

Supplements for:
Chapter 34. Person-Centered Research Strategies in Commitment Research

Alexandre J.S. Morin
 Institute for Positive Psychology and Education
 Australian Catholic University, Australia
Alexandre.Morin@acu.edu.au

Appendix A. Traditional Approaches to Person-Centered Research	S3
<i>Midpoint Splits</i>	S3
<i>Interaction Effects</i>	S3
<i>Cluster analyses</i>	S4
<i>References</i>	S4
Appendix B. Technical Considerations in the Estimation of Mixture Models	S7
<i>Mixture Indicators</i>	S7
<i>Random Starts</i>	S7
<i>Class Enumeration</i>	S8
<i>References</i>	S9
Appendix C. Mplus Input Code for the Estimation of Mixture Models	S11
<i>Data section</i>	S11
<i>Variable section</i>	S11
<i>Analysis section</i>	S11
<i>Output section</i>	S12
<i>Model section</i>	S12
<i>Latent Profile Analysis</i>	S13
<i>Latent Profile Analysis with Covariates</i>	S15
Predictors	S15
Outcomes	S16
Correlates	S16
<i>Factor Mixture Analysis</i>	S17
<i>Multiple-Group Latent Profile Analyses</i>	S18
Configural Similarity	S18
Structural Similarity	S19
Dispersion Similarity	S19
Distribution Similarity	S20
<i>Multiple-Group Latent Profile Analyses with Predictors</i>	S20
Relations Freely Estimated Across Subpopulations	S20
Predictive Similarity	S21
<i>Multiple-Group Latent Profile Analyses with Outcomes</i>	S22
Relations Freely Estimated Across Subpopulations	S22
Explanatory Similarity	S23
<i>Latent Transition Analysis</i>	S24
Configural Similarity	S24
Structural Similarity	S24
Dispersion Similarity	S25
Distributional Similarity	S25
<i>Latent Transition Analysis with Predictors</i>	S26
Relations freely estimated at both time points	S26
Predictive Similarity	S26
<i>Latent Transition Analysis with Outcomes</i>	S27
Relations freely estimated at both time points	S27
Explanatory Similarity	S28
<i>Mixture Regression Analysis</i>	S29
<i>Multiple Group Mixture Regression Analysis</i>	S30
Configural Similarity	S30

Regression Similarity	S31
Structural Similarity	S31
Dispersion Similarity	S32
Distribution Similarity	S32
<i>Multiple Group Mixture Regression Analysis with Predictors</i>	S33
Relations Freely Estimated Across Subpopulations	S33
Predictive Similarity	S33
<i>Multiple Group Mixture Regression Analysis with Outcomes</i>	S34
Relations Freely Estimated Across Subpopulations	S34
Explanatory Similarity	S35
Appendix D. Growth Mixture Analyses	S36
<i>Alternative Functional Forms (Shape) for the Trajectories</i>	S37
Linear GMA	S37
Quadratic GMA	S37
Piecewise GMA	S38
Latent Basis GMA	S38
Additional non-linear GMA specifications	S39
<i>Alternative parameterizations of GMA</i>	S40
<i>Input for a Linear GMA</i>	S42
<i>Input for a Quadratic GMA</i>	S43
<i>Input for a Piecewise GMA</i>	S43
<i>Input for a Latent Basis GMA</i>	S44
<i>References</i>	S44

Appendix A.

Traditional Approaches to Person-Centered Research

Midpoint Splits. At the most basic level, studies interested in the identification of employees presenting distinct profiles on a set of commitment components have relied on a midpoint split approach to identify subgroups (e.g., Baugh & Roberts, 1994; Carson, Carson, Roe, Birkenmeier, & Philips, 1999; Somers & Birnbaum, 2000). More precisely, these studies have divided employees into subgroups according to whether they presented a high or low level on a set of commitment components, defining high and low levels according to some indicator of the sample-specific midpoint on the various measures used to assess the commitment component. Although interesting for initial exploratory purposes, this approach is importantly limited by its reliance on artificially-created subgroups that may not exist in nature (as illustrated by the theoretical Invested profile described earlier presenting an AC/CC dominated profile) and may conceal potentially important subgroups (such as moderately committed employees).

Interaction Effects. Although anchored within a variable-centered approach, some studies have investigated whether the effects of specific commitment components on a set of outcomes changed as a function of their levels on other commitment components using tests of interactions (Marsh, Hau, Wen, Nagengast, & Morin, 2013) among various commitment components (e.g., Gellatly, Meyer, & Luchak, 2006; Johnson, Groff, & Taing, 2009; Snape & Redman, 2003). Although these studies do indeed provide an efficient test of some of the key questions considered previously within a person-centered perspective, they still present multiple limitations in contrast to the mixture modeling approach advocated in this chapter. First, interaction effects involving more than three components are likely to be impossible to properly interpret within the limitations of the human brain. In contrast, profiles can easily be identified, and interpreted, even if based on multiple components. Second, interactions effects still assume the linearity of the effects across levels of the interacting variables. It is possible to incorporate non-linear terms (which have been showed to be potentially important to consider in commitment research; e.g. Morin, Vandenberghe, Turmel, Madore, & Maïano, 2013) in addition to the interactions themselves and even interactions among non-linear terms (Edwards, 2007). However, in interpretative limits are then likely to be reached with as few as two interacting terms.

Cluster Analyses. Cluster analyses are naturally suited to the identification of profiles (e.g., Becker & Billings, 1993; Swailes, 2004; Tsoumbris & Xenikou, 2010). However, cluster analyses present multiple technical limitations (e.g., Morin, Morizot et al., 2011) that can be avoided by the reliance on more flexible mixture models. For instance, cluster analyses do not provide clear guidelines to help in the identification of the correct number of profiles present in the data. Similarly, cluster analyses results are highly sensitive to the distributions of the variables used in the clustering process, and to the retained classification algorithm. More importantly, cluster analyses rely on rigid assumptions that often fail to hold with real-life data and can easily be relaxed in the context of mixture models (e.g., Muthén, 2002; Vermunt & Magidson, 2002) such as conditional independence (i.e., the indicators are uncorrelated conditional on the classification; e.g., Uebersax, 1999), class-invariant variances (the variances of the indicators are the same across profiles; e.g., Morin, Maïano, et al., 2011; Peugh & Fan, 2013), and exact assignment whereby each individual is assumed to correspond entirely to a single profile (although recent clustering methods provide ways to circumvent at least some of these limitations, such as fuzzy clustering which allows participants to assume partial membership into multiple profiles, see Brusco, Steinley, Cradit, & Singh, 2012). Although it is true that simulation studies have shown cluster analyses to be quite efficient at recovering true classification patterns present at the population level (e.g., Steinley & Brusco, 2011), this efficiency is limited to situations where the only objective of the research is to achieve a classification of participants into distinct subgroups based on indicators presenting no form of residual relations with one another. In contrast, whenever these assumptions need to be relaxed to properly model the data (see later discussion of factor mixture models), or when there is a need to incorporate predictors or outcomes to the model, then mixture models are preferable to cluster analyses. Indeed, being solely a classification process, cluster analyses do not provide the possibility to directly incorporate predictors or outcomes into the model without relying on suboptimal two-steps strategies (e.g., Bolck, Croon, & Hageaars, 2004).

References for Appendix A

- Baugh, S.G., & Roberts, R.M. (1994). Professional and organizational commitment among engineers: Conflicting or complementing ? *IEEE Transactions on Engineering Management*, *41*, 108-114.
- Becker, T.E., & Billings, R.S. (1993). Profiles of commitment: An empirical test summary. *Journal of*

- Organizational Behavior*, 14, 177-190.
- Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12, 3-27.
- Brusco, M.J., Steinley, D., Cradit, J.D., & Singh, R. (2011). Emergent clustering methods for empirical OM research. *Journal of Operations Management*, 30, 454-466.
- Carson, K.D., Carson, P.P., Roe, C.W., Birkenmeier, B.J., & Philips, J.S. (1999). Four commitment profiles and their relationships to empowerment, service recovery, and work attitudes. *Public Personnel Management*, 28, 1-13.
- Edwards, J.R. (2007). Polynomial regression and response surface methodology. In C. Ostroff & T.A. Judge (Eds.), *Perspectives on organizational fit* (pp. 361-372). San Francisco: Jossey-Bass.
- Johnson, R. E., Groff, K. W., & Taing, M. U. (2009). Nature of the interactions among commitments: Complementary, competitive, or synergistic?. *British Journal of Management*, 20, 431-447.
- Gellatly, I. R., Meyer, J. P., & Luchak, A. A. (2006). Combined effects of the three commitment components on focal and discretionary behaviors. *Journal of Vocational Behavior*, 69, 331-345.
- Marsh, H.W., Hau, K.-T., Wen, Z., Nagengast, B., & Morin, A.J.S. (2013). 17. Moderation. In T.D. Little (Ed.), *Oxford Handbook of Quantitative Methods, Vol. 2* (pp. 361-386), New York, NY: Oxford University Press.
- Morin, A. J. S., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). Growth mixture modeling of adolescents trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling*, 18, 613-648.
- Morin, A.J.S., Morizot, J., Boudrias, J.-S., & Madore, I., (2011). A multifoci person-centered perspective on workplace affective commitment: A latent profile/factor mixture Analysis. *Organizational Research Methods*, 14, 58-90.
- Morin, A.J.S., Vandenberghe, C., Turmel, M.-J., Madore, I., & Maïano, C. (2013). Probing into Commitment's Nonlinear Relationships to Work Outcomes. *Journal of Managerial Psychology*, 28, 202-223.
- Muthén, B.O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81-117.
- Somers, M.J., & Birnbaum, D. (2000). Exploring the relationship between commitment profiles and

- work attitudes, employees withdrawal, and job performance. *Public Personnel Management*, 29, 353-365.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling*, 20, 616–639.
- Snape, E., & Redman, T. (2003). An evaluation of a three component model of Occupational commitment: dimensionality and consequences among United Kingdom human resource management specialists. *Journal of Applied Psychology*, 88, 152-159.
- Steinley, D., & Brusco, M.J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, 16, 63-79.
- Swales, S. (2004). Commitment to change: Profiles of commitment and in-role performance. *Personnel Review*, 33, 187-204.
- Tsoumbris, P., & Xenikou, A. (2010). Commitment profiles: The configural effects of the forms and foci of commitment on work outcomes. *Journal of Vocational Behavior*, 77, 401–411.
- Uebersax, J.S. (1999). Probit Latent Class Analysis with Dichotomous or Ordered Category Measures: Conditional Independence/Dependence Models. *Applied Psychological Measurement*, 23, 283-297.
- Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class models* (pp. 89-106). New York: Cambridge.

Appendix B.

Technical Considerations in the Estimation of Mixture Models

Mixture Indicators. Mixture models are usually estimated using scale scores on the various indicators (commitment components; i.e., taking the sum or average on the items used to assess a component and using this aggregated score as the indicator). Although it is well known that using latent variables controlled for measurement error (i.e., models where the items are used to estimate latent factors, which are then used as mixture indicators) provides a much stronger approach (e.g., Bollen, 1989), applications of fully latent mixture models are few (e.g., Morin, Scalas, & Marsh, 2015). In fact, given the complexity of mixture models it is often impossible in practice to implement a fully latent approach to their estimation. An alternative, which is becoming more frequent in recent applications of mixture models, is to rely on factor scores saved from preliminary measurement models (e.g., Kam, Morin, Meyer, & Topolnytsky, 2015; Morin & Marsh, 2015). Factors scores do not explicitly control for measurement errors the way latent variables do. However, by giving more weight to items presenting lower levels of measurement errors, they still provide a partial implicit control for measurement errors, making them a stronger alternative than scale scores. An added advantage of factors scores is that when they are estimated from more complex measurement models (including methodological controls, bifactor models, exploratory structural equation models, etc.) they tend to preserve the nature of the underlying measurement structure better than scale scores. More importantly, preliminary measurement models can be used to systematically assess the measurement invariance of the measures across time or groups (Millsap, 2011) and then save factor scores from the most invariant measurement model. This approach ensures comparability of the results over time or groups for multiple group, or longitudinal, applications of mixture models. Furthermore, factors scores are naturally standardized with a mean of zero and a variance of one, making them perfect for the application of the profile labeling scheme described in the chapter.

Random Starts. An important challenge in mixture models is to avoid converging on a local solution (i.e., a false maximum likelihood). More precisely, mixture models are estimated through an iterative process that risks converging on a local solution rather than on a true maximum likelihood when starts values are not adequate (Hipp & Bauer, 2006). To limit this risk, models should be

estimated with multiple sets of random start values (Hipp & Bauer, 2006; McLachlan & Peel, 2000). As far as random starts go, the basic rule is “The more, the better,” keeping in mind that more also requires more computational time. However, in published person-centered research, researchers all seem to possess their own interpretation of how many start values is optimal. In practice, our recommendation is to use at least 3000 sets of random starts, 100 iterations of each of these sets of start values, and to retain at least the 100 best sets of start values for final stage optimization. These values can be increased to 5000, 200, and 2000 when the final solution is not sufficiently replicated. We see these values as a minimum that can be increased as needed.

Class Enumeration. Typically, alternative solutions including increasing number of latent profiles (from one to a number typically above theoretical expectations but typically under 10) are contrasted in order to select the final solution in a mainly exploratory manner (but see Finch & Bronk, 2011, for confirmatory applications). As noted in the chapter, selection of the optimal number of profiles is determined based on inspection of: (a) the substantive meaning and theoretical conformity of the solution; (b) the statistical adequacy of the solution and (c) statistical indicators. We focus here on these statistical indicators. Although conventional indicators of the absolute fit of a model (e.g., RMSEA) are not available for mixture models, several indicators of relative fit are available: The Akaike Information Criterion (AIC), the Consistent AIC (CAIC), the Bayesian information criterion (BIC), and the sample-adjusted BIC (SABIC). A lower value on these indicators suggests a better-fitting model. Classical likelihood ratio tests (LRT) are appropriate for the comparisons of nested models based on the same variables and number of profiles, but inappropriate for class enumeration purposes. However, two LRT approximations are available for this purpose: The Lo, Mendel and Rubin’s (2001) LRT (the standard and adjusted version of this test typically yield the same conclusions), and the Bootstrap LRT (BLRT; McLachlan & Peel, 2000). These tests compare a k -profile model with a $k-1$ -profile model. Non-significant p values indicate that the $k-1$ profile model should be retained. Finally, although it should not be used for class enumeration purposes the entropy provides a useful indication of the precision with which the cases are classified into the profiles, with larger values (closer to 1) indicating fewer classification errors. Multiple simulation studies have looked at the relative efficacy of these various indicators, and show that the BIC, SABIC, CAIC and

BLRT appear particularly effective (Henson et al., 2007; McLachlan & Peel, 2000; Nylund, Asparouhov, & Muthén, 2007; Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2008; Tolvanen, 2007; Yang, 2006). When these indicators fail to retain the optimal model, the BIC and CAIC tend to underestimate the true number of profiles, while the AIC, ABIC, and BLRT tend to overestimate it. However, these indicators are heavily influenced by sample size (Marsh et al., 2009), which means that with a sufficiently large sample, they may keep on suggesting the addition of profiles without ever converging on a preferable solution. When this happens, it is recommended to present these indicators in the format of “elbow plots” (Morin, Maïano, et al., 2011; Petras & Masyn, 2010). These plots illustrate the gains in fit associated with the addition of profiles, and the point after which the slope flattens is typically indicates the optimal number of profiles.

References for Appendix B

- Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Finch, W.H., & Bronk, K.C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling, 18*, 132-151.
- Henson, J.M., Reise, S.P., & Kim, K.H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling, 14*, 202-226.
- Hipp, J.R., & Bauer, D.J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*, 36-53.
- Kam, C., Morin, A.J.S., Meyer, J.P., & Topolnytsky, L. (2015, In Press). Are commitment profiles stable and predictable? A latent transition analysis. *Journal of Management*.
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-778.
- McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor & Francis.
- Morin, A.J.S., Maïano, C., Nagengast, B., Marsh, H.W., Morizot, J., & Janosz, M. (2011). Growth mixture modeling of adolescents trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling, 18*, 613-648.

- Morin, A.J.S., & Marsh, H.W. (2015). Disentangling Shape from Levels Effects in Person-Centred Analyses: An Illustration Based University Teacher Multidimensional Profiles of Effectiveness. *Structural Equation Modeling*, 22 (1), 39-59.
- Morin, A.J.S., Scalas, L.F., & Marsh, H.W. (2015, In Press). Tracking the elusive actual-ideal discrepancy model within latent subpopulations. *Journal of Individual Differences*.
- Nylund, K.L., 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.
- Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A.R. Piquero, & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 69-100). New York, NY: Springer.
- Peugh, J. & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling*, 20, 616-639.
- Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*, 20, 640-657.
- Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. In G.R. Hancock & K.M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 317-341). Charlotte, NC: Information Age.
- Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study*. PhD dissertation, Department of Mathematics, University of Jyväskylä, Jyväskylä, Finland.
- Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50, 1090–1104.

Appendix C.

Mplus Input Code for the Estimation of Mixture Models

In Mplus syntax, text sections preceded by an exclamation mark (!) are annotations.

DATA:

The first part of the Mplus syntax identify the data set to be used in the analysis. If the data set is in the same folder as the input file, only the name of the data set needs to be indicated. If the data set is in another folder, then the full path needs to be specified. Here, the data set is labeled “Commitment.dat” and is located in the same folder.

```
DATA:
FILE IS Commitment.dat;
```

VARIABLE:

In the VARIABLE section, the NAMES function identifies all variables included in the data set, in order of appearance. The USEVARIABLES function defines the variables to be used in the analysis. The MISSING function defines the missing data code (we typically use the same code for all variables). The IDVARIABLE function defines the unique identifier for participants. The CLASSES function defines the number of latent profiles to be estimated (here 3). The USEOBS function can be used to limit the estimation to a subset of participants (here, we limit the estimation to employees (identified in the variable Status, coded 1 for employees, and 2 for supervisors). The CLUSTER function can be used to define the unique identifier for the clustering (level 2) variable to be controlled in the analysis (e.g., the work unit).

```
VARIABLE:
NAMES = ID Status unit Pred1 Pred2 Cor1 Cor2 Out1 Out2 AC1 NC1 CC1 AC2 NC2 CC2;
USEVARIABLES = AC1 NC1 CC1;
MISSING = all (999);
IDVARIABLE = ID;
CLASSES = c (3);
! CLUSTER = P5Code;
! USEOBS Status EQ 2;
```

ANALYSIS:

The ANALYSIS section described the analysis itself. Here, we request the estimation of a mixture model (TYPE = MIXTURE) including a correction for the nesting of employees within work unit (TYPE = COMPLEX) and using the robust maximum likelihood estimator (ESTIMATOR = MLR). STARTS = 3000 100 requests 3000 sets of random start values, with the best 100 of these starts retained for final stage optimization. STITERATIONS = 100 requests that all random starts be allowed a total of 100 iterations. PROCESS = 3 requires that the model be estimated using 3 of the available processors to speed up the estimation.

```
Analysis:
TYPE = MIXTURE COMPLEX;
ESTIMATOR = MLR;
PROCESS = 3;
STARTS = 3000 100; STITERATIONS = 100;
```

The best way to ensure that the final solution represents a true maximum likelihood rather than a local solution is to increase the number of starts values. As part of the output, Mplus provides the loglikelihood values associated with all of the random starts retained for the final stage optimization. It will also indicate how many of the start value runs did not converge.

```
RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
15 perturbed starting value run(s) did not converge.
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
-6095.887 991399      1433
-6095.887 165268     2436
-6095.887 551639      55
-6095.887 58353      1723
```

-6095.887	168648	1788
-6095.887	445692	2796
-6095.887	973369	202
-6095.887	623887	2611
-6096.828	163110	584
-6096.828	762461	425
-6096.828	62715	2599
-6104.647	264951	1130
-6104.647	930872	277
-6104.647	670998	1876
-6104.698	98068	998
-6105.491	361131	2563
-6105.491	293428	2473
-6105.491	18741	2970
-6106.248	415502	194
-6106.682	338858	1342
-6106.682	824956	607

...

In this example, the best loglikelihood value was replicated 8 times (in bold, the number of times the value of -6095.887 appears in the first column), which is satisfactory. Although no clear-cut rule exists, we suggest that solutions should be replicated at least 5 times. Failing to do so, additional tests should be conducted while increasing the number of start values and/or iterations or using user-defined starts values (for instance, using the starts values from the best fitting solution provided when requesting SVALUES in the output section of the syntax and using these starts values in the model while keeping the random starts function active – we provide an example below). The second column provides the model seed associated with each random start. Using the seed provides an easy way to replicate the final solution (or any other solution) while decreasing computational time. To do so, the following ANALYSIS section can be used to replicate the above solution. This seed however will not ensure that the solution is replicated if additional covariates are added to the model (SVALUES then need to be used).

```
Analysis:
TYPE = MIXTURE COMPLEX;
ESTIMATOR = MLR;
PROCESS = 3;
STARTS = 0;
OPTSEED = 991399;
STITERATIONS = 100;
```

OUTPUT:

The last section of the syntax covers specific sections of the output that are requested. Here we request standardized model parameters (STDYX), sample statistics (SAMPSTAT), confidence intervals (CINTERVAL), the starts values corresponding to the solution (SVALUES), the residuals (RESIDUAL), the arrays of parameter specifications and starting values (TECH1), the profile-specific sample characteristics (TECH7), the LMR and aLMR (TECH11), and the BLRT (TECH14).

```
OUTPUT:
STDYX SAMPSTAT CINTERVAL SVALUES RESIDUAL TECH1 TECH7 TECH11 TECH14;
```

MODEL:

In between the ANALYSIS and OUTPUT section, the MODEL section describes the specific analysis to be conducted. We provide code, in sequence, for each of the models described in the manuscript.

Latent Profile Analysis

The MODEL section includes an %OVERALL% section describing the global relations estimated among the constructs, and profile specific statements (here %c#1% to %c#3%, where c corresponds to the label used to define the categorical latent variable in the CLASSES command of the VARIABLE: section, and the number 1 to k refers to the specific value of this variable (the specific profile). Here, no relations are estimated between the variables so nothing appears in the %OVERALL% section. The profile specific sections request that the means (indicated by the name of the variable between brackets []) and variances (indicated by the names) of the indicators be freely estimated in all profiles.

```
MODEL:
%OVERALL%
%c#1%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#2%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#3%
[AC1 NC1 CC1];
AC1 NC1 CC1;
```

Variances that are equal across profiles:

```
MODEL:
%OVERALL%
%c#1%
[AC1 NC1 CC1];
%c#2%
[AC1 NC1 CC1];
%c#3%
[AC1 NC1 CC1];
```

With correlated uniquenesses (correlations are identified by WITH) among all profile indicators (not recommended) where the correlations are invariant across profiles:

```
MODEL:
%OVERALL%
AC1 WITH NC1 CC1; NC1 WITH CC1;
%c#1%
[AC1 NC1 CC1]; AC1 NC1 CC1;
%c#2%
[AC1 NC1 CC1]; AC1 NC1 CC1;
%c#3%
[AC1 NC1 CC1]; AC1 NC1 CC1;
```

With correlated uniquenesses among all profile indicators (not recommended) where the correlations are freely estimated in all profiles:

```
MODEL:
%OVERALL%
AC1 WITH NC1 CC1; NC1 WITH CC1;
%c#1%
[AC1 NC1 CC1]; AC1 NC1 CC1;
AC1 WITH NC1 CC1; NC1 WITH CC1;
%c#2%
[AC1 NC1 CC1]; AC1 NC1 CC1;
AC1 WITH NC1 CC1; NC1 WITH CC1;
%c#3%
[AC1 NC1 CC1]; AC1 NC1 CC1;
AC1 WITH NC1 CC1; NC1 WITH CC1;
```

When SVALUES are requested, the output will include a section similar to the following, this section can be cut-and-pasted and used as a replacement of the MODEL section represented above in conjunction with the STARTS function set to 0 (STARTS = 0) to replicate the final solution. This function is particularly useful when one wants to include covariates in a model while making sure that the final unconditional LPA solution remains unchanged.

MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

```
%OVERALL%
[ c#1*-0.93515 ];
[ c#2*0.49113 ];
% C#1%
[ CC1*0.39664 ];
[ NC1*-0.93155 ];
[ AC1*-1.59367 ];
CC1*1.59137;
NC1*2.12542;
AC1*0.53789;
% C#2%
[CC1*-0.15201 ];
[NC1*-0.29812 ];
[AC1*-0.18248 ];
CC1*0.72513;
NC1*0.43116;
AC1*0.42376;
% C#3%
[CC1*0.08333 ];
[NC1*0.86526 ];
[AC1*0.94111 ];
CC1*1.13489;
NC1*0.48030;
AC1*0.44053;
```

Latent Profile Analysis with Covariates

Predictors:

Direct inclusion of predictors to the model is done by way of a multinomial logistic regression where the predictors are specified as having an impact on profile membership (c#1-c#4 ON Pred1 Pred2) in the %OVERALL% section.

```
MODEL:
%OVERALL%
c#1-c#4 ON Pred1 Pred2;
%c#1%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#2%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#3%
[AC1 NC1 CC1];
AC1 NC1 CC1;
```

To make sure that the nature of the profiles remains unchanged by the inclusion of predictors, the SVALUES from the final solution can also be used.

```
%OVERALL%
[ c#1*-0.93515 ]; [ c#2*0.49113 ];
c#1-c#4 ON Pred1 Pred2;
%c#1%
[ CC1*0.39664 ]; [ NC1*-0.93155 ]; [ AC1*-1.59367 ];
CC1*1.59137; NC1*2.12542; AC1*0.53789;
%c#2%
[CC1*-0.15201 ]; [NC1*-0.29812 ]; [AC1*-0.18248 ];
CC1*0.72513; NC1*0.43116; AC1*0.42376;
%c#3%
[CC1*0.08333 ]; [NC1*0.86526 ]; [AC1*0.94111 ];
CC1*1.13489; NC1*0.48030; AC1*0.44053;
```

Among available AUXILIARY approaches, the R3STEP (see Asparouhov & Muthén, 2014) appears the most naturally suited to the exploration of predictors. This approach is similar to the multinomial logistic regression described above, but explicitly tests whether including the predictors resulted in a change in the nature of the profile. When this occurs (and the previous direct approaches did not work), then predictors needs to be treated as correlates. To use this approach, the following line of code needs to be included to the VARIABLE section 9 (in bold):

```
VARIABLE:
NAMES = ID Status unit Pred1 Pred2 Cor1 Cor2 Out1 Out2 AC1 NC1 CC1 AC2 NC2 CC2;
USEVARIABLES = AC1 NC1 CC1;
MISSING = all (999);
IDVARIABLE = ID;
CLASSES = c (3);
AUXILIARY = Pred1 (R3STEP) Pred2 (R3STEP);
```

With the Model section remaining unchanged (start values may help).

Outcomes:

The direct inclusion of distal outcomes to the model involves adding them as additional mixture indicators, preferably while using the SVALUES from the final solution to ensure stability in the nature of the profiles. Parameter labels are given to the means of the outcome variables in each profile (in parentheses), and these labels can be used with the MODEL CONSTRAINT command to provide tests of significance of differences between profiles on the various outcomes.

```
%OVERALL%
[ c#1*-0.93515 ]; [ c#2*0.49113 ];
%C#1%
[ CC1*0.39664 ]; [ NC1*-0.93155 ]; [ AC1*-1.59367 ];
CC1*1.59137; NC1*2.12542; AC1*0.53789;
[Out1] (oa1); [Out2] (ob1);
%C#2%
[CC1*-0.15201 ]; [NC1*-0.29812 ]; [AC1*-0.18248 ];
CC1*0.72513; NC1*0.43116; AC1*0.42376;
[Out1] (oa2); [Out2] (ob2);
%C#3%
[CC1*0.08333 ]; [NC1*0.86526 ]; [AC1*0.94111 ];
CC1*1.13489; NC1*0.48030; AC1*0.44053;
[Out1] (oa3); [Out2] (ob3);
MODEL CONSTRAINT:
! New parameters are created using this function and reflect pairwise mean differences between
! profiles. e.g. y12 reflect the differences between the means of profiles 1 and 2 on Out1
NEW (y12);
y12 = oa1-oa2;
NEW (y13);
y13 = oa1-oa3;
NEW (y23);
y23 = oa2-oa3;
NEW (z12);
z12 = ob1-ob2
NEW (z13);
z13 = ob1-ob3;
NEW (z23);
z23 = ob2-ob3;
```

Three alternative Auxiliary approaches (see Asparouhov & Muthén, 2014) are available. The first one is similar to the R3STEP approach and tests the degree to which continuous outcomes change the nature of the profiles. This approach can either allow for the variances of the outcomes to be freely estimated in all profiles (DU3STEP) or invariant across profiles (DE3STEP). A more recent alternative (BCH) has been shown to outperform these approaches, while ensuring the stability of the profile solutions. Finally, a last approach also ensures the stability of the profile solution, while accommodating continuous (DCON) and categorical (DCAT) outcomes. Our recommendation, based on current knowledge, would be to rely on the BCH approach for continuous outcomes and the DCAT approach for categorical outcomes.

```
! Pick between these alternatives:
AUXILIARY = Out1 (DU3STEP) Out2 (DU3STEP);
AUXILIARY = Out1 (DE3STEP) Out2 (DE3STEP);
AUXILIARY = Out1 (BCH) Out2 (BCH);
AUXILIARY = Out1 (DCON) Out2 (DCON);
AUXILIARY = Out1 (DCAT) Out2 (DCAT);
```

Correlates:

Correlates can be incorporated to the model via the Auxiliary A approach.

```
AUXILIARY = Cor1 (e) Cor2 (e);
```


Factor Mixture Analysis

Here we present the input for the factor mixture analytic model described in the main manuscript that provided a way to control for global levels shared among the indicators in order to estimate clearer latent profiles. Although this input is based on commitment mindsets, we reinforce that such a model would not be suitable for this context, and more appropriate to the estimation of profile analyses based on multiple targets of commitment to control for global mindsets tendencies (e.g. Morin, Morizot et al., 2011). The only difference with the previous LPA models is the introduction of a common factor model in the %OVERALL% section. This factor model is specified as invariant across profiles. This common factor is labeled G, and defined by the same indicators (BY defines factor loadings). All loadings on this factor are freely estimated (the * associated with the first indicators overrides the default of constraining the loading of the first factor to be 1). The factor variance thus needs to be fixed to 1 for identification purposes (the @ is used to fix a parameter to a specific value). Because the intercepts of the indicators of this factor will be freely estimated across profiles, the factor means needs to be fixed to 0 for identification purposes.

```
MODEL:
%OVERALL%
G BY AC1* NC1 CC1;
G@1;
[G@0];
%c#1%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#2%
[AC1 NC1 CC1];
AC1 NC1 CC1;
%c#3%
[AC1 NC1 CC1];
AC1 NC1 CC1;
```

Multiple-Group Latent Profile Analyses: Configural Similarity

In the VARIABLE section, the multiple groups needs to be defined using the KNOWCLASS function, which uses a label (here we use cg) to define this new grouping variable, and the levels of this new grouping variables are defined as: (a) including participants with a value of 1 (Employees) on the variable Status, and (b) including participants with a value of 2 (supervisor) on the variable Status. There are now two “latent” grouping variables, C estimated as part of the model estimation (the profiles) and having k levels (here we are still working with a 3-profile solution) and CG reflecting the observed subgroups (Status) with 2 levels. Participants are allowed to be cross classified.

```
KNOWCLASS = cg (Status = 1 Status = 2);
CLASSES = cg (2) c (3);
```

The %OVERALL% section is used to indicate that the class sizes are freely estimated in all observed samples (employees and supervisors) using the ON function (reflecting regressions) indicating that profile membership is conditional on status. $k-1$ statements are required (i.e., 2 for a 3-profile model). Profile-specific statements are then defined using a combination of the known classes CG and the estimated classes C. Labels in parentheses identify parameters that are estimated to be equal across groups. Here, none of the labels are shared between groups, so that the means and variances are freely estimated in all combinations of profiles and gender. Lists of constraints (m1-m3) apply to the parameters in order of appearance (m1 applies to AC1, m2 to NC1, m3 to CC1).

```
%OVERALL%
c#1 on cg#1; c#2 on cg#1;

%cg#1.c#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (v1-v3);
%cg#1.c#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (v4-v6);
%cg#1.c#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (v7-v9);

%cg#2.c#1%
[AC1 NC1 CC1] (mm1-mm3);
AC1 NC1 CC1 (vv1-vv3);
%cg#2.c#2%
[AC1 NC1 CC1] (mm4-mm6);
AC1 NC1 CC1 (vv4-vv6);
%cg#2.c#3%
[AC1 NC1 CC1] (mm7-mm9);
AC1 NC1 CC1 (vv7-vv9);
```

Multiple-Group Latent Profile Analyses: Structural Similarity

The only difference between this model and the previous one is that the means are constrained to be equal across status within each profile using identical labels in parentheses.

```
%OVERALL%
c#1 on cg#1; c#2 on cg#1;

%cg#1.c#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (v1-v3);
%cg#1.c#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (v4-v6);
%cg#1.c#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (v7-v9);

%cg#2.c#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (vv1-vv3);
%cg#2.c#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (vv4-vv6);
%cg#2.c#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (vv7-vv9);
```

Multiple-Group Latent Profile Analyses: Dispersion Similarity

The only difference between this model and the previous one is that the variances are constrained to be equal across status within each profile using identical labels in parentheses.

```
%OVERALL%
c#1 on cg#1; c#2 on cg#1;

%cg#1.c#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (v1-v3);
%cg#1.c#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (v4-v6);
%cg#1.c#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (v7-v9);

%cg#2.c#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (v1-vv3);
%cg#2.c#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (v4-vv6);
%cg#2.c#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (v7-vv9);
```

Multiple-Group Latent Profile Analyses: Distribution Similarity

The only difference between this model and the previous one is that nothing appears in the %OVERALL% section of the input to reflect the fact that the sizes of the profiles are no longer conditional on status.

```
%OVERALL%

%cg#1.c#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%cg#1.c#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%cg#1.c#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);

%cg#2.c#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-vv3);
%cg#2.c#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-vv6);
%cg#2.c#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-vv9);
```

Multiple-Group Latent Profile Analyses with Predictors: Relations Freely Estimated Across Subpopulations

This models uses the SVALUES associated with the previous model of dispersion similarity (specified as * followed by the value of the estimated parameters), and simply include predictor effects on profile membership (c#1-c#2 ON Pred1 Pred2;). To allow these effects to be freely estimated across status, they need to be constrained to 0 in the %OVERALL% section, and freely estimated in both status groups in a new section of the input specifically referring to CG. See all sections in bold.

```
%OVERALL%
[ cg#1*-0.00217 ]; [ c#1*-0.93515 ]; [ c#2*0.49113 ];
c#1-c#2 ON Pred1@0 Pred2@0;
%CG#1.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
%CG#1.C#2%
[ CC1*-0.15201 ] (m4); [ NC1*-0.29812 ] (m5); [ AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
%CG#1.C#3%
[ CC1*0.08333 ] (m7); [ NC1*0.86526 ] (m8); [ AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
%CG#2.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
%CG#2.C#2%
[ CC1*-0.15201 ] (m4); [ NC1*-0.29812 ] (m5); [ AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
%CG#2.C#3%
[ CC1*0.08333 ] (m7); [ NC1*0.86526 ] (m8); [ AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
MODEL cg:
%cg#1%
c#1-c#2 ON Pred1 Pred2;
%cg#2%
c#1-c#2 ON Pred1 Pred2;
```

Multiple-Group Latent Profile Analyses with Predictors: Predictive Similarity

This model is almost identical to the previous one. In order for the effects of the predictors to be constrained to similarity across genders, they simply need to be specified as freely estimated in the %OVERALL% section (c#1-c#4 ON Pred1 Pred2;), while taking out the gender specific sections.

```
%OVERALL%
[ cg#1*-0.00217 ]; [ c#1*-0.93515 ]; [ c#2*0.49113 ];
c#1-c#2 ON Pred1 Pred2;
%CG#1.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
%CG#1.C#2%
[ CC1*-0.15201 ] (m4); [ NC1*-0.29812 ] (m5); [ AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
%CG#1.C#3%
[ CC1*0.08333 ] (m7); [ NC1*0.86526 ] (m8); [ AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
%CG#2.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
%CG#2.C#2%
[ CC1*-0.15201 ] (m4); [ NC1*-0.29812 ] (m5); [ AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
%CG#2.C#3%
[ CC1*0.08333 ] (m7); [ NC1*0.86526 ] (m8); [ AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
```

Multiple-Group Latent Profile Analyses with Outcomes: Relations Freely Estimated Across Subpopulations

This model also uses the SVALUES associated with the model of dispersion similarity. Here, we simply request the free estimation of the distal outcome means in all profiles x status ([Out1 Out2]). We also use labels in parentheses to identify these new parameters, which will then be used in a new MODEL CONSTRAINT section to request tests of the significance of mean differences between profiles and genders.

```
%OVERALL%
[ cg#1*-0.00217 ]; [ c#1*-0.93515 ]; [ c#2*0.49113 ];
%CG#1.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
[Out1] (oa1);
[Out2] (ob1);
%CG#1.C#2%
[CC1*-0.15201 ] (m4); [NC1*-0.29812 ] (m5); [AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
[Out1] (oa2);
[Out2] (ob2);
%CG#1.C#3%
[CC1*0.08333 ] (m7); [NC1*0.86526 ] (m8); [AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
[Out1] (oa3);
[Out2] (ob3);
%CG#2.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
[Out1] (oaa1);
[Out2] (obb1);
%CG#2.C#2%
[CC1*-0.15201 ] (m4); [NC1*-0.29812 ] (m5); [AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
[Out1] (oaa2);
[Out2] (obb2);
%CG#2.C#3%
[CC1*0.08333 ] (m7); [NC1*0.86526 ] (m8); [AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
[Out1] (oaa3);
[Out2] (obb3);

MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2;
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;
NEW (yy12); yy12 = oaa1-oaa2;
NEW (yy13); yy13 = oaa1-oaa3;
NEW (yy23); yy23 = oaa2-oaa3;
NEW (zz12); zz12 = obb1-obb2;
NEW (zz13); zz13 = obb1-obb3;
NEW (zz23); zz23 = obb2-obb3;
```

Multiple-Group Latent Profile Analyses with Outcomes: Explanatory Similarity

This model is almost identical to the previous one except that the parameter labels are used to constrain the outcome means to be invariant across status. As a result, less lines of code are required in the MODEL COSNTRAIINT section.

```

%OVERALL%
[ cg#1*-0.00217 ]; [ c#1*-0.93515 ]; [ c#2*0.49113 ];
%CG#1.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
[Out1] (oa1);
[Out2] (ob1);
%CG#1.C#2%
[CC1*-0.15201 ] (m4); [NC1*-0.29812 ] (m5); [AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
[Out1] (oa2);
[Out2] (ob2);
%CG#1.C#3%
[CC1*0.08333 ] (m7); [NC1*0.86526 ] (m8); [AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
[Out1] (oa3);
[Out2] (ob3);
%CG#2.C#1%
[ CC1*0.39664 ] (m1); [ NC1*-0.93155 ] (m2); [ AC1*-1.59367 ] (m3);
CC1*1.59137 (v1); NC1*2.12542 (v2); AC1*0.53789 (v3);
[Out1] (oa1);
[Out2] (ob1);
%CG#2.C#2%
[CC1*-0.15201 ] (m4); [NC1*-0.29812 ] (m5); [AC1*-0.18248 ] (m6);
CC1*0.72513 (v4); NC1*0.43116 (v5); AC1*0.42376 (v6);
[Out1] (oa2);
[Out2] (ob2);
%CG#2.C#3%
[CC1*0.08333 ] (m7); [NC1*0.86526 ] (m8); [AC1*0.94111 ] (m9);
CC1*1.13489 (v7); NC1*0.48030 (v8); AC1*0.44053 (v9);
[Out1] (oa3);
[Out2] (ob3);

MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;

```

Latent Transition Analysis: Configural Similarity.

The estimation of a latent transition model is highly similar to the estimation of a multiple-group LPA with the exception that the latent categorical variable representing the observed group defined using the KNOWCLASS function (CG in the previous examples) is replaced by another unknown latent categorical variable representing profiles groups estimated at the second time point.

```
CLASSES = c1 (3) c2 (3);
```

Because of the similarity of the inputs, we do not comment the sequence of similarity tests in the next sections. In the basic LTA model, the %OVERALL% section states that membership into the profiles at the second time point (C2) is conditional on membership in the profiles estimated at the first time points (C1). This is necessary to estimate the individual transition probabilities over time. Then two sections of the inputs are used to define the profiles estimated at the first (MODEL C1:) and second (MODEL C2:) time points, where the profiles are defined by distinct variables reflecting the mixture indicators measured at either the first (e.g., AC1) or second (e.g., AC2) time point.

```
%OVERALL%
c2 on c1;
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3);
AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6);
AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9);
AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (mm1-mm3);
AC2 NC2 CC2 (vv1-vv3);
%c2#2%
[AC2 NC2 CC2] (mm4-mm6);
AC2 NC2 CC2 (vv4-vv6);
%c2#3%
[AC2 NC2 CC2] (mm7-mm9);
AC2 NC2 CC2 (vv7-vv9);
```

Latent Transition Analysis: Structural Similarity.

```
%OVERALL%
c2 on c1;
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (vv1-vv3);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (vv4-vv6);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (vv7-vv9);
```


Latent Transition Analysis: Dispersion Similarity

```

%OVERALL%
c2 on c1;
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);

```

Latent Transition Analysis: Distributional Similarity.

Labels are used to request that the sizes of the profiles be invariant over time.

```

%OVERALL%
c2 on c1;
[ c1#1] (p1);
[ c1#2] (p2);
[ c2#1] (p1);
[ c2#2] (p2);
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);

```

Latent Transition Analysis with Predictors: Relations freely estimated at both time points

To ensure stability, starts values from the previously most invariant solution should be used.

```
%OVERALL%
c2 on c1;
[ c1#1] (p1); [ c1#2] (p2);
[ c2#1] (p1); [ c2#2] (p2);
c1 ON Pred1;
c2 ON Pred2;
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);
```

Latent Transition Analysis with Predictors: Predictive Similarity.

To ensure stability, starts values from the previously most invariant solution should be used.

```
%OVERALL%
c2 on c1;
[ c1#1] (p1); [ c1#2] (p2);
[ c2#1] (p1); [ c2#2] (p2);
c1 ON Pred1 (pr1-pr2); ! one less label than the number of latent profiles
c2 ON Pred2 (pr1-pr2);
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);
```

Latent Transition Analysis with Outcomes: Relations freely estimated at both time points

To ensure stability, starts values from the previously most invariant solution should be used.

```

%OVERALL%
c2 on c1;
[ c1#1] (p1); [ c1#2] (p2);
[ c2#1] (p1); [ c2#2] (p2);
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
[Out1] (oa1);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
[Out1] (oa2);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
[Out1] (oa3);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
[Out2] (ob1);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
[Out2] (ob2);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);
[Out2] (ob2);
MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;

```

Latent Transition Analysis with Outcomes: Explanatory Similarity

To ensure stability, starts values from the previously most invariant solution should be used.

```

%OVERALL%
c2 on c1;
[ c1#1] (p1); [ c1#2] (p2);
[ c2#1] (p1); [ c2#2] (p2);
MODEL C1:
%c1#1%
[AC1 NC1 CC1] (m1-m3); AC1 NC1 CC1 (v1-v3);
[Out1] (oa1);
%c1#2%
[AC1 NC1 CC1] (m4-m6); AC1 NC1 CC1 (v4-v6);
[Out1] (oa2);
%c1#3%
[AC1 NC1 CC1] (m7-m9); AC1 NC1 CC1 (v7-v9);
[Out1] (oa3);
MODEL C2:
%c2#1%
[AC2 NC2 CC2] (m1-m3); AC2 NC2 CC2 (v1-v3);
[Out2] (oa1);
%c2#2%
[AC2 NC2 CC2] (m4-m6); AC2 NC2 CC2 (v4-v6);
[Out2] (oa2);
%c2#3%
[AC2 NC2 CC2] (m7-m9); AC2 NC2 CC2 (v7-v9);
[Out2] (oa2);
MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;

```

Mixture Regression Analysis

Mixture regression analyses specify a regression model in the %OVERALL% section of the input indicating here that (for illustration purposes) AC2 (AC measured at Time 2) is regressed (ON) a series of predictors from (AC1 NC1 CC1). Then, the profile-specific sections of the input request that these regression coefficients be freely estimated in all profiles. In the basic mixture regression model, the mean and variance of the outcome(s) also need to be freely estimated in each profiles as these respectively reflect the intercept sand residuals of the regressions.

```
%OVERALL%
AC2 ON AC1 NC1 CC1;
%c#1%
AC2 ON AC1 NC1 CC1;
AC2;
[AC2];
%c#2%
AC2 ON AC1 NC1 CC1;
AC2;
[AC2];
%c#3%
AC2 ON AC1 NC1 CC1;
AC2;
[AC2];
```

A more flexible (and perhaps realistic) representation also freely estimates the means (and variances) of the predictors in each profiles, resulting in a model that combines LPA (for predictors) and mixture regressions and provides results indicating how the regression differs as a function of latent profiles of employees defined based on their configuration on the predictors.

```
%OVERALL%
AC2 ON AC1 NC1 CC1;
%c#1%
AC2 ON AC1 NC1 CC1;
AC2 AC1 NC1 CC1;
[AC2 AC1 NC1 CC1];
%c#2%
AC2 ON AC1 NC1 CC1;
AC2 AC1 NC1 CC1;
[AC2 AC1 NC1 CC1];
%c#3%
AC2 ON AC1 NC1 CC1;
AC2 AC1 NC1 CC1;
[AC2 AC1 NC1 CC1];
```

Multiple Group Mixture Regression Analysis: Configural Similarity

This set-up is highly similar to the one used for the multiple groups LPA models. Here again, the KNOWCLASS option is used to define the status groups in the VARIABLE section:

```
KNOWCLASS = cg (status = 1 status = 2);
CLASSES = cg (2) c (3);
```

Then the Model section describes the model of configural similarity.

```
%OVERALL%
c#1 on cg#1;
c#2 on cg#1;
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4);
[AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14);
[AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24);
[AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (rr1-rr3);
AC2 AC1 NC1 CC1 (mm1-mm4);
[AC2 AC1 NC1 CC1] (vv1-vv4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (rr11-rr13);
AC2 AC1 NC1 CC1 (mm11-mm14);
[AC2 AC1 NC1 CC1] (vv11-vv14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (rr21-rr23);
AC2 AC1 NC1 CC1 (mm21-mm24);
[AC2 AC1 NC1 CC1] (vv21-vv24);
```

Multiple Group Mixture Regression Analysis: Regression Similarity

```

%OVERALL%
c#1 on cg#1; c#2 on cg#1;
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (mm1-mm4); [AC2 AC1 NC1 CC1] (vv1-vv4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (mm11-mm14); [AC2 AC1 NC1 CC1] (vv11-vv14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (mm21-mm24); [AC2 AC1 NC1 CC1] (vv21-vv24);

```

Multiple Group Mixture Regression Analysis: Structural Similarity

```

%OVERALL%
c#1 on cg#1; c#2 on cg#1;
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (vv1-vv4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (vv11-vv14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (vv21-vv24);

```

Multiple Group Mixture Regression Analysis: Dispersion Similarity

```

%OVERALL%
c#1 on cg#1; c#2 on cg#1;
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);

```

Multiple Group Mixture Regression Analysis: Distribution Similarity

The statements making profile membership conditional on status are simply taken out.

```

%OVERALL%
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);

```


Multiple Group Mixture Regression Analysis with Predictors: Relations Freely Estimated Across Subpopulations

To ensure stability, starts values from the previously most invariant solution should be used.

```
%OVERALL%
AC2 ON AC1 NC1 CC1;
c#1-c#2 ON Pred1@0 Pred2@0;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
MODEL cg:
%cg#1%
c#1-c#2 ON Pred1 Pred2;
%cg#2%
c#1-c#2 ON Pred1 Pred2;
```

Multiple Group Mixture Regression Analysis with Predictors: Predictive Similarity

To ensure stability, starts values from the previously most invariant solution should be used.

```
%OVERALL%
AC2 ON AC1 NC1 CC1;
c#1-c#2 ON Pred1 Pred2;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
```

Multiple Group Mixture Regression Analysis with Outcomes: Relations Freely Estimated Across Subpopulations

To ensure stability, starts values from the previously most invariant solution should be used.

```

%OVERALL%
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
[Out1] (oa1);
[Out2] (ob1);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
[Out1] (oa2);
[Out2] (ob2);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
[Out1] (oa2);
[Out2] (ob2);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
[Out1] (oaa1);
[Out2] (obb1);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
[Out1] (oaa2);
[Out2] (obb2);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
[Out1] (oaa3);
[Out2] (obb3);
MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2;
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;
NEW (yy12); yy12 = oaa1-oaa2;
NEW (yy13); yy13 = oaa1-oaa3;
NEW (yy23); yy23 = oaa2-oaa3;
NEW (zz12); zz12 = obb1-obb2;
NEW (zz13); zz13 = obb1-obb3;
NEW (zz23); zz23 = obb2-obb3;

```

Multiple Group Mixture Regression Analysis with Outcomes: Explanatory Similarity

To ensure stability, starts values from the previously most invariant solution should be used.

```

%OVERALL%
AC2 ON AC1 NC1 CC1;
%cg#1.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
[Out1] (oa1);
[Out2] (ob1);
%cg#1.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
[Out1] (oa2);
[Out2] (ob2);
%cg#1.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
[Out1] (oa2);
[Out2] (ob2);
%cg#2.c#1%
AC2 ON AC1 NC1 CC1 (r1-r3);
AC2 AC1 NC1 CC1 (m1-m4); [AC2 AC1 NC1 CC1] (v1-v4);
[Out1] (oa1);
[Out2] (ob1);
%cg#2.c#2%
AC2 ON AC1 NC1 CC1 (r11-r13);
AC2 AC1 NC1 CC1 (m11-m14); [AC2 AC1 NC1 CC1] (v11-v14);
[Out1] (oa2);
[Out2] (ob2);
%cg#2.c#3%
AC2 ON AC1 NC1 CC1 (r21-r23);
AC2 AC1 NC1 CC1 (m21-m24); [AC2 AC1 NC1 CC1] (v21-v24);
[Out1] (oa3);
[Out2] (ob3);
MODEL CONSTRAINT:
NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;

```

Appendix D

Growth Mixture Analyses

For y observed indicators and k latent profiles, LPA is expressed as (e.g., Peugh & Fan, 2013):

$$\sigma_y^2 = \sum_{k=1}^K \pi_k (\mu_{yk} - \mu_y)^2 + \sum_{k=1}^K \pi_k \sigma_{yk}^2$$

This model decomposes the variance into between-profile (the first term) and within-profile (the second term) components, where the profile-specific means (μ_{yk}) and variances (σ_{yk}^2) are expressed as a function of the density function π_k reflecting the proportion of participants in each profile.

GMA are built from the combination of this LPA model with latent curve models (LCM, see Bentein, this volume; Bollen & Curran, 2006; McArdle & Epstein, 1987; Meredith & Tisak, 1990). GMA thus identify subgroups (i.e., profiles) that follow distinct trajectories (e.g., Morin, Maïano et al. 2011). A linear GMA for the repeated measure y_{it} for individual i at time t is estimated within k distinct levels ($k = 1, 2, \dots, K$) of the unobserved latent categorical variable c representing the profiles, with each individual having a probability of membership in the k levels.

$$y_{it} = \sum_{k=1}^K p_k [\alpha_{iyk} + \beta_{iyk} \lambda_t + \varepsilon_{yitk}] \quad (1)$$

$$\alpha_{iyk} = \mu_{\alpha yk} + \zeta_{\alpha yik} \quad (2)$$

$$\beta_{iyk} = \mu_{\beta yk} + \zeta_{\beta yik} \quad (3)$$

The k subscript indicates that most parameters are allowed to differ across profiles. This means that each profile can be defined by its own LCM. In this equation, α_{iyk} and β_{iyk} represent the random intercept and random linear slope of the trajectory for individual i in profile k ; $\mu_{\alpha yk}$ and $\mu_{\beta yk}$ represent the average intercept and linear slope in profile k and $\zeta_{\alpha yik}$ and $\zeta_{\beta yik}$ represent the variability of the intercepts and slopes across cases within profiles. ε_{yitk} represents the time-individual- and class- specific residual. These errors are generally allowed to vary across time. p_k

defines the probability that an individual i belongs to class k with all $p_k \geq 0$ and $\sum_{k=1}^K p_k = 1$. The

variance parameters ($\zeta_{\alpha_{iyk}}, \zeta_{\beta_{iyk}}$) have a mean of zero and a Φ_{yk} variance-covariance matrix:

$$\Phi_{yk} = \begin{bmatrix} \psi_{\alpha\alpha_{yk}} & \\ \psi_{\alpha\beta_{yk}} & \psi_{\beta\beta_{yk}} \end{bmatrix} \quad (4)$$

In these models, Time is represented by λ_t , the factor loading matrix relating the time-specific indicators to the linear slope factor. Time is coded to reflect the passage of time and is thus a function of the intervals between measurement points. Assuming a study including four equally spaced monthly measurement points of newcomers starting a new employment, it is reasonable to set the intercept at Time 1 [$E(\alpha_{iyk}) = \mu_{y1k}$]. Thus, for a linear GMM, time would be coded $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$, $\lambda_4 = 3$. For additional discussion of time codes, the readers should consult Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004) and Metha and West (2000).

Alternative Functional Forms (Shape) for the Trajectories

Linear GMA. In the estimation of GMA, a first decision that has to be taken is related to the expected shape of the trajectories. The most common functional forms are part of the polynomial family, with the two most common being the linear and quadratic forms. The linear GAM model, presented in equations 1 to 4, assumes that all longitudinal trajectories will be linear, characterized either by a steady increase over time, a steady decrease over time, or longitudinal stability. This is the most basic form of GMA, and its estimation requires a minimum of three measurement points.

Quadratic GMA. A curvilinear (quadratic) GMA adds one quadratic slope factor to the model (e.g., Diallo, Morin, & Parker, 2014), providing a way to model U-shape, or inverted U-Shape trajectories. The estimation of a quadratic slope requires four time points. In this model, α_{iyk} remains defined as in equation 2, λ_t remains coded as in the linear model, and β_{1iyk} and β_{2iyk} respectively represent the random linear and quadratic slopes for individual i in profile k .

$$y_{it} = \sum_{k=1}^K p_k [\alpha_{iyk} + \beta_{1iyk} \lambda_t + \beta_{2iyk} \lambda_t^2 + \varepsilon_{yitk}] \quad (5)$$

$$\beta_{1iyk} = \mu_{\beta 1,yk} + \zeta_{\beta 1,yik} \quad (6)$$

$$\beta_{2iyk} = \mu_{\beta 2,yk} + \zeta_{\beta 2,yik} \quad (7)$$

$$\Phi_{yk} = \begin{bmatrix} \psi_{\alpha\alpha yk} & & \\ \psi_{\alpha\beta 1,yk} & \psi_{\beta 1\beta 1,yk} & \\ \psi_{\alpha\beta 2,yk} & \psi_{\beta 1\beta 2,yk} & \psi_{\beta 2\beta 2,yk} \end{bmatrix} \quad (8)$$

Piecewise GMA. Piecewise GMA are particularly useful when there is an expected transition point over the course of the study (transition, job change, intervention, etc.). Piecewise models allow for the estimation of a change in the direction of longitudinal trajectories before and after a transition point. Typically, this is reflected by the integration of two linear slopes to the model, the first representing the pre-transition slope, and the second representing the post-transition slope (e.g., Diallo & Morin, 2015):

$$y_{it} = \sum_{k=1}^K p_k [\alpha_{iyk} + \beta_{1iyk} \lambda_{1t} + \beta_{2iyk} \lambda_{2t} + \varepsilon_{yik}] \quad (9)$$

In this model, α_{iyk} is defined as in equation 2, Φ_{yk} is defined as in equation 8, β_{1iyk} and β_{2iyk} are linear slopes reflecting the growth occurring before and after the transition. In this model, two distinct sets of time scores λ_{1t} and λ_{2t} are used to reflect the passage of time before, and after the transition. Assuming a study including six equally spaced measurements points with an intercept located at Time 1 [$E(\alpha_{iyk}) = \mu_{y1k}$], and a transition point located after the third time point, λ_{1t} would be coded {0, 1, 2, 2, 2, 2} for time $\lambda_{1t=1}$ to $\lambda_{1t=6}$ (reflecting linear growth for the first three time points after which the equal loadings allow the remaining growth information to be absorbed by the second linear slope factor) and λ_{2t} would be {0, 0, 0, 1, 2, 3} for time $\lambda_{2t=1}$ to $\lambda_{2t=6}$ (reflecting linear growth between the last three time points). Piecewise GMA require at least 2 measurement points before and after the transition point, with a total of five measurement point in total (e.g., Diallo & Morin, 2015). With additional time points, the piecewise model may be extended to model curvilinear trends.

Latent Basis GMA. A limitation of typical implementations of GMA is that the same functional form needs to be estimated in all profiles. This requirement is not, however, as restrictive as it seems given that it is easy to constrain one or many of the slope factors to be zero in specific profiles. For

instance, by constraining the quadratic slope to be zero in one profile within a quadratic GMA, one constrains this profile to follow a linear trajectory. Similarly, by constraining the second slope to be equal to the first one in a piecewise model, one constrains the trajectories to follow the same course before and after the transition point. More restrictive is the fact that all of these models involve estimating profiles that remain within the same family of polynomial trajectories (i.e., it is not possible to combine the estimation of exponential, logistic, and quadratic trajectories within the same model). The latent basis GMA provides an interesting way around this problem. In LCM/GMA, only two time codes λ_t need to be fixed to 0 and 1 for identification purposes, while the remaining codes can be freely estimated (Ram & Grimm, 2009). As a result, the slope factor represents the total amount of change occurring between these two time points, and the loadings reflect the proportion of the total change occurring at specific time points – thus providing a way to estimate GMA without imposing any functional form on the trajectories. When these loadings are freely estimated across profiles, the latent basis model thus provides a way to estimate completely distinct trajectories across profiles (for an example, see Morin, Maïano et al., 2013), without imposing any functional shape.

The latent basis model is expressed as in equations 1 to 4, but $t-2$ time codes are freely estimated in λ_t . This model further provides the possibility to freely estimate these $t-2$ times codes in all profiles so that λ_t becomes λ_{tk} , allowing for the extraction of trajectories differing completely in shape across profiles. In this model, $\mu_{\beta_{yk}}$ then reflects the total change occurring between the two time points codes 0 and 1, and the freely estimated loadings represent the proportion of change ($\mu_{\beta_{yk}}$) occurring at each specific time point (and come with significance tests).

Additional non-linear GMA specifications. For a description of additional functional forms, interested readers should consult: Blozis, 2007; Browne and DuToit, 1991; Grimm, Ram, and Hamagami, 2011; Grimm et al., 2010; Ram and Grimm, 2007, 2009.

Alternative Parameterizations of GMA

As discussed above, LPA models can be estimated using more or less restricted parameterizations depending on whether the variances of the profile indicators are constrained to be invariant across profiles. In GMA, the situation is even more complex as the profiles can be defined while allowing any

or all parameters from a LCM (the mean of the growth factors, the variances-covariances of the growth factors, the time-specific residuals of the growth factors) to be freely estimated in all profiles.

However, fully variant GMA are seldom estimated, potentially because more complex models run more frequently into estimation and convergence problems. Among simpler parameterizations, Nagin's (1999) group-based latent class growth analysis (LCGA) constrains the variances of the growth factors (e.g., $\alpha_{iyk}, \beta_{1iyk}, \beta_{2iyk}$) to be zero, thus taking out the latent variance-covariance matrix from the model ($\Phi_{yk} = 0$). LCGA thus force all members to follow the exact same trajectory. LCGA also typically assumes the time-specific residuals to be equal across profiles ($\varepsilon_{yitk} = \varepsilon_{yit}$). Another common restricted GMA is linked to the defaults of the Mplus package (Muthén & Muthén, 2014), which freely estimate $\mu_{\alpha yk}, \mu_{\beta 1 yk}$ and $\mu_{\beta 2 yk}$ in all profiles but constrain the latent variance-covariance parameters and the residuals to be equal across the profiles ($\Phi_{yk} = \Phi_y$ and $\varepsilon_{yitk} = \varepsilon_{yit}$).

Although these restrictions are common, simulation studies have shown that similar restrictions could result in the over-extraction of latent classes and biased parameter estimates in the context of mixture models (e.g., Bauer & Curran, 2004; Enders & Tofighi, 2008; Lubke & Muthén, 2007; Lubke & Neale, 2006, 2008; Magidson & Vermunt, 2004). In discussing these restrictions, Morin, Maïano et al. (2011) presented them as untested invariance assumptions that are unlikely to hold in real life and generally fails to be supported when empirically tested. Using a real data set, they further showed that these restricted parameterizations could result in drastically different conclusions. Unfortunately, arguments supporting the adequacy of these restricted parameterizations are seldom provided in applied research, and tests of these assumptions (which are easy to conduct using the information criteria and LRTs) are almost never implemented.

As for LPA, more flexible models are likely to provide a much richer perspective, although the ability to estimate more flexible models is likely to be limited with smaller samples, or fewer time points. For this reason, we recommend to always start with the estimation of more complex models allowing the profiles to be defined based on the complete LCM, and then to slowly impose constraints when less restricted models fail to converge on proper solutions. It should be noted that, as for LCM,

sample size issues for GMA are not limited to the number of participants, but it is also necessary to take into account the number of measurement points so that more measurement occasions can offset sample size limitations (Diallo & Morin, 2015; Diallo, Morin, & Parker, 2014).

In sum, whenever possible, we suggest that GMA be estimated with a fully independent within-profile LCM: $\mu_{\alpha yk}$, $\mu_{\beta 1 yk}$, $\mu_{\beta 2 yk}$, $\zeta_{\alpha yik}$, $\zeta_{\beta 1 yik}$, $\zeta_{\beta 2 yik}$, Φ_{yk} , ε_{yitk} , and even λ_{tk} in latent basis models. Should GMM users face convergence problems, we suggest that the following sequence of constraints should be implemented: (1) $\varepsilon_{yitk} = \varepsilon_{yit}$; (2) $\Psi_{\alpha\beta 1 yk}$, $\Psi_{\alpha\beta 2 yk}$, $\Psi_{\beta 1\beta 2 yk} = \Psi_{\alpha\beta 1 y}$, $\Psi_{\alpha\beta 2 y}$, $\Psi_{\beta 1\beta 2 y}$; (3) $\Phi_{yk} = \Phi_y$; (4) $\Phi_{yk} = 0$. However, this sequence should not be followed blindly and should be adapted to the specific research question and context.

Input for a Linear GMA

In LCM, the “I S |” function serves as a shortcut to define longitudinal intercepts and slope parameters and are generally followed by a specification of the time-varying indicators and their time codes (loadings on the slope factor). In this input, we request the means of the intercepts and slope factors ([I S]);, their variances (I S;) and covariances (I WITH S;) and all time specific residuals (AC1, AC2, AC3, AC4, AC5, AC6;) be freely estimated in all profiles. We assume here six repeated measures of AC, equally spaced, with an intercept located at Time 1.

```
%OVERALL%
I S | AC1@0 AC2@1 AC3@2 AC4@3 AC5@4 AC6@5;
%c#1%
I S; [I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#2%
I S; [I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#3%
I S; [I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
```

The following function provides plots of the trajectories.

```
PLOT:
TYPE IS PLOT3;
SERIES = CP4-CS4(*);
```

Input for a Quadratic GMA

A quadratic slope factor (Q) is simply added to the previous model.

```
%OVERALL%
I S Q | AC1@0 AC2@1 AC3@2 AC4@3 AC5@4 AC6@5;
%c#1%
I S Q;
[I S Q];
I WITH S Q;
S WITH Q;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#2%
I S Q;
[I S Q];
I WITH S Q;
S WITH Q;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#3%
I S Q;
[I S Q];
I WITH S Q;
S WITH Q;
AC1 AC2 AC3 AC4 AC5 AC6;
```

Input for a Piecewise GMA

Two linear slope factors (S1 and S2) are defined to represent change before and after the transition.

```
%OVERALL%
I S1 | AC1@0 AC2@1 AC3@2 AC4@2 AC5@2 AC6@2;
I S2 | AC1@0 AC2@0 AC3@0 AC4@1 AC5@2 AC6@3;
%c#1%
I S1 S2;
[I S1 2];
I WITH S1 S2;
S WITH S2;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#2%
I S1 S2;
[I S1 2];
I WITH S1 S2;
S WITH S2;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#3%
I S1 S2;
[I S1 2];
I WITH S1 S2;
S WITH S2;
AC1 AC2 AC3 AC4 AC5 AC6;
```

Input for a Latent Basis GMA

In a latent basis model, two loadings (typically the first and last) need to be respectively fixed to 0 and 1 (@0 and @1) while the others are freely estimated. Here, we also request that these be freely estimated in all profiles by repeating this function in the profile-specific sections.

```
%OVERALL%
I S | AC1@0 AC2* AC3* AC4* AC5* AC6*;
%c#1%
I S | AC1@0 AC2* AC3* AC4* AC5* AC6*;
I S;
[I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#2%
I S | AC1@0 AC2* AC3* AC4* AC5* AC6*;
I S;
[I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
%c#3%
I S | AC1@0 AC2* AC3* AC4* AC5* AC6*;
I S;
[I S];
I WITH S;
AC1 AC2 AC3 AC4 AC5 AC6;
```

References used in Appendix D

- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3-29.
- Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The Role of Coding Time in Estimating and Interpreting Growth Curve Models. *Psychological Methods, 9*(1), 30-52.
- Blozis, S. A. (2007). On fitting nonlinear latent curve models to multiple variables measured longitudinally. *Structural Equation Modeling, 14*, 179-201.
- Bollen, K.A., & Curran, P.J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.
- Browne, M. W., & du Toit, S. H. C. (1991). Models for learning data. In L. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47–68). Washington, DC: APA.
- Diallo, T.M.O, & Morin, A.J.S. (2015, In Press). Power of Latent Growth Curve Models to Detect

Piecewise Linear Trajectories. *Structural Equation Modeling*.

Diallo, T.M.O., Morin, A.J.S., & Parker, P.D. (2014). Statistical Power of Latent Growth Curve Models to Detect Quadratic Growth. *Behavior Research Methods*, *46*, 357-371

Enders, C.K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling*, *15*, 75-95.

Grimm, K.J., Ram, N. & Estabrook, R. (2010). Nonlinear structured growth mixture models in Mplus and OpenMx. *Multivariate Behavioral Research*, *45*, 887-909.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357-1371.

Lubke, G. & Muthén, B. (2007). Performance of factor mixture models as a function of model size, criterion measure effects, and class-specific parameters. *Structural Equation Modeling*, *14*, 26–47.

Lubke, G. & Neale, M. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, *41*, 499-532

Lubke, G. & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models? *Multivariate Behavioral Research*, *43*, 592-620

Magidson, J., & Vermunt, J.K. (2004). Latent class models. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 175-198). Newbury Park, CA: Sage.

McArdle, J.J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.

Metha, P.D., & West, S.G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*, 23-43.

Morin, A.J.S., Maïano, C., Marsh, H.W., Nagengast, B., & Janosz, M. (2013). School life and adolescents' self-esteem trajectories. *Child Development*, *84*, 1967-1988.

Morin, A.J.S., Maïano, C., Nagengast, B., Marsh, H.W., Morizot, J., & Janosz, M. (2011). Growth mixture modeling of adolescents trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling*, *18*, 613-648.

- Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4*, 139-157.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling, 20*, 616–639.
- Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development, 31*, 303-316.
- Ram, N., & Grimm, K.J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*, 565-576.