults	53 (15.9%)
Patients	38 (11.4%)
General population	33 (9.9%)
Children	24 (7.2%)
Adolescents	22 (6.6%)
Employees	19 (5.7%)
Twins	13 (3.9%)
Young adults & Teens	8 (2.4%)
Others	72 (21.6%)
Sample size, median	*M*=3562.42, *SD*=7662.53, Med=888
Number of items/indicators	*M*=17.34, *SD*=17.71, Med=12
Item type	
Likert	209 (62.6%)
Dichotomous/binary	78 (23.4%)
Continuous	24 (7.2%)
Mixed	11 (3.3%)
Not reported/unclear	10 (3.0%)
0-10 Rating scale	2 (0.6%)
Modeling approach	
Exploratory	275 (82.3%)
Confirmatory	59 (17.7%)
FMM Type	
FMM-1/LCFA	109 (32.6%)
FMM-2	46 (13.8%)
FMM-3	33 (9.9%)
FMM-4	19 (5.7%)
ML-FMM	10 (3.0%)
Not clear	155 (46.4%)

Table 1 Continued

Study Characteristics	*N* (%), or *M* (SD), Median
Models estimated before FMM analyses	
CFA	141 (42.2%)
LCA	109 (32.6%)
EFA	79 (23.7%)
LPA	23 (6.9%)
ESEM	15 (4.5%)
SEM	13 (3.9%)
IRT	13 (3.9%)
LTA	11 (3.3%)
Others	14 (4.2%)
Not reported	87 (26.0%)
Software package	
Mplus	259 (77.5%)
Latent GOLD	37 (11.1%)
R packages	12 (3.6%)
Others (mdltm, Mx, OpenMx, WINBUGS)	8 (2.4%)
Not reported	19 (5.7%)
Type of estimator	
Frequentist	245 (73.4%)
Bayesian	1 (0.3%)
Not reported/unclear	88 (26.3%)
Missing data methods	
FIML	68 (20.4%)
Pairwise/Listwise/Excluded	58 (17.4%)
Imputation (single, multiple, nonparametric, mean, recorded as zero)	12 (3.6%)
Complete data	16 (4.8%)
Not reported	180 (53.9%)
Number of classes in final models	*M*=2.96, *SD*=1.28, Med=3
Number of factors in final models	*M*=2.17, *SD*=1.48, Med=2
Class percentages are reported in final models	285 (85.3%)
Classes are labeled in final models	208 (62.3%)
Profile plots are presented	129 (38.6%)
Fit indices reported for model selection	322 (96.4%)
Multiple fit values applied for model selection	286 (85.6%)
Interpretability and theory considered for model selection	116 (34.7%)
AIC/BIC difference used for model selection	35 (10.5%)
Elbow plots used for model selection	9 (2.7%)
Class sizes considered for model selection	42 (12.6%)
Applied model fit values	
BIC	296 (88.6%)
AIC	180 (53.9%)
Entropy	165 (49.4%)
SABIC	161 (48.2%)
LMR-LRT	99 (29.6%)
BLRT	81 (24.3%)
LRT/aLRT	52 (15.6%)
VLMR	44 (13.2%)
CAIC	20 (6.0%)
Others	61 (18.3%)
Most frequent covariates included	
Psychological and behavioral disorders	117 (35.0%)
Gender	104 (31.1%)
Age	92 (27.5%)
Education level	39 (11.7%)
Ethnicity	20 (6.0%)
Marital status	18 (5.4%)
Income	12 (3.6%)
SES	11 (3.3%)
BMI	10 (3.0%)
Employment status	7 (2.1%)
Language	4 (1.2%)

Table 1 Continued

Study Characteristics	*N* (%), or *M* (SD), Median
Statistical analyses after FMM	
Chi-square test	48 (14.4%)
Regression	39 (11.7%)
Logistic regression	34 (10.2%)
t-test or non-parametric versions	26 (7.8%)
ANOVA	26 (7.8%)
Correlation	13 (3.9%)
Multiple comparison/Mean comparison	13 (3.9%)
MANOVA	11 (3.3%)
R3STEP	9 (2.7%)
ANCOVA	7 (2.1%)
Odds ratio	6 (1.8%)
ROC Curves	5 (1.5%)
Wald test	4 (1.2%)
Cross-tabs	4 (1.2%)
Kappa classification agreement	4 (1.2%)
SEM/ESEM	4 (1.2%)
Fisher's exact test	3 (0.9%)
Others	33 (9.9%)
Not reported	134 (40.1%)
Use of FMM	
Identifying the latent classes/clusters/profiles/patterns	157 (47%)
Investigating the latent structure	89 (27%)
Model comparison	17 (5%)
Analyzing population heterogeneity	14 (4%)
Determining the best fitting model	14 (4%)
Exploring the response process/styles	11 (3%)
Examining the validity	6 (2%)
Testing measurement invariance/equivalence	6 (2%)
Others	20 (6%)

Use of FMM

Based on preliminary analysis, FMM applications in Table 1 were divided into nine subcategories. These included identifying the latent classes/clusters/profiles/patterns (47%), investigating the latent structure (27%), analyzing population heterogeneity (4%), model comparison (5%), determining the best fitting model (4%), exploring the response process or response styles (3%), examining the validity (2%), and testing measurement invariance or equivalence (2%). The remaining six percent of the studies included applications examining measurement assumptions, evaluating the appropriateness of latent class factor analysis, model building, investigating the covariate effect, investigating Spearman's law of diminishing returns, and testing for performance and structural differences.

Outcome Measured

Different topics covered in these studies included alcohol use disorder (n=14), posttraumatic stress disorder (n=10), anxiety sensitivity (n=7), panic attack symptoms (n=7), schizotypal personality disorders (n=5), tobacco dependence (n=5), autism spectrum disorder (n=5), borderline personality disorder (n=4), life satisfaction (n=4), job stress and job resources (n=4), mathematics (n=4), and reading (n=4).

Number of items and item type

There were five different item types in the studies reviewed. As can be seen in Table 1, the most frequent item type was the Likert item (n=209, 62.6%), followed by the dichotomous (n=81, 23.4%), the continuous (n=24, 7.2%), mixed (n=11, 3.3%), and rating scale items (n=2, 0.6%). Item type was not specified in 3% (n=10) of the studies. The number of items (or indicators) used varied greatly from k=1

(Ma et al., 2022) to *k*=165 (Grove et al., 2015). Only four studies used measurement tools with more than 100 items. The average number of items across 334 applications was 17.34, with a median of 12, and an SD of 17.71.

Sample size and population type

The studies reviewed included samples of participants from a variety of populations. These populations were grouped into 10 categories. Frequencies and percentages for each category are presented in Table 1. Populations in the reviewed studies were identified as students (18.9%), adults (15.9%), patients (11.4%), general population (9.9%), children (7.2%), adolescents (6.6%), employees (5.7%), twins (3.9%), and young adults and teens (2.4%). Almost twenty-two percent of the studies included different types of participants including veterans, soldiers, dancers, gamers, athletic performers, educators, households, immigrants, job applicants, current smokers or drinkers, and individuals with autism spectrum disorder.

Sample size is an important consideration in applications of FMM. Of the studies reviewed, sample sizes varied greatly from *N*=50 to *N*=261,747. Apart from four studies with very large sample sizes (261747, 212674, 177480, and 116543), the remaining studies had sample sizes of less than 50,000. Only two studies used samples of fewer than 100 individuals, and fifteen studies had sample sizes between 100 and 200. The mean sample size across 330 applications was 3,562, with a median of 888, and an SD=7,662.53, excluding the four outlier studies with the very large sample sizes.

Missing data

Researchers often have to deal with missing data. When this occurs, there are a number of different methods that can sometimes be used to deal with missing data. These include data deletion (pairwise or listwise), imputation (single or multiple), or FIML (full information maximum likelihood) methods (see Enders, 2022). In handling missing data, most studies (*n*=68, 20.4%) preferred the FIML estimation method. Missing data were excluded or deleted pairwise or listwise in 58 studies (17.4%). Imputation was used in 12 studies (3.6%), including single imputation, multiple imputation, nonparametric imputation, mean imputation, and recording missingness as zero. Only 16 of the studies (4.8%) reviewed reported using a complete data set. The remaining 180 studies with missing data (53.9%) did not report how missing data was addressed.

Analyses applied before FMM application

Clark et al. (2013, p. 691) recommend the following for the initial step (Step 0) of the FMM analysis: "*Fit latent class analysis and factor analysis models for later comparison and to determine the ending point combination of number of class and factors when fitting factor mixture models*". Additional analyses prior to the FMM included confirmatory factor analysis (CFA; *n*=141, 42.2%) and LCA (*n*=109, 32.6%), which were used in most often, followed by EFA (*n*=79, 23.7%), LPA (*n*=23, 6.9%), exploratory structural equational model (ESEM; *n*=15, 4.5%), SEM (*n*=13, 3.9%), IRT (*n*=13, 3.9%), LTA (*n*=11, 3.3%), and others (*n*=14, 4.2%). In 87 studies (26.0%), the type of analysis conducted before FMM was not reported. More than one preliminary analysis was performed in 144 (43.1%) of the studies. EFA and LCA were used together in 50 (15.0%) studies, CFA and LCA were used together in 39 (11.7%) studies, and EFA-CFA was used together in 15 (4.5%) studies.

Modeling strategy

An exploratory approach to determine the number of latent classes was used in 275 of the 334 studies reviewed (82.3%). For the remaining six studies, a single latent class solution (i.e., 2 or 3 latent classes) was used (1.78%). A model with a single latent class solution was the final model in 53 studies (4.5%). In these latter 53 studies, no information was provided as to the number of different latent class solutions tried. A number of different models were reported in the studies reviewed. Different labels were used for the FMM analysis: 67.9% used the label FMM; 1.4% used the label exploratory FMM. Of those, 32.6% used FMM-1 or latent class factor analysis; 13.8% used FMM-2; 9.9% used FMM-3; and 5.7% used FMM-4. The specific type of FMM used in 46.4% of the studies could not be understood from the information presented. Multilevel extensions of FMM were used in 3.0% of the studies. In the remaining studies, different FMM labels were used including non-normal FMM, twin FMM, multimodality FMM, discrete FMM, constrained FMM, confirmatory FMM, MTMM mixture modeling, repeated measures LCA, and mixtures of factor analyzer.

Estimation methods and software

FMM analyses can be conducted with several statistical software packages, including Mplus (L. K. Muthén & Muthén, 2017), Latent GOLD (Vermunt & Magidson, 2003), mdltm (von Davier, 2006), Mx (Neale et al., 2006), WINBUGS (Spiegelhalter et al., 2003), and R packages such as FactMixtAnalysis (Viroli, 2011) and mclust (Scrucca et al., 2016). Of the papers included in this review, Mplus (*n*=259, 77.5%) was the most commonly reported statistical software package for FMM applications followed by Latent GOLD (*n*=37, 11.1%) and two R packages, FactMixtAnalysis and mclust (*n*=12, 3.6%). Eight (2.4%) studies reported other statistical software packages, including mdltm, Mx, OpenMx, and WINBUGS. There were studies (*n*=19, 5.7%) that did not report the name of the software used.

In the present review, we distinguish between two types of parameter estimation methods for FMMs: frequentist and Bayesian estimation. Most of the included studies (*n*=245; 73.4%) were conducted with frequentist estimation methods. Only one study (Cho et al., 2014) reported using Bayesian estimation. Eighty-eight studies (26.3%) did not report the estimation method used. For the frequentist estimation methods, we differentiate between maximum likelihood (ML), robust maximum likelihood (MLR), full maximum likelihood (FIML), marginal maximum likelihood (MML), linear approximation of ML, and weighted least squares (WLS)-based estimations (WLS, WLSM, and WLSMV). In this review, MLR (*n*=152, 45.5%) was the most commonly reported estimation method for FMM applications, followed by ML (*n*=72, 21.6%), FIML (*n*=9, 2.7%), and MML (*n*=4, 1.2%). Linear approximation ML was used in two studies. Apart from ML based methods, six studies used WLS (*n*=2, 0.6%), WLSM (*n*=1, 0.3%), and WLSMV (*n*=3, 0.9%) estimation methods.

Random starting values are another important issue to consider in ML estimation of FMM parameters due to the local maxima problem. In the present review, random starting values were explicitly mentioned in 87 studies (26.0%). Of these, software defaults were used in one study, and 20 of the reviewed studies reported that “random starting values were used” without reporting the number of random starting values. The number of random starting values used varied greatly in the reviewed studies, ranging from 2 to 200,000. The most commonly used random starting values were 500 (*n*=14), 100 (*n*=11), 5000 (*n*=8), 1000 (*n*=8), and 2000 (*n*=6).

Model fit statistics

In the present review, at least one fit index was reported for model selection in most of the studies (*n*=322, 96.4%). All but 48 studies (*n*=286, 85.6%) reported multiple fit indices with a majority reporting between two and five indices (74.2%). For the studies reporting fit indices, BIC (*n*=296, 88.6%) was the most commonly reported fit index for FMM applications, followed by AIC (*n*=180, 53.9%), entropy (*n*=165, 49.4%), and SABIC (*n*=161, 48.2%). These four indices were followed by likelihood ratio-based tests, including LMR-LRT (*n*=99, 29.6%), BLRT (*n*=81, 24.3%), LRT/aLRT (*n*=52, 15.6%), and VLMR (*n*=44, 13.2%). CAIC was reported in 20 studies (6.0%). Apart from these indices, other indices were also reported in 61 studies (18.2%). These indices include values such as L^2 value (*n*=10), classification error (*n*=9), bivariate residuals (*n*=8), chi-square diff test/the Pearson chi-square (*n*=8), ICL-BIC (*n*=5), AICC (*n*=3), AIC3 (*n*=3), Cressie-Read (*n*=2), delta AIC/BIC (*n*=2), Akaike weight (*n*=2), bootstrap *p* (*n*=2), DIC (*n*=1), ACPP (*n*=1), IC1000 (*n*=1), the log penalty AIC (*n*=1), AWE (*n*=1), and ratio of distance measure (*n*=1). In the present review, only 169 studies (50.6%) reported BIC and one of the likelihood ratio tests. Selecting best among several models should take into account theoretical factors in addition to fit indices (Muthén, 2006). Methods other than fit indices were also used in the studies reviewed. Studies also considered interpretability and theory (*n*=116, 34.7%), class sizes (*n*=42, 12.6%), AIC or BIC difference (*n*=35, 10.5%), and elbow plot (*n*=9, 2.7%) for model selection. Loglikelihood (LL) values and degrees of freedom (df) are other statistics to be reported with fit indices. LL was reported in 183 studies (54.8%) and df value was reported in 165 studies (49.4%). Only 44.9% of the studies (*n*=150) reported both statistics together. However, neither LL nor df were reported in 137 studies (40.7%).

Numbers of factors and classes

As suggested above, the best FMM model should be decided upon on the basis of model fit indices and theory. The number of factors and the number of latent classes also need to be reported for the final model. In the present review, the number of factors varied between 1 and 11, while the number of latent classes varied between 0 and 7. The mean number of classes across the 334 applications was 2.96, with a median of 3, and an SD=1.28. A majority of the studies reported a two-class (*n*=130), three-class (*n*=87) or four-class (*n*=62) FMM solution. The mean number of factors was 2.17, with a median of 2, and an SD=1.48. A majority of the studies reported one- (*n*=138), two- (*n*=104) or three-factor (*n*=48) models. Labeling for multiple latent classes in the final model, reporting the percentage or ratio of each latent class, and drawing a profile plot of the latent classes on the items are among the common practices in FMM analyses. In the present review, class percentages or proportions were reported in 85.3% (*n*=285), latent classes were labeled in 62.3% of the studies (*n*=208), and profile plots were drawn in 129 studies (38.6%).

Covariates and further analyses

Another important issue is the use of covariates in FMM analyses. Because latent classes are unobserved, variables (such as gender and race) are frequently linked to the latent class variable in order to better understand and characterize latent classes (Wang et al., 2022). A significant covariate effect would specifically mean that this covariate could explain the latent class membership. There are two options to examine the covariate effect in FMM analyses: adding the covariate variable directly to the model (sometimes referred to as a one-step approach) or performing a regression analysis with the latent classes obtained from the model (sometimes referred to as a three-step approach). In the three-step analysis, researchers first estimate an unconditional FMM without adding a covariate, then assign each respondent to one of the latent classes in the final model, after which a multinomial regression analysis is applied with the class membership and covariates (e.g., gender) specified by the researcher. In the present review, covariates in the estimation of FMMs were included in 199 studies (59.6% of the 334 applications). Only 28 studies added covariates directly to the FMM. The remaining 171 studies followed a two- or three-step approach. The most frequent covariates included were psychological and behavioral disorders (*n*=117, 35.0%), gender (*n*=104, 31.1%), age (*n*=92, 27.5%), and education level (*n*=39, 11.7%). Additional covariates included ethnicity (*n*=20, 6.0%), marital status (*n*=18, 5.4%), income (*n*=12, 3.6%), SES (*n*=11, 3.3%), body mass index (*n*=10, 3.0%), employment status (*n*=7, 2.1%), and language (*n*=4, 1.2%). Most studies included more than one covariate.

Once the best-fitting model was determined, classes were compared across different covariates. Applications in this review suggest that researchers were interested in analyzing the categorical latent variable (i.e., latent class) with further statistical analyses in order to investigate the relationship between class membership and auxiliary observed variables. A number of different analyses are used in FMM studies for covariate effect. The chi-square test (*n*=48, 14.4%), linear regression (*n*=39, 11.7%) and logistic regression (*n*=34, 10.2%) were the most commonly used analyses. Additional analyses included *t*-test and its non-parametric version (*n*=26, 7.8%), ANOVA (*n*=26, 7.8%), correlation (*n*=13, 3.9%), multiple comparison and comparison of means (*n*=13, 3.9%), MANOVA (*n*=11, 3.3%), R3STEP (*n*=9, 2.7%), ANCOVA (*n*=7, 2.1%), odds ratio (*n*=6, 1.8%), ROC curves (*n*=5, 1.5%), Wald test (*n*=4, 1.2%), cross-tabs (*n*=4, 1.2%), kappa classification agreement (*n*=4, 1.2%), SEM/ESEM (*n*=4, 1.2%), MIMIC model (*n*=3, 0.9%), and Fisher's exact test (*n*=3, 0.9%). Other statistical methods included cluster analysis (*n*=2), taxometric (*n*=2), IFA (*n*=2), PLS (*n*=1), bifactor analysis (*n*=1), MAXCOV (*n*=1), latent factor (*n*=1), and confirmatory MIRT (*n*=1).

Conclusion

In this study, a systematic review was conducted to first summarize the state of the use of FMMs as found in peer-reviewed journals and then to describe the trends in use based on these studies. There were a total of 304 peer-reviewed articles with 334 applications retrieved from databases including Web of Science, PubMed, ERIC, and PscyInfo. Relatively few studies using FMMs were published in fields other than psychology. In future studies, more emphasis may need to be placed on FMM analyses, particularly in the areas of health and education. Further, most studies were conducted in North American countries, including the U.S. (*n*=130), Canada (*n*=13), and European countries such as the Netherlands (*n*=19), Germany (*n*=14), Italy (*n*=12), and Finland (*n*=12). It is important that researchers in other countries take advantage of the FMM and seize the chance to answer new research questions.

Studies varied in their use of FMMs, outcome measured, methods for handling missing data, and reporting of methods and results. An interesting finding is that most studies used Mplus software and the MLR estimation method. Latent Gold was also used, particularly in LCFA studies. Not surprisingly, a majority of the reviewed studies used an exploratory approach, as FMMs are mainly exploratory in nature. As is the case with CFA, however, FMMs can also be applied in a confirmatory fashion. When multiple populations are believed to underlie the data, researchers may want to use a confirmatory version of the FMMs (Gagné, 2004) where restrictions are added in advance. The theoretically more grounded confirmatory approach may enable researchers to obtain more accurate findings in future studies. As Clark et al. (2013) has suggested, most of the FMM studies started with either EFA or CFA and LCA. In this review, however, more than a quarter of the studies did not apply these methods. Clark et al. (2013, p. 690) notes that "…the FMM-1 and FMM-2 often do not fit real data well because the specification of invariant factor loadings and thresholds are likely to be too restrictive for certain items." In this review, however, most of the studies relied on simple FMMs such as FMM-1 or LCFA, and FMM-2 methodology, and other models were not reported that often. One barrier to other models would be access to example code/syntax. As was noted in this review, FMMs can be applied with varying numbers of items and sample sizes. Typical applications include sample sizes of around 900 and approximately 12 items. The number of studies applying FMM analysis with small samples is not small. Studies with small samples can decide on the adequacy of the sample according to the power analysis that can be performed with Monte Carlo simulation analysis.

In the present review, 169 studies (50.3%) reported BIC and one of the likelihood ratio tests for model fit followed this suggestion. According to Muthén (2006), selecting among several models should take into account both theoretical and statistical factors. In the present review, 116 studies followed this suggestion (34.5%). Based on evidence from the studies in this review, it is recommended reporting multiple fit indices and also taking the theory into account, when selecting the final model. AIC, BIC, and SABIC were the most frequently reported IC indices. LMR and BLRT were the most reported LR based tests. This review also found that other considerations, including class proportions and entropy, were also considered for model selection. Although entropy is not a model selection index, it was the 3rd most reported index in the studies reviewed. Although previous studies (e.g., Lubke & Muthén, 2007) have indicated that entropy should not be used to determine the number of classes, there were a number of studies in this review that did use entropy in model selection. The least used indices for model selection were the AIC or BIC difference and the elbow plot methods. While most of the studies reported the percentage of latent classes in the final model, only 62% labeled the latent classes. Only 39% of the studies presented a profile plot over the final latent classes. This shows that after deciding on the optimal number of classes, researchers ignore some of the information they should give to inform the reader.

A majority of the studies reviewed included covariates in the estimation of FMMs. This approach can be useful for accounting for the uncertainty in class membership and for helping interpret FMM results. Findings of this review also showed that demographic variables such as gender, age, education level, and marital status, were used often in addition to psychological and behavioral disorders. A possible limitation of this review is that the focus was solely on peer-reviewed research published in English language journals. This ignores the gray literature, which includes unpublished studies, especially theses.

It is clear from this review that not all studies using FMMs have provided the same level and standard of study details. In the literature, some researchers (Clark et al., 2013; Lubke, 2019) have made some suggestions on how mixture model and FMM studies should be reported. It is expected that both the suggestions of these researchers and the results found in this review study will improve the reporting quality of future FMM studies. We hope that this review will contribute to future FMM research.

Declarations

Author Contribution: S.Ş. conceived of the presented idea. S.Ş. reviewed the studies and analyzed the data. A.S.C. verified the analytical methods. Both authors contributed to the final version of the manuscript.

Conflict of Interest: The authors have no conflicts of interest to declare.

Ethical Approval: Ethical approval was not required because this study retrieved and synthesized data from already published studies.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Baron, E., Bass, J., Murray, S. M., Schneider, M., & Lund, C. (2017). A systematic review of growth curve mixture modelling literature investigating trajectories of perinatal depressive symptoms and associated risk factors. *Journal of Affective Disorders, 223*, 194–208. https://doi.org/10.1016/j.jad.2017.07.046

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology, 39*(2), 174–187. https://doi.org/10.1093/jpepsy/jst084

Brown, T. A. (2013). Latent variable measurement models. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, pp. 257–280). Oxford University Press.

Cho, S. J., Cohen, A. S., & Kim, S. H. (2014). A mixture group bifactor model for binary responses. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 375–395. https://doi.org/10.1080/10705511.2014.915371

Clark, S. L., Muthén, B. O., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structural underlying psychological disorders. *Structural Equation Modeling, 20*, 681–703. https://doi.org/10.1080%2F10705511.2013.824786

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Gagné, P. E. (2004). *Generalized confirmatory factor mixture models: A tool for assessing factorial invariance across unspecified populations* [Unpublished doctoral dissertation]. University of Maryland, College Park.

Grove, R., Baillie, A., Allison, C., Baron-Cohen, S., & Hoekstra, R. A. (2015). Exploring the quantitative nature of empathy, systemising and autistic traits using factor mixture modelling. *The British Journal of Psychiatry, 207*(5), 400–406. https://doi.org/10.1192/bjp.bp.114.155101

Hofmans, J., Wille, B., & Schreurs, B. (2020). Person-centered methods in vocational research. *Journal of Vocational Behavior, 118*, 103398. https://doi.org/10.1016/j.jvb.2020.103398

Killian, M. O., Cimino, A. N., Weller, B. E., & Hyun Seo, C. (2019). A systematic review of latent variable mixture modeling research in social work journals. *Journal of Evidence-Based Social Work, 16*(2), 192–210. https://doi.org/10.1080/23761407.2019.1577783

Kim, E., Wang, Y., & Hsu, H. Y. (2023). A systematic review of and reflection on the applications of factor mixture modeling. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000630

Krawietz, C. E., & Pett, R. C. (2023). A systematic literature review of latent variable mixture modeling in communication scholarship. *Communication Methods and Measures, 17*(2), 83–110. https://doi.org/10.1080/19312458.2023.2179612

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

Lin, Y., & Mâsse, L. C. (2021). A look at engagement profiles and behavior change: A profile analysis examining engagement with the Aim2Be lifestyle behavior modification app for teens and their families. *Preventive Medicine Reports, 24*, 101565. https://doi.org/10.1016/j.pmedr.2021.101565

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778. https://www.jstor.org/stable/2673445

Lubke, G. (2019). Latent variable mixture models. In G. R., Hancock, L. M., Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 202–213). Routledge.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14*, 26–47. https://doi.org/10.1080/10705510709336735

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*, 21–39. https://doi.org/10.1037/1082-989X.10.1.21

Ma, X., Wang, M., Ma, J., Zhang, Z., Hao, Y., & Yan, N. (2022). The association between lifestyles and health conditions and the choice of traditional Chinese medical treatment in China: A latent class analysis. *Medicine, 101*(51), e32422. https://doi.org/10.1097/md.0000000000032422

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots, and related graphical displays. *Sociological Methodology, 31*(1), 223–264. http://dx.doi.org/10.1111/0081-1750.00096

Masyn, K. E., Henderson, C. E., & Greenbaum, P. E. (2010). Exploring the latent structures of psychological constructs in social development using the dimensional–categorical spectrum. *Social Development, 19*(3), 470–493. https://doi.org/10.1111/j.1467-9507.2009.00573.x

McDonald, R. P. (2003). A review of multivariate taxometric procedures: Distinguishing types from continua. *Journal of Educational and Behavioral Statistics, 28*, 77–81. http://dx.doi.org/10.3102/10769986028001077

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*(2), 195–215. https://psycnet.apa.org/doi/10.1007/BF02295283

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*, e1000097. https://doi.org/10.1136/bmj.b2535

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology, 44*(1), 369–399. https://doi.org/10.1177/0081175013516114

Morin, A. J., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

Muthén, B. (2006). Should substance use disorders be considered as categorical or dimensional?. *Addiction, 101*, 6–16. https://doi.org/10.1111/j.1360-0443.2006.01583.x

Muthén, L. K., & Muthén, B. O. (1998/2017). *Mplus user's guide* (Eight ed.). Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463–469. https://doi.org/10.1111/j.0006-341x.1999.00463.x

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H., (2006). *Mx: Statistical Modeling*, 7th ed. Medical College of Virginia, Richmond.

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

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271–282. https://doi.org/10.1177/014662169001400305

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://www.jstor.org/stable/2958889

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal, 8*(1), 289–317.

Sen, S., & Cohen, A. S. (2019). Applications of mixture IRT models: A literature review. *Measurement: Interdisciplinary Research and Perspectives, 17*(4), 177–191. https://doi.org/10.1080/15366367.2019.1583506

Spearman, C. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology, 5*, 201–293.

Spiegelhalter, D., Thomas, A., & Best, N. (2003). *WinBUGS (Version 1.4) [Computer software]*. Cambridge, UK: Biostatistics Unit, Institute of Public Health.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Ulbricht, C. M., Chrysanthopoulou, S. A., Levin, L., & Lapane, K. L. (2018). The use of latent class analysis for identifying subtypes of depression: A systematic review. *Psychiatry Research, 266*, 228–246. https://doi.org/10.1016/j.psychres.2018.03.003

Vermunt, J. K., & Magidson, J. (2003). *Latent Gold 3.0*. Belmont, MA. URL http://www. statisticalinnovations.com.

Viroli, C. (2011). *FactMixtAnalysis: Factor Mixture Analysis with covariates*.

von Davier, M. (2006). *Multidimensional Latent Trait Modelling (MDLTM) [Computer Software]*. Educational Testing Service.

Wang, Y., Cao, C., & Kim, E. (2022). Covariate inclusion in factor mixture modeling: Evaluating one-step and three-step approaches under model misspecification and overfitting. *Behavior Research Methods*, 1–16. https://doi.org/10.3758/s13428-022-01964-8

Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297–330. https://doi.org/10.1007/BF02294554