Getting started with Latent Profile Analysis (LPA) in the psychological sciences: A tutorial using the tidyLPA R package that provides an interface to open-source and commercial software

Joshua Rosenberg[1], Caspar van Lissa[2], Jennifer Schmidt[3], Patrick Beymer[3], Daniel Anderson[4], Matthew Schell[3], Rebecca Steingut[5], & Stephanie Wormington[6]

[1] University of Tennessee, Knoxville

[2] Utrecht University

[3] Michigan State University

[4] University of Oregon

[5] Columbia University

[6] University of Virginia

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Joshua Rosenberg, 1122 Volunteer Blvd., Knoxville, TN, 37996. E-mail: jmrosenberg@utk.edu

Abstract

Keywords: Latent Profile Analysis, mixture models, finite mixture models, tutorial, R, MPlus, mclust

Word count: X

Getting started with Latent Profile Analysis (LPA) in the psychological sciences: A tutorial using the tidyLPA R package that provides an interface to open-source and commercial software

Background

Person-oriented analysis involves a statistical approach to exploring psychological constructs in a way that captures and simplifies complexity. This approach can be used to consider the way in which psychological constructs, for example, are experienced together and at once. Thus, conceptually, this approach emphasizes how individuals go about interacting, socializing, and behaving in this way, experiencing these activities as a whole person. Person-oriented analysis, developed within developmental science, focuses on how individuals have experiences in a holistic way and also how these ways can be identified through common groups, or profiles.(Bergman & El-Khouri, 1997; Magnusson & Cairns, 1996). One way to distinguish the approach is to consider it in contrast to variable-centered approaches, such as those for which the covariability between variables, rather than the relations between profiles of individuals, is the focus (Bergman & Trost, 2006).

Though studies examining learning from a person-oriented perspective are not very common, some examples include studies of intrinsic and extrinsic motivation (Corpus & Wormington, 2014; Hayenga & Corpus, 2010), profiles of achievement goals (see Wormington & Linnenbrink-Garcia, advance online publication, for a review), and epistemic cognition (Trevors, Kendeou, Braten, & Braasch, 2017). Recently, too, there have been excellent reviews of person-oriented analyses in motivation and educational psychology (Linnenbrink-Garcia & Wormington, 2017).

In this article, our goal is to review person-oriented analysis focused on getting started (on both knowing about and doing person-oriented analyses) in an informed way. More specifically, we start with an overview of the methodology, describe specific software tools, explain analytic choices made in the course of their use, and conclude with general recommendations One challenge facing those trying to learn about and apply person-oriented analysis in their research is the multitude of terms used for similar analytic approaches or ideas. Throughout this manuscript, we aim to point these out and explain when and why a variety of terms have been used. Thus, this manuscript is written for an informed and curious reader at any level interested in understanding this particular approach.

Overview of Methodologies

At present, there are a number of approaches to carrying out person-oriented analysis when the indicators, of variables, used to find the groups are continuous. This section is focused upon two approaches: cluster analysis and Latent Profile Analysis (LPA). We focus on these because they are relatively simple and accessible forms of person-oriented analysis that can also accommodate cross sectional and longitudinal data alike. Just as a brief point of contrast, when the indicators for the groups are dichotomous (or categorical and can therefore be "dummy-coded" to be dichotomous), an approach closely related to LPA, Latent Class Analysis, can be used (cite PSU folks).

Cluster Analysis

Cluster analysis is an approach that has been frequently applied to analysis of topics in educational psychology (such as motivation and engagement) from a more holistic, whole-person approach. One particular way in which cluster analysis has been used has been as part of an approach termed I-States-as-Objects Analysis (ISOA). The purpose of ISOA is to identify profiles at specific time points, and then to explore how individuals do or do not shift between profiles across time points (needs cite). In their early work, Bergman and Magnusson (1999) and colleagues (1999) described a two-step approach to cluster analysis as part of their I-States-as-Objects analysis. In this approach, all of the observations are clustered. The key idea behind the approach is that each observation represents not a permanent group, but rather a status, or an i-state. These i-states, are considered independently for the purposes of constructing profiles. Then, shifts between these i-states across time (say, two or three time points) are evaluated.

Commonly, cluster analysis carried out as part of ISOA in fact involves two distinct clustering steps. This is because different clustering algorithms have different strengths–as well as weaknesses. In ISOA, the first step is Ward's hierarchical cluster analysis (Bergman & El-Khouri, 1999). In this step, observations are split, step-by-step, based upon their multivariate similarity. The analyst provides the number of groups which will be used and interpreted, though technically most algorithms will continue splitting the groups until each observation makes up its own. This type of clustering is deterministic: The same results will be obtained each time it is run. In the second step of the cluster analysis, an iterative, k-means cluster analysis is carried out (Bergman and El-Khouri, 1999). For this step, the algorithm partitions the observations into groups, the success of the clustering based on a criterion such as the variability in the data explained by the groups, and then "moves," when possible, the observations that are not in the group they are most similar to to the one that they are most similar to. When no more such moves are possible, the algorithm stops. On its own, k-means is not deterministic: Depending on the starting values, it can obtain different results the different times that it is run. So, the results from the hierarchical cluster analysis, the first step, are used as start values to initialize the k-means clustering.

The rationale behind using these two approaches is that the hierarchical clustering provides reasonable guesses for the profile membership, and then the iterative clustering optimizes the clustering solution. To sum up, then, cluster analysis as part of ISOA in fact involves two separate cluster analyses. The merits of this approach are to leverage the benefits of each: the iterative clustering starts with reasonable cluster solutions generated from the hierarchical clustering and optimizes them.

Latent Profile Analysis

Another approach what is becoming increasingly common for carrying out person-oriented analysis is Latent Profile Analysis (LPA). LPA is a type of finite mixture model (Harring and Hodis, 2017). The goal of this approach is estimate the parameters for a number of distributions (often multivariate) from a single dataset. The approach can be thought of as seeking an answer to the question, Do the data in a sample come from more than one population? Thus, such an approach is model-based, and some descriptions in the literature refer to it as model-based clustering (Hennig, Meila, Murtagh, & Rocci, 2015; Scrucca, Fop, Murphy, & Raftery, 2017). Thus, one distinction between LPA and cluster analytic approaches is that LPA is model-based; instead of using algorithms to group together cases, LPA seeks to estimate parameters - in terms of variances and covariances and how they are the same or different across profiles - that best characterize the different distributions. Then, this approach seeks to assign to each observation a probability that the observation is a sample from the population associated with each profile (or mixture component).

Because LPA is model-based, a number of different model parameterizations can be estimated. These models differ in terms of whether–and how–parameters are estimated across the profiles. These parameters are the means for the different profiles, which, in this approach, always are estimated freely across the profiles; the variances for the variables used to create the profiles, which can be estimated freely or can be estimated to be the same, or equal, across profiles; and the covariances of the variables used to create the profiles, which can be freely-estimated, estimated to be equal, or fixed to be zero. One challenge facing the analyst using LPA is that these parameters and the distinct model parameterizations that can be estimated is the different terminology used. As one example, Scrucca et al. (2017) refer to these parameterizations not in terms of whether and how parameters are estimated, but rather in terms of the geometric properties of the distributions that result from particular parameterizations. Muthen and Muthen (1997-2017) and others (Pastor et al., 2007) commonly refer to local independence to mean that the covariances are fixed to zero (also described as the specification of the covariance matrix as "diagonal," because only the diagonal components, or the variances, are estimated). More information on the model parameterizations are discussed in the context of the software tool tidyLPA that we have developed.

This section outlines several LPA models and guidelines/examples for when they might be most appropriate in person-centered analysis. In general, as more parameters are estimated (i.e., those that are fixed to zero are estimated as being equal across profiles; or those estimated as being equal across profiles are freely-estimated across them), the model becomes more complex; the model may fit better, but also be overfit, meaning that the profiles identified may be challenging to replicate with another, separate data set. Even still, flexibility in terms of which models can be estimated also has affordances. For example, the varying means, equal variances, and covariances fixed to 0. A researcher might choose this model specification if she wants to model the variables to be used to create profiles that are independent. This model is very simple, as no covariances are estimated and the variances are estimated to be the same across profiles. As we estimate more parameters (and decrease the degrees of freedom), we are more likely to fit the data, but less likely to be able to replicate the model with a second set of data. In other words, more parameters may mean a loss of external validity. As we progress toward more complex models (with increasingly complex parameterization, i.e. models 3 and 4 above), then we are more likely to fit the data better.

```
A number of notes regarding person-oriented analyses are important to raise at the same
```

Specific Software Tools

SPSS is a common tool to carry out cluster analyses. SPSS provides functionality to carry out both hierarchical and k-means cluster analysis. While somewhat straightforward to carry out, particularly in SPSS's graphical user interface (GUI), there are some challenges to use of this approach. The GUI in SPSS can be challenging, even for the most able analyst, to be able to document every step with syntax, and so reproducing the entire analysis efficiently can be a challenge, both for the analyst exploring various solutions and for the reviewer looking to replicate this work. Additionally, SPSS is commercial software (and is expensive), and so analysts without the software cannot carry out this analysis.

Accordingly, we worked to develop tools, initially used by the research team and now shared more widely, to provide basic functionality for each of the two primary approaches described in this section, cluster analysis and Latent Profile Analysis (LPA). In addition to being freely-available and open-source (i.e., anyone can inspect the source code), the tools described lend themselves to an open science. It is especially important to have an open transparent methodology when using person-centered approaches because of the high dependence on researcher judgment at multiple points in the process.

Accordingly, we set out to develop packages that a) provided sensible defaults and were easy to use, but provided the option to access and modify all of the inputs to the model (i.e., low barrier, high ceiling), b) interfaced to existing tools,, and are able to translate between what existing tools are capable of and what researchers and analysts carrying-out person-oriented analyses would like to specify, c) made it easy to carry-out fully-reproducible analyses and d) were well-documented. Both of these packages–one for two-step cluster analysis and one for LPA–are designed to do one thing well, namely, quickly and easily estimate profiles using cluster analysis or a model-based approach (Latent Profile Analysis).

One way that tidyLPA is designed to be easy to use is that it assumes a "tidy" data structure (Wickham, 2014). This means that it emphasizes the use of a data frame as both the primary input and output of functions for the package. Because data is passed to and returned (in amended form, i.e., with the latent profile probabilities and classes appended to the data) from the function, it makes it easy to create plots or use results in subsequent analyses. Another noteworthy feature of tidyLPA is that it provides the same functionality through two different tools, one that is open-source and available through R, the mclust package (Scrucca et al., 2017) and one that is available through the commercial software MPlus (Muthen & Muthen, 1997-2017). Moreover, as both tools use the same maximum likelihood estimation procedure, they are benchmarked to produce the same output (see here). Also, note that we have described the model specifications with descriptions of what is estimated in terms of the variances and covariances, the common names for the models (i.e., class-varying unrestricted), and the covariance matrix associated with the parameterizationfor the six models that are possible to be estimated on the website for tidyLPA (see here).

Analytic Choices

There are a number of analytic choices that need to be made when carrying out person-oriented analyses. Because such person-oriented approaches are often more subjective (in practice) than other approaches (Linnenbrink-Garcia and Wormington, 2017), there is no one rule for determining the solution obtained. This solution is obtained on the basis of multiple decisions, such as the number of profiles selected or the modeling decisions such as what specific options are used for the cluster analysis (i.e., the distance metric used to calculate the similarity of the observations as part of the Ward's hierarchical clustering) or what parameters are estimated and how as part of LPA.

Given the subjectivity involved, it is important that researchers be transparent and work as part of a team to obtain clustering solutions. Transparency about the design and analytic choices is important so that readers can appropriately interpret the report. Researchers can enhance transparency and reproducibility by sharing detailed descriptions of methodology and document it through the use of syntax (and, if possible, data) that we share with others. Working as part of a team can help to serve as a check on several of the choices researchers make, such as over-fitting or under-fitting the model to the data. Each decision depends on multiple factors and balancing tensions. We discuss each of the key decisions listed in an analysis.

How the Data are Centering or Scaled

The data can be transformed by centering or scaling. Typically, these are done after the profiles are created, so that differences between profiles can be more easily explored. They can also be done prior to the analysis, which can be helpful for obtaining solutions when the variables are on very different scales..

Whether to Use Cluster Analysis or LPA

Whether to use cluster analysis or LPA depends in part on the nature of the data being investigated (cite the 2010 debate on this topic between Vermunt and others; MacLachlan, 2011; Steinley & Brusco, 2011a; Steinley & Brusco, 2011b; Vermunt, 2011). In some cases, cluster analysis and LPA solutions are very similar (when a constrained LPA model is fit). Broadly, if the variables used can be considered independent of one another, then cluster analysis may be justified in terms of reaching an optimal solution. However, LPA provides other benefits, such as the availability of information criteria (i.e., the Aikake Information Criteria [AIC] and the Bayesian Information Criteria [BIC] and statistical tests for the number of profiles (when a model is chosen, with a bootstrapped likelihood-ratio test). LPA also results in probabilistic assignments; this information can be used in subsequent analyses rather than the simple assignment (Collins & Lanza, 2010).

How to Choose the Number of Profiles

In the case of choosing the number of profiles (and the specification of the model / profile solution), multiple criteria, such as the BIC or the proportion of variance explained are recommended for decision-making, but also interpretability in light of theory, parsimony, and evidence from cross-validation should be considered. For cluster analysis, the proportion of variance explained can be helpful for selecting a number of candidate profiles. Then, the analyst should focus on the interpretability of the solution as well as concerns of parsimony; if two profile solutions are similar, then the more simple solution may be preferred.

How to Choose Options for the Clustering Procedure or to Select the Model Parameterization

This applies most to LPA, but in cluster analysis, there are also options about the distance metric used. In LPA, we can determine which parameters - variance and covariance - are estimated to be the same or different across all profiles, and, in the case of covariance, whether it is fixed to zero. Of course, for the profiles we are most interested in, the mean is allowed to vary across profiles. The models with more parameters freely-estimated use more degrees of freedom; thus, similar to the number of profiles selected, the balance between adding more parameters (and having a more complex model) and interpretability and parsimony must be balanced. If a more complex model fits better and is interpretable and is justifiable in terms of the data collected, then it may be preferred to simpler models. If the two models are similar in terms of their fit, then the more simple, parsimonious model should be selected.

More General Recommendations

This section highlights three general recommendations. First, scholars taking a person-oriented approach should emphasize reproducibility in carrying out analyses. This is in part due to the exploratory and interpretative nature of person-oriented approaches. To this end, presenting multiple models, as in Pastor et al. (2007), should be encouraged, rather than presenting one solution. In addition, having multiple analysts review the solutions found is encouraged. As part of this recommendation, we suggest that researchers consider how their analyses can be reproduced by other analysts outside of the research team: Sharing code and data is an important part of this work. Also related, researchers should consider how profiles are replicated across samples.

A second general recommendation considers more flexibly incorporating time, when data are collected across multiple time points, into analyses. ISOA groups all time points and doesn't make distinctions. Other approaches perform analysis separately at different timepoints (Corpus & Wormington, 2014). Some integrate time as a part of the profiles, i.e. growth mixture modeling (groups of patterns), i.e. within-person growth modeling, where there are individual growth patterns. Research to date has yet to consider additional challenges in applying person centered approaches to longitudinal data. For instance, Schmidt et al (2018) use an Experience Sampling Method (ESM) approach to collecting data and used a person-centered approach to generate profiles of students in science class. This work does not account for student-level effects. In other words, it did not model the shared variance of multiple observations of the same student.

Third, best practices in within-person or longitudinal research call for modeling the nesting structure. However, researchers have yet to successfully incorporate this practice into the person centered approach. One way to approach this is to use cross-classified mixed effects models, as in Strati, Schmidt, and Maier (2017) and in Rosenberg (2018). In such approaches, dependencies in terms of, for example, individuals responding to (ESM) surveys at the same time and repeated responses being associated with the same individuals can both be modeled, although the effects of these two sources of dependencies are not nested as in very common uses of multi-level models, but rather are cross-classified. West, Welch, & Galecki (2014) have a description of the use of multi-level models with cross-classified data and tools (including those freely available through R) that can be used to estimate them.

tidyLPA update

Rationale for the update

tidyLPA has received a major update. Your old function calls might not work as expected anymore. The new version of tidyLPA aims to make Latent Profile Analysis (LPA) even more accessible to a broad audience, by means of a user-friendly interface, and to fit more seemlessly into a "tidyverse" workflow.

As before, **tidyLPA** offers a parallel workflow for the open-source package Mclust, and the commercial package *Mplus*.

These changes necessitated some simplification of the existing functions. Advanced options are still available, but require slightly more work on the user's part. The upside of this is that, for most users, the tidyLPA workflow and documentation are substantially simplified.

Major changes

There have been major changes to the two primary functions, `estimate_profiles()` and `compare_solutions()`

- `estimate_profiles()` now estimates any number of profiles, for any number of models, using any algorithm. It has a simplified user interface. This makes it a more general function, of which the former version (which could only estimate one number of profiles for one model) is a special, restricted case.
- `compare_solutions()` now compares solutions estimated by estimate_profiles(), suggesting the optimal model and number of classes. This replaces its dual role of estimating AND comparing solutions. *compare_solutions()`gives you the same set of fit indices for Mclust and Mplus. It provides a broader range of`

`fit indices than before` *compare_solutions()`calls`AHP()“, a function which uses the Automatic Hierarchical Process to determine the optimal number of classes, based on several reputable fit indices

- `plot_profiles()` now incorporates best practices for plotting latent profile analyses: Visualizing cluster centroids, standard errors, cluster variances, and classification uncertainty by means of weighted raw data
- `impute_missing()` offers two easy-to-use single imputation options, for use with the Mclust package (which does not natively handle missing data). Mplus handles partially missing data by default.

The best way to understand these changes to **tidyLPA**, though, is by explaining the new workflow.

Example analysis using estimate_profiles()

A full analysis might look like the following:

```
## Warning: package 'tibble' was built under R version 3.5.2
```

```
## Warning: package 'purrr' was built under R version 3.5.2
```

```
## Warning: package 'stringr' was built under R version 3.5.2
```

```
## tidyLPA analysis using mclust:
##
##  Model Classes AIC     BIC     Entropy prob_min prob_max n_min n_max
##  1     3       632.652 669.124 0.795   0.891    0.937    0.030 0.630
##  BLRT_p
##  0.010
```

You can use Mplus simply by changing the package argument for `estimate_profiles()`:

A simple summary of the analysis is printed to the console (and its posterior probability). The resulting object can be further passed down a pipeline to other functions, such as `plot`, `compare_solutions`, `get_data`, `get_fit`, etc. This is the "tidy" part, in that the function can be embedded in a tidy analysis pipeline.

If you have Mplus installed, you can call the version of this function that uses MPlus in the same way, by adding the argument `package = "MplusAutomation`.

We can plot the profiles by piping the output to `plot_profiles()`.

Example analysis using compare_solutions()

The function `compare_solutions()` compares the fit of several estimated models, with varying numbers of profiles and model specifications:

```
## Compare tidyLPA solutions:
##
##  Model Classes AIC     BIC
##  1     1       678.635 694.266
##  1     2       642.228 668.280
##  1     3       638.735 675.207
##  6     1       624.765 648.212
##  6     2       625.524 675.023
##  6     3       612.808 688.358
##
## Best model according to AIC is Model 6 with 3 classes.
## Best model according to BIC is Model 6 with 1 classes.
##
```

```
## An analytic hierarchy process, based on the fit indices AIC, AWE, BIC, CLC, and KIC (
```

Deprecated functionality:

- `compare_solutions()` no longer estimates profiles. It merely compares solutions estimated by `estimate_profiles()`
- all Mplus-specific functions are deprecated. `estimate_profiles()` offers a unified interface to Mclust and Mplus. These Mplus-specific functions are deprecated:
  - `estimate_profiles_mplus()`
    * `compare_profiles_mplus()`
    * `plot_profiles_mplus()`
- estimate_profiles and compare_solutions no longer select variables, passed as unquoted variable names, from "df". Instead, they use all variables in df. To select variables prior to analysis, use:
  - The dplyr function select(df, Your, Selected, Variable, Names)
    * The base R function df[, c("Your", "Selected", "Variable", "Names")]
- estimate_profiles and compare_solutions no longer center and scale the raw data. To center and scale variables prior to analysis, use scale(df, center = TRUE, scale = TRUE)
- estimate_profiles no longer accepts arguments specific to Mplus or Mclust. Instead, you can specify additional arguments directly to the core functions `mplusModeler()` and `Mclust()`. See the help file for these functions for more details. Deprecated arguments, whose functionality is supplanted by passing arguments directly to these core functions, are:

For Mclust:

* `prior_control`

For Mplus:

* `starts`
* `m_iterations`
* `st_iterations`
* `convergence_criterion`
* `optseed`
* `cluster_ID`
* `include_VLMR`
* `include_BLRT (but BLRT is returned by default now)`

tidyLPA function with _mplus() appended require Mplus to use

Conclusion

Person-oriented analysis ia way to consider how psychological constructs are experienced (and can be analyzed) together and at once. Though described in contrast to a variable-centered approach, scholars have pointed out how person-oriented approaches are complementary to variable-centered analyses (Marsh, Ludtke, Trautwein, & Morin, 2009). A person-oriented approach can help us to consider multiple variables together and at once and in a more dynamic way, reflected in the methodological approaches for cluster analysis and LPA that identify profiles of individuals responses.

This manuscript provided an outline of how to get started with person-oriented analyses in an informed way. We provided a general overview of the methodology and described tools to carry out such an analysis. We also described specific tools, emphasizing freely-available open-source options that we have developed. Because of the inherently exploratory nature of person-oriented analysis, carrying out the analysis in a trustworthy and open way is particularly important. In this way, the interpretative aspect of settling on a solution shares some features of quantitative and qualitative research: The systematic nature of quantitative research methods (focused upon agreed-upon criteria such as likelihood-ratio tests) and qualitative research methods (focused upon the trustworthiness of both the analysis and the analyst) are important to consider when carrying out person-oriented analysis. Lastly, we made some general recommendations for future directions–and also highlighted some situations for which person-oriented approaches may not be the best and some cautions raised in past research regarding how such approaches are used.

In conclusion, as use of person-oriented approaches expand, new questions and opportunities for carrying out research in a more holistic, dynamic way will be presented. Analyzing constructs together and at once is appealing to researchers, particularly those carrying out research in fields such as education for which communicating findings to stakeholders in a way that has the chance to impact practice is important. Our aim was not to suggest that such an approach is always the goal or should always be carried out, but rather to describe how researchers may get started in an informed way as researchers seek to understand how individuals interact, behave, and learn in ways that embraces the complexity of these experiences.

References

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. Development and psychopathology, 9(2), 291-319.

Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds?. Merrill-Palmer Quarterly, 52(3), 601-632. Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). John Wiley & Sons. Corpus, J. H., & Wormington, S. V. (2014). Profiles of intrinsic and extrinsic motivations in elementary school: A longitudinal analysis. The Journal of Experimental Education, 82(4), 480-501. Corpus, J. H., & Wormington, S. V. (2014). Profiles of intrinsic and extrinsic motivations in elementary school: A longitudinal analysis. The Journal of Experimental Education, 82(4), 480-501. Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. Educational Psychologist, 51(3-4), 354-367. Hayenga, A. O., & Corpus, J. H. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. Motivation and Emotion, 34(4), 371-383. Linnenbrink-Garcia, L., & Wormington, S. V. (2017). Key challenges and potential solutions for studying the complexity of motivation in schooling: An integrative, dynamic person-oriented perspective. British Journal of Educational Psychology monograph series II. Psychological aspects of education: Current trends—the role of competence beliefs in teaching and learning. Chichester, UK: Wiley. Magnusson, D., & Cairns, R. B. (1996). Developmental science: Toward a unified framework. Cambridge, England: Cambridge University Press.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. Structural Equation Modeling, 16(2), 191-225. McLachlan, G. J. (2011). Commentary on Steinley and Brusco

(2011): Recommendations and cautions. Muthen, L. K., & Muthen, B. O. (1997-2017). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén. Rosenberg, J. M., Schmidt, J. A., Beymer, P. N., & Steingut, R. R. (2018). Interface to mclust to easily carry out Latent Profile Analysis [Statistical software for R]. https://github.com/jrosen48/tidyLPA Rosenberg, J. M., Schmidt, J. A., Beymer, P. N., & Steingut, R. R. (2017). prcr: Person-Centered Analysis. R package version 0.1.5. https://CRAN.R-project.org/package=prcr Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2017) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, pp. 205-233 Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). Handbook of cluster analysis. CRC Press. Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. Psychological Methods, 16(1), 63. Steinley, D., & Brusco, M. J. (2011). K-means clustering and mixture model clustering: Reply to McLachlan (2011) and Vermunt (2011). Trevors, G. J., Kendeou, P., Bråten, I., & Braasch, J. L. (2017). Adolescents' epistemic profiles in the service of knowledge revision. Contemporary Educational Psychology, 49, 107-120. Vermunt, J. K. (2011). K-means may perform as well as mixture model clustering but may also be much worse: Comment on Steinley and Brusco (2011). West, B. T., Welch, K. B., & Galecki, A. T. (2014). Linear mixed models: a practical guide using statistical software. Chapman and Hall/CRC. Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23.