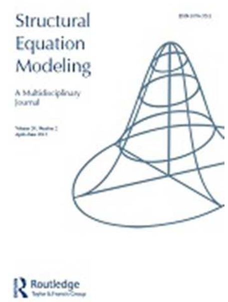

The GRoLTS-checklist: Guidelines for Reporting on Latent Trajectory Studies

Journal:	*Structural Equation Modeling: A Multidisciplinary Journal*
Manuscript ID	HSEM-2016-0031.R2
Manuscript Type:	Teacher's Corner
Keywords:	mixture modelling, SEM, LGMM/LCGA, latent classes
Abstract:	Estimating models within the mixture model framework, like Latent Growth Mixture Modeling (LGMM) or Latent Class Growth Analysis (LCGA), involves making various decisions throughout the estimation process. This has led to a high variety of how results of latent trajectory analysis are reported. To overcome this issue, using a four-round Delphi study, we developed Guidelines for Reporting on Latent Trajectory Studies (GRoLTS). The purpose of GRoLTS is to present criteria that should be included when reporting the results of latent trajectory analysis across research fields. We have gone through a systematic process to identify key components that, according to a panel of experts, are necessary when reporting results for trajectory studies. We applied GRoLTS to 38 papers where LGMM/LCGA was used to study trajectories of post-traumatic stress after a traumatic event.

The GRoLTS-checklist: Guidelines for Reporting on Latent Trajectory Studies

Rens van de Schoot[1,2] – a.g.j.vandeschoot@uu.nl

Marit Sijbrandij[3,4] - e.m.sijbrandij@vu.nl

Sonja D. Winter[1] - swinter@ucmerced.edu

Sarah Depaoli[5] - sdepaoli@ucmerced.edu

Jeroen K. Vermunt[6] - J.K.Vermunt@uvt.nl

[1] Utrecht University, Department of Methods and Statistics, The Netherlands

[2] Optentia Research Program, Faculty of Humanities, North-West University, South Africa

[3] Department of Clinical Psychology, VU University Amsterdam, The Netherlands

[4] EMGO Institute for Health and Care Research, The Netherlands

[5] University of California, Merced, Psychological Sciences, USA

[6] Tilburg University, Department of Methodology and Statistics, The Netherlands

Acknowledgement: The first and last author were supported by a grant from the Netherlands organization for scientific research: NWO-VIDI-452-14-006 and NWO-VICI-453-10-002, respectively. We would also like to thank the experts who provided feedback during the different stages of our Delphi study (in alphabetical order): Heather Armstrong, Daniel Bauer, George Bonanno, Jan Boom, Patrick Curran, Isaac Galatzer-Levy, Christian Geiser, Kevin Grimm, Joop Hox, John Hipp, Loes Keijsers, Lynda and Dan King, Todd Little, Gitta Lubke, Peter Lugtig, Katherine Masyn, Bengt Muthen, Daniel Nagin, Karen Nylund, Cecile Proust-Lima, Quinten Raaijmakers, Sophia Rabe-Hesketh, Jost Reinecke, Paula Schnurr, Anders Skrondal, Geert Smid.

Abstract

Estimating models within the mixture model framework, like Latent Growth Mixture Modeling (LGMM) or Latent Class Growth Analysis (LCGA), involves making various decisions throughout the estimation process. This has led to a high variety of how results of latent trajectory analysis are reported. To overcome this issue, using a four-round Delphi study, we developed Guidelines for Reporting on Latent Trajectory Studies (GRoLTS). The purpose of GRoLTS is to present criteria that should be included when reporting the results of latent trajectory analysis across research fields. We have gone through a systematic process to identify key components that, according to a panel of experts, are necessary when reporting results for trajectory studies. We applied GRoLTS to 38 papers where LGMM/LCGA was used to study trajectories of post-traumatic stress after a traumatic event.

Keywords: LGMM, LCGA, mixture modelling, SEM, latent classes

Methods to estimate latent trajectories[1] are becoming ever more popular across social, behavioural, and biomedical research areas. Estimating models within the mixture model framework involves making various decisions throughout the estimation process. Such decisions can impact the results, even leading to different conclusions. Despite latent trajectory analysis becoming very popular— currently being the dominant tool to analyze longitudinal data in many different fields, there is no standard for how to report results for latent trajectory models. This has led to a high variety of how results of latent trajectory analysis are reported in papers. Inadequate or incomplete reporting of the results for latent trajectory analysis hampers interpretation and critical appraisal of results, as well as comparison of results between studies.

This paper describes Guidelines for Reporting on Latent Trajectory Studies (GRoLTS). The ultimate goal of GRoLTS is to enhance the uniformity of reporting latent trajectory studies so that the results presented are fully transparent (i.e., have a high quality) and can be used for comparisons, replications, systematic reviews, and meta-analyses. In what follows we first describe the development of GRoLTS; we have gone through a systematic process, using a four round Delphi study, to identify key components that, according to a panel of experts, are necessary when reporting results for trajectory studies. Next, we provide a detailed description of each item. Finally, we present our experiences with administering GRoLTS to a set of 38 studies applying latent trajectory analyses to assess change in posttraumatic stress symptoms (PTSS) after traumatic experience. Additional information is available on the Open Science Framework (Go to: https://osf.io/vw3t7/): (1) all the details for the Delphi study; (2) additional information for some of the items which can be used for teaching purposes; and (3) the data set with the screening of the 38 PTSS papers.

[1] With latent trajectory analysis, we refer to person-centred techniques to estimate membership of unobserved subgroups of individuals developing over time (e.g., B. O. Muthén & Muthén, 2000a). To estimate trajectory membership, a conventional latent growth model (e.g., Raudenbush & Bryk, 2002) is combined with a mixture component (e.g., Vermunt, 2010b). The basic idea of latent growth modelling is the assumption that all individuals are drawn from one population. When combined with mixture modeling it is assumed that growth parameters (i.e., intercept, slope, etc.) vary across a number of pre-specified, unobserved subpopulations. This is accomplished using categorical latent variables, which allow for groups of individual growth trajectories and results in separate latent growth models for each (unobserved) group, each with its unique set of growth parameters.

The Development of GRoLTS

The development process of GRoLTS involved the following stages (cf., Streiner, Norman, & Cairney, 2014): (1) preliminary conceptual decisions; (2) item generation; (3) assessment of face validity; (4) field trials to access consistency and construct validity; and (5) the creation of the final, refined checklist. At the start of the project, we decided that GRoLTS would need to meet the following basic requirements:

- be targeted at papers where latent trajectory analyses have been used in an exploratory way to answer a substantive research question;
- summarize the requirements of what to report on latent trajectory analyses;
- allow for consistent and reliable use for researchers with different backgrounds; and
- be short and simple to complete, but at the same time it should include all of the aspects needed to guarantee replicability and transparency of findings.

During the development phase, face validity of the generated item set was assessed by a three-round Delphi procedure and a fourth round with field trials. We used the Delphi procedure to obtain consensus among experts on which criteria should be included in GRoLTS, as well as the phrasing of the items. In total, 27 experts (see acknowledgement for a list of experts) were invited to take part in the expert panel and were provided the aims of GRoLTS and its desired features. The specific details of each step, including all earlier versions of the GRoLTS are provided on the Open Science Framework (Go to: https://osf.io/vw3t7/).

User Guide to the GRoLTS

The GRoLTS is a list of 16 items (some with sub-items); see Table 1. Each item should be scored 0 (not reported) or 1 (reported). The use of GRoLTS is recommended for:

- Researchers preparing to submit a manuscript;
- Editors/Reviewers/Grant panelists to check whether all essential aspects are reported;
- Lecturers teaching their students which topics are of importance.

In what follows we explain each of the items and provide an overview of the discussions in the literature (if there are any) especially for the more complicated items. More detailed information for items 1, 2, 7 and 14 can be found on the Open Science Framework (Go to: https://osf.io/vw3t7/).

Item 1. Is the metric of time used in the statistical model reported?

The coding of time in any type of growth model has important implications for the interpretation of the results. As was shown by, for example, Eggleston, Laub, and Sampson (2004), the number of latent trajectories and their shapes appeared not to be robust to the length of the follow-up period specified; longer ranges result in more groups. Moreover, Piquero (2008) found in his systematic review of LGMM/LCGA papers applied to delinquency data that the spacing in between time points also affects the number of trajectories found. Therefore, it is of importance that the metric of time is not only transparently reported, but also that it is correctly specified. The fit of the model or the significance of the growth parameter estimates should never be used to determine the specification of the metric of time. Rather, the metric of time should be decided on prior to running the analyses, and it is completely determined by the research design. For a more in depth discussion about the metric of time we refer to the online materials which can be found on the Open Science Framework (Go to: https://osf.io/vw3t7/), and to Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004) or Duncan, Duncan, and Strycker (2013).

Item 2. Is information presented about the mean and variance of time within a wave?

In longitudinal studies, it is inevitable that there will be some variation across individuals' time intervals due to logistical reasons of data collection. This variation in assessment is called time-unstructured data, or within-wave variability. The counterpart is a time-structured study, where all individuals are assessed at exactly the same time intervals. Most longitudinal data are time-unstructured, at least to some degree. That is, not all participants are assessed at exactly the same time, see Palardy and Vermunt (2010) for an application. However, such data are often analyzed as if they were time-structured. Ignoring time-unstructuredness can lead to serious substantive misinterpretations. Singer and Willett (2003, chapter 5) found that the linear slope was overestimated when using the planned age instead of the actual age, as were the variances of the intercept and linear slope. This has been replicated by Mehta and West (2000), Hertzog and Nesselroade (2003), and in several simulation studies (see, e.g., Aydin, Leite, & Algina, 2014; Coulombe, Selig, & Delaney, 2015). We suggest including a timestamp in the dataset for each assessment that notes the exact time in between observations. This way, the degree of time variance can be computed, and reported in the methods section. Thus, random factor loadings may be used with individually-varying times of observations instead of fixed factor loadings, see Coulombe et al. (2015) for more details. See for a more detailed explanation and some graphical illustrations the online materials which can be found on the Open Science Framework (Go to: https://osf.io/vw3t7/).

Item 3a. Is the missing data mechanism reported?

Most longitudinal studies are plagued with missing data and/or dropout of participants. When describing missing data and dropout, the missing data mechanism should be reported first. Three types of mechanisms can be distinguished (Schafer & Graham, 2002): (1) Missing Completely at Random (MCAR), which means that all missing data occurred independently of all observed and non-observed variables; (2) Missing at Random (MAR), which means that the missing data may depend on observed variables but do not depend on unobserved variables; and (3) Missing not at Random (MNAR), which means that attrition is related to unobserved variables. We can never know whether we are in a MAR or a NMAR situation (i.e., this cannot be tested), and one can only do as much as they can to ensure the missing data fall under the MAR assumption. Statistical models like LGMM/LCGA assume the situation of MAR. As long as attrition is not systematic in a specific way, MAR is quite realistic with longitudinal data because we already have several measurements for all persons (and missingness is assumed to be random given a person’s score on these observed measurements).

Item 3b. Is a description provided of what variables are related to attrition/missing data?

As showed by Asendorpf, Van De Schoot, Denissen, and Hutteman (2014), even small and non-significant selective dropout effects from wave-to-wave can accumulate over the course of a longitudinal study such that the results become increasingly biased (see also Rubin & Little, 2002). Therefore, researchers should compare individuals who have dropped out to individuals who completed the study on relevant characteristics. The variables related to attrition, also called auxiliary variables, can then be included as either covariates in the model (and fit under MAR) or used in the multiple imputation (MI) model. The advantage of MI is that one can separate the missing data treatment from the model of interest;

Item 3c. Is a description provided for how missing data were handled in the analyses?

The way missing data is dealt with in the analyses is the third thing to report about missing data. See, among many other papers, Peeters, Zondervan-Zwijnenburg, Vink, and van de Schoot (2015) for a comparison of different imputation methods. Currently, a rather general and flexible method for dealing with missing data is to implement multiple imputation using chained equations, also called Predictive Mean Matching (Van Buuren & oudshoorn, 2005); see Pietrzak et al. (2014) for an application to LGMM data.

Item 4.Is information about the distribution of the observed variables included?

The dependent variables in latent trajectory analyses can take different forms. Often it is just assumed that the variables are measured on a continuous scale and are normally distributed within classes. However, this is not always the case. It may be the case that the dependent variables are not measured on a continuous scale, but are categorical (e.g., a Likert-type scale with 5 answering categories), count data (e.g., counting the number of symptoms someone has), or zero-inflated (e.g., 80-90% of the participants have a zero score). As stated by Vermunt (2011), one should be critical with regard to the within-cluster normal distribution assumption. Vermunt advises not to use a mixture model for continuous responses, but instead to use a mixture model for discrete responses assuming multinomial within-cluster distributions (opposed to normal). Bauer and Curran (2003a, 2003b, 2004) showed that when assumptions about the distribution of the variables are violated, that is, when the actual outcome distribution is non-normal, then a model with multiple trajectory groups could be preferred even though only one group was actually present (see also, Hoeksma & Kelderman, 2006). The latent trajectory framework can easily deal with these types of variables by simply 'telling' the software the scale of the outcomes and overextraction of latent classes may be avoided. Another option is to use latent variables (for an application in LGMM, see, Nash et al., 2014), where the measurement structure is taken into account. That is, the individual items are used instead of sum scores. If latent variables are to be meaningfully implemented in the model, then the measurement structure(s) of the latent factor(s) and the survey items should be stable over time. That is, the measurement structure needs to be "time-invariant". This is called measurement invariance (see, e.g., Van de Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015), which is a crucial assumption to check because it can have a large impact on results and it does not always hold (see, e.g., Lommen, Van De Schoot, & Engelhard, 2014).

Item 5. Is the software mentioned?

There are several different software packages that can be used to estimate latent trajectory studies: LatentGold (Vermunt & Magidson, 2013), Mplus (L. K. Muthén & Muthén, 2013), SAS Proc Traj (Jones, Nagin, & Roeder, 2001), Stata GLLAMM (Rabe-Hesketh, Skrondal, & Pickles, 2004), the R package LCMM (Proust-Lima, Philipps, & Liquet, 2015), the R package OpenMx (Boker et al., 2011), etc. All of these software packages have different ways of how the default model is specified. For example, in Mplus the default setting is for covariances and (residual) variances to be constrained across classes. In contrast, this is not the case in LatentGold, which uses posterior-mode-estimation using priors for the residual variances to prevent these from becoming zero. For replicability purposes, it is of utmost importance to provide information about which software has been used, as well as the version (because the algorithms under the hood might have been adjusted in version-updates). In the next item, we discuss the specification of the variance-covariance matrix in more detail.

Item 6a. Are alternative specifications of <u>within</u>-class heterogeneity considered (e.g. LGCA versus LGMM) and clearly documented?

In setting-up the latent trajectory model, there are many choices to be made for how exactly the model can be specified. The first method deals with within-class heterogeneity, which is in reference to the variance around the growth parameters within the latent classes. There are two types of latent growth models that account for unobserved groups. If variance around the growth parameters is estimated within a latent trajectory, then this modeling flexibility is called Latent Growth Mixture Modeling (LGMM; Muthen, 2001; B. O. Muthén, 2003, 2006; B. O. Muthén & Muthén, 2000a, 2000b; B. O. Muthén & Shedden, 1999). If all individual growth trajectories within a class are assumed to be homogeneous, and the variance and covariance estimates for the growth factors within each class are assumed to be fixed to zero, then this is called Latent Class Growth Analysis (LCGA; Nagin, 1999, 2005; Nagin & Land, 1993; Nagin & Tremblay, 2001). The difference between LGMM and LCGA is nicely summarized by Croudace, Jarvelin, Wadsworth, and Jones (2003), Erosheva, Matsueda, and Telesca (2014), Feldman, Masyn, and Conger (2009), Jung and Wickrama (2008), Kreuter and Muthén (2008), or Twisk and Hoekstra (2012).

Nagin (1999); (Nagin & Land, 1993) took a theoretical approach and introduced two conceptualizations of latent trajectory models: (1) as approximations of a continuous but unknown distribution of population heterogeneity, or (2) as concrete trajectories that can be treated as substantively important entities. In this latter approach the trajectories are given descriptive names, and discussed as distinct entities. Most researchers take the latter approach, as was found in a systematic review by Erosheva et al. (2014, pp. 325-326). However, these authors also found that true discovery of distinct groups of trajectories was relatively rare. As one expert in our study noted: "*I have yet to encounter a developmental theory so well-articulated that it* *would dictate, a priori, the parameterization of the within-class var/cov structure for the growth factors.*" Twisk and Hoekstra (2012) argue that the choice for one method is based on pragmatic arguments, namely LCGA is often preferred because of computation difficulties with LGMM. The latter method is more flexible, because it takes the earlier mentioned heterogeneity regarding the variation within a class into account, but a price has to be paid for this flexibility: it is more computationally demanding, often leads to convergence issues, and needs larger samples. The discussion about which parameterization to use is heavily debated in the literature, see for example the discussion in the journal *Infant Child Development* (Connell & Frye, 2006a, 2006b; Hoeksma & Kelderman, 2006; B. O. Muthén, 2006). In the current paper, we do not take a stand in this discussion. We only stress that the selection of the final model should be discussed in the paper, and ideally both models should be fitted to the data and compared. We make this recommendation because substantive results can vary depending on the model specification implemented, and it is important to examine each method in order to understand the impact on final model interpretations.

Item 6b. Are alternative specifications of the between-class differences in variance-covariance matrix structure considered and clearly documented?

In addition to the differences between LGCA and LGMM, a second issue surrounds constraining (vs. estimating freely in different classes) the error structures. This issue interacts with the across-class heterogeneity (vs. homogeneity) of growth factors' variance-covariance matrix. That is, are the residual variances and the variance-covariance matrix fixed across latent classes or are these estimated freely? There are reasons why residual variances would be left invariant across classes, and also times where they would be specified to be class specific. Fixing the residual variances across latent classes assumes that there is no difference in variability of the groups in their deviation from the growth curves. Making the residual variances different assumes that some groups (classes here) may show more variability along their growth curve than others. Class-specific residual variances might be more realistic, but since such a model contains many more parameters it might cause estimation problems. Also, residual variances may go to zero, which is typically seen if analyzing discrete data with a continuous data model. We encourage the researcher to make this decision based on the specific substantive information they have, as well as any estimation issues that arise during the analysis process

Whether the variance-covariance matrix is constrained across classes is more a substantive decision. However, when each latent class is allowed to have its own variance-covariance matrix, the model contains much more parameters to be estimated and subsequently requires larger sample sizes to avoid convergence issues. Typically, local solutions are obtained with smaller sample sizes *and* separate variance-covariance matrices being estimated. Often researchers decide to constrain the variance-covariance matrix to simplify the model (or to deal with error messages about local maxima). Whatever decision is made about the between-class variance-covariance matrix, it should be explicitly reported in the paper due to the impact the decision can have on substantive conclusions. Specifically, the researcher would want to indicate clearly which method they used and why (e.g., "theory suggests variation in growth factors is constant across subgroups, so we held this matrix fixed", etc.). Then the researcher should interpret results according to the assumption being made, but be aware that findings may be altered if the variance-covariance matrix is redefined. For example, if the covariance matrix was shifted (e.g., to be freely estimated across the classes), then the latent class solution could shift and create completely different substantive interpretations.

Item 7. Are alternative shape/functional forms of the trajectories described?

One of the main ways in which trend lines can differ is in the growth functions specified to capture change over time. Growth models based on polynomial functions are commonly implemented to assess change that is linear, quadratic, cubic, and so on (see, e.g., B. O. Muthén & Shedden, 1999). However, growth need not be defined in this manner. Many models that are nonlinear in the parameters are also commonly implemented; for example, logistic, Gompertz, and Richards growth curves (e.g., Grimm & Ram, 2009). Semi-parametric models implementing smoothing functions such as the generalized additive model can also be used to estimate growth (Zuur, Ieno, & Smith, 2007); likewise, piecewise models can be specified (e.g., Kohli, Hughes, Wang, Zopluoglu, & Davison, 2015; Palardy & Vermunt, 2010). It is not only important to report what shape each of the trajectories has in the final model, but we also advise to test this model against alternative specifications—for example, comparing a linear growth model with another model that includes a quadratic effect. See the online materials available on the Open Science Framework (go to: https://osf.io/vw3t7/) for an explanation of how specifying a different form of growth function affects the interpretation of the growth parameters.

Item 8. If covariates have been used, can analyses still be replicated?

Predictors (or covariates) can be added to the model at three different places, see Figure 1: (1) as time-varying or time-invariant covariates at the level of the dependent variables to control for variability in specific time points; (2) on the level of the growth parameters to find latent classes that cannot be explained by individual differences on the covariates (like age, food intake, SES); or (3) to predict class membership. If covariates are specified as part of the model, then this is often called a conditional model—whereas an unconditional model is one that explores the number of latent classes without consideration of covariates. Note that predictors can be directly observed or latent, regardless of where they appear in the model. When predicting class membership, there are currently several methods available, which we will describe now.

One-step method. The predictors of class membership are added into a joint model in which the class solution *and* the prediction for class membership are estimated simultaneously. There are two major disadvantages of the one-step method. First, the make-up of the latent class structure can be inappropriately modified by the inclusion of predictors. In theory, any change to the model can impact classification of individuals into the latent classes. Adding the predictor directly into the model can lead to flawed results because the covariates may affect the latent class formation, and the latent class variable may lose its meaning as the latent variable measured by the indicator variables (Asparouhov & Muthén, 2013, p. 329). This effect is nicely described by La Greca et al. (2013, p. 360). In this specific situation, the number of classes should also be reconsidered rather than adhering to the number determined without the inclusion of covariates. Whether the effect of changing class solutions is wanted or unwanted, it is not so clear whether one should decide about the number of classes in a model with or without covariates; see Palardy and Vermunt (2010) who compared models with and without covariates also in terms of the required number of classes. In conclusion, users do not want to mix up the problem of selecting the main predictors with the problem of finding the number of relevant classes.

Second, there is an index referred to as entropy that is impacted (see also Item 13). The goal of entropy is to aid in determining the accuracy of classification of individuals into the different latent classes. If entropy is near 1.0, then classification of individuals is said to be adequate. If entropy is near 0, then classification is assumed to be poor. An artifact of the one-step method for including predictors is that the entropy index is artificially over-estimated, which inappropriately inflates confidence in the classification of individuals. Moreover, the meaning of entropy itself changes. It indicates how well one can predict class membership based an individual's trajectory *and* covariate values.

Standard three-step method: Saving most likely class membership and analyze these data separately. When following this strategy, one first determines the number of latent classes without the predictors on class membership (step 1). Then the most likely class membership is saved, merged with the original data (step 2), and analyzed separately from the latent trajectory model using a multinomial regression analysis (step 3). This method was, for example, used and clearly described by Andersen, Karstoft, Bertelsen, and Madsen (2014, Suppl.Materials p.2) and by Pietrzak et al. (2014, p. 208). Although this strategy of using the most likely class membership solves various of the issues associated with the one-step approach, it ignores the uncertainty about one's class allocation. That is, it is assumed that class allocation are obtained without classification errors. What results is that the prediction based on covariates will be an under-estimation of the true effect. However, it may be possible for entropy to aid in this assessment. The higher the entropy index, the lower the classification errors, and the less bias in the prediction of class membership (Celeux & Soromenho, 1996). The strategy of saving most likely class memberships should only be employed with a high enough entropy and if authors acknowledge the attenuation effect.

Three-step approach using the pseudo-class method. A method developed by Wang, H., and Bandeen-Roche (2005b) first estimates the latent class model, and then the latent class variable is handled through multiple imputation using the posterior distribution obtained by the model. This process is followed by analyzing the imputed class variables together with the covariates using the multiple imputation technique developed by Rubin (1987), see also Asparouhov and Muthén (2007). This strategy is applied and nicely described in Peutere, Vahtera, Kivimäki, Pentti, and Virtanen (2015, p. 17). Just like the other methods described above, if the pseudo-class method is used, then it should be explicitly described, and authors should acknowledge that multiple imputation was implemented for the latent class variable.

Three-step approach with adjustment for classification errors. This method was developed by Vermunt (2010a); (2010b), expanding on ideas by Bolck, Croon, and Hagenaars (2004), see also Bakk, Tekle, and Vermunt (2013). It differs from the three-step approaches discussed above in that the analyses in the third step takes into account that the class allocations contain classification errors; that is, that these are not the true class memberships. In fact, again a latent class model is estimated, but now with the assigned class memberships from step 2 as the single indicator, with classification error probabilities fixed to their estimates from steps 1 and 2 (Asparouhov & Muthén, 2013, p. 330). This approach allows covariates to predict class membership as in a standard latent class model, but also to have distal outcomes which are predicted by class membership (Bakk & Vermunt, 2016).

The bias adjusted three-step approach has the same advantages as the simpler three-step approaches discussed above; that is, the building of a meaningful latent trajectory model for the response variable(s) of interest can be separated from the modeling of the relationship of the latent classes with external variables. However, one should be aware of the fact that also this three-step approach makes certain assumptions, among others that external variables and class indicators are conditionally independent and when the external variables are distal outcomes that the class-specific distribution of the distal outcomes is specified correctly. Note that these assumptions are also made when adopting a one-step approach, though the conditional independence assumption could then be relaxed. As far as the class-specific distribution of distal outcomes is concerned, Bakk and Vermunt (2016) showed that the. So-called, BCH-variant (Bock, Croon, & Hagenaars, 2004) is robust for violation of distributional assumption, whereas the ML (maximum likelihood) variant is not.

In conclusion, covariates can be added to the LGMM in three different places within the model (see Figure 1), and there are at least four methods that can be used to include covariates when the goal is to predict class membership, which is the most often used reason to include covariates. Since the way covariates are used and the specific method employed has a strong impact on how the result of the model can be interpreted, without stating a clear preference for one method or another, we do stress that it is of utmost importance to be completely transparent about the followed procedure.

Item 9.Is information reported about the number of random start values and final iterations included?

If maximum likelihood (ML) has been used to estimate the latent trajectory model, then it is important to know if the final class solution has been converged to the maximum of the ML distribution and not on so-called local maxima. This is because the ML-function is not always a function with only one maximum. Rather, there may be several maxima present, and finding the 'true' (i.e., absolute) maximum often depends on the starting values used for the model parameters. A solution based on a local maximum (opposed to the true maximum) can be highly different from the optimal solution. It is therefore strongly advised to re-run the model with many different starting values to ensure the optimal solution has been found. The importance of using multiple sets of starting values when estimating a mixture model has been discussed in great detail in the statistical literature. Hipp and Bauer (2006), for example, discuss at length the impact of inappropriate, or too few, sets of starting values. They show that when starting values are not selected properly, the results obtained can be substantively erroneous. They also suggest determining starting values for each parameter "judiciously" based on the substantively appropriate parameter space for each parameter being estimated. Starting values for parameters can be generated randomly but, in mixture modeling contexts, it is often advantageous to select these based on some theory. Finch and Bronk (2011) discuss selecting starting values for thresholds in the context of latent class analysis based on theory in order to avoid the estimation algorithm searching in the wrong parameter space. It is also recommended that the number of starting values be increased to at least 50-100 sets for each parameter in order to fully explore the parameter space and avoid converging to local maxima (Hipp & Bauer, 2006). When user-specified starting values are provided based on theory or previous research, then these sets of starting values represent random perturbations of the substantively relevant starting values that were specified by the user; this helps ensure that all sets cover the probable parameter space.

Item 10. Are the model comparison tools described that were used for model selection?

To determine which model fits best with the data, that is, to answer the questions about how many latent classes should be used, several statistical criteria can be used. As was investigated in a large scale simulation study by Nylund, Asparouhov, and Muthén (2007), the Bayesian Information Criterion (BIC; Schwarz, 1978) outperformed other model selection tools like the Akaike Information Criterion (AIC; Akaike, 1973) in the context of LGMMs. Both are model selection tools for assessing relative model adequacy based on the log-likelihood and the number of parameters as a penalty of model complexity. The model with the lowest BIC-value is the preferred model in terms of the number of trajectories, see the results for Model 2 in Figure 2.

Many variations to the BIC have been published and, among these, the sample size adjusted BIC is sometimes used in latent trajectory studies.

Another model selection tool often used is the Lo–Mendel–Rubin–likelihood ratio test (LMR–LRT) developed by Lo and Rubin (2001). The LMR-LRT tests the fit of $k-1$ classes against k classes, where a significant result thereby indicates that the null hypothesis of $k-1$ classes should be rejected in favor of at least k classes. However, as was indicated by Jeffries (2003, p. 901), "the result is not proven and simulation studies suggest that it may not be correct." Subsequently, Nylund et al. (2007, p. 538) replied that early simulation studies in the original Lo et al. (2002) paper showed that despite this supposed analytic inconsistency, as outlined by Jeffries, the LMR-LRT may still be a useful empirical tool for class numeration. Given the potential inconsistencies in the literature, we would advise researchers not to base the final decision on the number of classes solely on the LMR-LRT. More recently, the bootstrap likelihood ratio test (McLachlan & Peel, 2004) has in simulation studies been shown to be a good indicator for choosing the optimal number of classes (Nylund et al., 2007), but it often appears to be always significant when applied to empirical data.

Although there is discussion about which fit measure to use, there seems to be consensus among our expert panel that the BIC is the most favorite one. When the optimal dimensionality identified by model selection tools and entropy index is large, when these tools are in conflict with each other, or when they conflict with theory, applied researchers tend to reduce the number of latent trajectories to a lower number that would still be theoretically meaningful. For example, researchers typically remove trajectories that appear to account for only minor variations (e.g., Galatzer-Levy et al., 2013), or they decide to reject a model with convergence issues (e.g., Orcutt, Bonanno, Hannan, & Miron, 2014).

In sum, we urge researchers to be transparent in how they select the final model. With the current state of affairs in the literature, we would have a preference for using the BIC, but we suggest that authors include more than one comparison tool to avoid "cherry picking", see Table

2 and Figure 2 for examples of how this can be done (note that the values reported in this table are hypothetically derived for illustrative purposes). If fit indices disagree on the optimal number of classes, akin to the case in Table 2, then this finding should be acknowledged. Authors should report on all models tested and then make a case for the model they selected, preferably in combination with theory (see also Item 14). Note that there are many alternative indices proposed in the literature (e.g., Wang, H., & Bandeen-Roche, 2005a), and this field is rapidly developing (see, e.g., Klijn, Weijenberg, Lemmens, van den Brandt, & Passos, 2015), so researchers should always be aware of new developments.

Item 11. Are the total number of fitted models reported, including a 1-class solution?

The goal of trajectory-based analyses is to find the optimal number of latent classes that describe the variability in the data set. To find the optimal number of classes, we suggest a forward modeling approach starting with a one-class solution, which is the best-fitting non-mixture latent growth model. Such a model simply assumes that there are no subgroups and all individuals follow, more or less, the same trajectory over time. Often researchers do not report the 1-class solution, but it might very well be the case that a non-mixture model actually fits the data best. This is illustrated in Figure 2 where, for Model 1 without the 1-class model, one would select the 3-class solution. However, when including the 1-class model, the BIC points to this model as being optimal. As such, the conclusion should be that there are no latent classes (i.e., a single class solution is best). After fitting the 1-class model, one should incrementally add extra classes one at a time to investigate which model fits the data best. This process does not end at the moment the model fit indices stop improving. Instead, one should fit at least one or two additional models to ensure the full gamut of possible models has been examined.

Item 12. Are the number of cases per class reported for each model?

The decision on the final number of classes should not be solely based on statistical criteria. It could, for example, be that the statistically optimal solution is a solution with trajectories that contain very few subjects.[2] When two clusters (i.e., latent classes) differ drastically in size (e.g., when one cluster is much larger in size compared to another), then the larger cluster can overwhelm the smaller cluster, thus resulting in inaccurate estimates of cluster sizes and corresponding growth trajectories (Depaoli, 2013). Moreover, the model may not properly detect clusters that are small in size because there is not enough substantive information to properly identify these clusters. Instead, the trajectories might be based on outliers, or other random fluctuations, rather than substantive clusters (Bauer & Curran, 2003a; B. O. Muthén, 2003; Rindskopf, 2003). Therefore, researchers should provide information about the number of cases allocated to each of the latent classes per model, see Table 2 for an example of how this can be achieved.

Item 13. If classification of cases in a trajectory is the goal, is entropy reported?

If the goal of the analyses is to classify individuals, which is typically the case with latent trajectory studies, then it is essential to report on the performance of this classification. One tool that can be used for this purpose is the relative entropy value, with higher values indicating that individuals are classified with more confidence. That is, the solution is able to clearly classify persons in a specific class, and there is adequate separation between the classes.[3] The relative entropy is also called a measure of "fuzziness" of the derived latent classes (Jedidi, Ramaswamy, & DeSarbo, 1993; Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). The relative entropy takes on a value of 0 when all of the posterior probabilities are equal for each subject (i.e., all participants have posterior probabilities of .33 for each of three latent classes). When each participant perfectly fits in one latent class only, the relative entropy receives a maximum value of 1, which indicates that the latent classes are completely discrete partitions. Therefore, an

[2] When small latent groups are of great substantive interest, one might want to use Bayesian estimation which has shown to outperform ML estimation in LGMM/LCGA models with small sample sizes, see for example Depaoli (2013) and for an application Van de Schoot et al. (in press).

[3] Class separation refers to how different the latent classes are statistically or substantively. Class separation may be based on a variety of different trajectory characteristics, including: having very different intercepts or slopes, trajectory shapes varying across classes (linear versus nonlinear growth), the covariance structure underlying the latent growth factors may differ across classes, etc. See Depaoli (2013) for more details on latent trajectory class separation.

entropy value that is too low is cause for concern as it implies that people or cases were not well classified, or assigned to latent classes. Thus, as stated by Celeux and Soromenho (1996) the relative entropy can be regarded as a measure of the ability of the latent trajectory model to provide a relevant partition of the data; a nice explanation is provided by Greenbaum, Del Boca, Darkes, Wang, and Goldman (2005, p. 233). The relative entropy should, however, not be used to select the number of latent classes (Jedidi et al., 1993; Kaplan & Depaoli, 2011; Tein, Coxe, & Cham, 2013). As suggested by Ram and Grimm (2009), models with higher entropy are only favored when selecting among models with similar relative fit indices (e.g., BIC). Nonetheless, we advise authors to report entropy values (see for an example Table 3), or the number of misclassifications as was done by Greenbaum et al. (2005, p. 233), for each of the models.

Item 14a. Is a plot included with the estimated mean trajectories of the final solution?* And *Item 14b. Are plots included with the estimated mean trajectories for each model?

As discussed before, many researchers use substantive arguments solely—or in combination with model selection tools—to decide on the number of classes. When assessing model results for different class solutions, it is quite helpful to examine trajectory plots. A first type of graph to include represents the mean trajectories, not only for the final model, but also for each model under investigation (e.g., the models being compared during the model building/assessment phase of analysis). In the online materials (go to: https://osf.io/vw3t7/) we provide an example. It might seem a little challenging given the potential for a large number of models to be fit, but if theoretical argument are used to decide on the number of classes, then all solutions have to be presented. Note that if the journal does not allow for such a large figure then the information can be provided as online supplementary materials. We feel it is essential to provide all the information needed to replicate the decision on the final number of classes.

Item 14c. Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?

Aside from reporting the estimated mean trajectories for each model, it is also important to study the final estimated mean of the trajectories in combination with the observed individual trajectories. As argued by Erosheva et al. (2014), it allow us to visualize the extent to which individual variability is explained by latent group trajectories, as well as the extent of overlap between observations of individuals from different groups. As illustrated in Figure 3, it might be that all individuals follow the mean trajectory and maybe LCGA can be applied (Figure 3A); note that although this plot shows variation in individual trajectories, they all follow basically the same pattern of growth over time. In contrast, Figure 3B shows great variability in the individual trajectories, and the mean trajectory does not really reflect what is going on in the data. In Figure 3C none of the individual trajectories actually follow the mean trajectory and it could be questioned whether this result should be interpreted at all, even if the fit statistics are adequate. Figure 3D is even worse because there appears to be a quadratic effect, but this is completely based on missing data.

Item 15. Are characteristics of the final class solution numerically described?

Solely presenting plots of the latent trajectories obtained from the various models examined is not sufficient. It is important to also include a table of results for the final model, which would include the following for each model parameter: estimated means, standard deviations, *p*-values, confidence intervals, and the sample size used to estimate each model parameter (noting any missing data). Having access to all the information in a table helps aid in interpretation for the reader; even if numerical results are not fully described in the text, they still have access to the full model results. Including such a table also contributes to full transparency of results, making perfect replication possible since full model results have been presented in a table.

Item 16. Are the syntax files available?

There is a growing awareness that open and transparent research is essential to maintaining and improving the quality of science (Asendorpf et al., 2013; Miguel et al., 2014; Nosek et al., 2015; Wicherts, 2013; Wicherts, Bakker, & Molenaar, 2011). One of the ways in which transparency can be reached is through sharing data, syntax, and other supporting material (on the Open Science Framework (Go to: https://osf.io/vw3t7/). These are important components to any paper and provide the means for other researchers to reproduce, or alter, the reported data analysis. Moreover, it gives other researchers the opportunity to detect potential errors in the analysis, or even fraudulent results. Ensuring that syntax files are available is a first step toward fully open data and material. There are many ways in which the syntax can be made available to readers. In some cases, it is possible to include the syntax in the article via an appendix. Alternatively, syntax can be made available through online supplementary materials, or on the online data repository stores. Preferably, this information is not solely made available on personal websites since these are not permanent and may go offline one day. Furthermore, some recently created tools can aid in the availability of syntax. Examples include an online collaboration tool developed by the Center for Open Science, which allows research teams to publicize any part of their study materials as a way to support open communication. The Center for Open Science has also created certain "badges" to certify papers that adhere to open materials requirements. These developments mean that there is truly no reason to keep your syntax files from your audience. Our files are available on the Open Science Framework (Go to: https://osf.io/vw3t7/).

Application of GRoLTS in a systematic review on latent trajectory studies

To evaluate the consistency, validity, and usability of GRoLTS, we pilot-tested the questionnaire on 38 studies that all applied latent trajectory analyses (i.e., LGMM or LCGA) to assess change in posttraumatic stress symptoms (PTSS) after a traumatic event. Completing GRoLTS took an average of 20 minutes per paper. Two independent assessors administered GRoLTS to each of the 38 papers included. When scores were conflicting, consensus about the score was easily obtained after shortly discussing the assessors' rationales behind their respective scores. The complete list of references and all of the details surrounding these papers can be found on the Open Science Framework (Go to: https://osf.io/vw3t7/).

Figure 4 displays the sum scores for the GRoLTS items across all papers, but not one single paper came close to the maximum score, which is 21 (M=9.47; SD=1.97; Range=5-15). Upon examining the specific GRoLTS items that were reported, see Figure 5, we found that some items were almost always reported while others were hardly ever reported. We highlight the top six most/least frequently reported items.

Frequently reported items (top 6)

1. All papers provided a plot with the estimated mean trajectories of the final solution (item 14a).
2. The model selection tools used were reported (item 10) in almost all papers (97%) and always included the BIC. Other model selection tools were mentioned in roughly two-thirds of the papers: SS-BIC=60.5%; AIC=63,2%; LRT=65.8%; BLRT=60.5%. It was mentioned in 12 papers that the model fit indices disagreed, in nine papers the AIC/BIC kept decreasing, in 10 papers the preferred model based on the statistical criteria did not make sense, and in five papers the best model contained a class with only a couple of individuals allocated to one of the subgroups. Of the papers where model selection tools did not provide an easy solution, 13 papers cited theory instead of statistics to choose between models.
3. Software was reported in 95% of the papers (item 5)—Mplus being the most popular (29 papers), followed by SAS Proc Traj (7 papers); and 30 papers (79%) reported which version was used.
4. Entropy level was reported in 95% of the papers (item 13) with a median entropy value reported across all of these papers of .85.

5. The fifth most often reported item was item 3c (how missing data was dealt with; reported in 89.5%). The most popular method for handling missing data was FIML (24 papers), but only one study combined this approach with auxiliary variables. Three studies indicated to have used multiple imputation for handling missing data.
6. A clear description on the use of covariates was reported in 86% of the papers. The one-step method was applied in 15 papers, while the standard three-step method (i.e., saving the most likely class membership and analyzing data separately) was used in 14 papers. The recently proposed biased adjusted three-step method of Vermunt was only applied in three papers.

Infrequently reported items

1. As described in GRoLTS, we recommend not only including a graph representing the mean trajectories for the final model (item 14a), but also for each model under investigation (item 14b). In the PTSS papers, none of the papers presented the latter type of plots. Reporting the final estimated mean of the trajectories in combination with the observed individual trajectories (item 14c) was reported in just 5 papers (13%). The failure to report these plots (14b&c) is not limited to the field of PTSS literature—the systematic reviews of Piquero (2008) and Erosheva et al. (2014) found zero out of 87 and eight out of 200 (4%) trajectory studies reporting such graphs.
2. Another item that was not described at all is the between-class variance-covariance matrix structure (item 6b). Although this is considered a very important topic from a statistical perspective, apparently researchers take the between-class variance-covariance matrix structure for granted, and probably stay with the default setting of the software used.
3. Only one study reported on the exact number of starting sets used (item 9).
4. Ensuring that syntax files are available (item 16) is a first step toward fully open data and material. This information was only provided in two of the papers we examined.

5. The missing data mechanism (item 3a) was only reported in two papers; they concluded MAR.
6. Although 74% of the papers reported on the metric of time, only three reported on the variability of time in detail (item 2).

Discussion and Conclusion

We have developed GRoLTS, a tool for reporting on latent trajectory studies (LGMM or LCGA). We have gone through a systematic process to identify key components that, according to a panel of expert statisticians and senior users, are necessary when reporting results for latent trajectory studies. Reporting standards are important for any statistical model implemented in the applied literature. Other reporting standards, such as the CONSORT checklist for the reporting on randomized controlled trials have successfully been implemented. A systematic review has shown that the use of reporting standards such as CONSORT do improve the quality of reporting (Plint, Moher, Morrison, & Schulz, 2006). We expect that especially for the field of latent growth trajectory models, reporting standards are an irrefutable component to abide by when presenting model results. Substantive interpretations rely heavily on a variety of components embedded within the specification and estimation of these models.

We recommend that GRoLTS should be used in any application of latent growth trajectory modeling to ensure proper dissemination of results. Note that GRoLTS- does not aim to measure the quality of the paper itself, but rather the quality of reporting on key issues of latent trajectory models. GRoLTS has been designed to be thorough, yet easy (and concise) to implement. Although GRoLTS is relatively detailed, many of these components can be adequately handled in just a few extra sentences added to the text of the paper, or with the use of online supplementary material. GRoLTS may be used by authors who prepare their manuscript for submission, and may be endorsed by journals as a standard for reporting on LGMM or LCGA studies. Naturally, GRoLTS should be regularly updated and revised because it is a rapidly evolving method that is used heavily across different fields. New advances may necessitate the addition or removal of GRoLTS items. We recognize that there is a great deal of variability in field-standards and types of research questions addressed using trajectory-based methods. As a result, it is important to consider whether additional points not addressed on GRoLTS need to be considered when reporting trajectory results.

We would like to end our paper with a quote by Daniel Bauer (2007, p. 782):

> *"The fundamental question I sought to address is whether these models* [he refers to LGMM/LCGA models] *are likely to advance psychological science. My firm conviction is that, if these models continue to be applied as they have been so far, the answer is clearly no. [...] I therefore believe that direct applications of GMMs should be refrained from unless both the theory and data behind the analysis are uncommonly strong. Otherwise, the application of GMMs in psychological research is likely to lead to more blind alleys than ways forward".*

We agree with Bauer's quote and feel that the way these models have been reported on in the past has not been as transparent and consistent as would be needed to produce trustworthy and replicable findings. If researchers across all fields have a strong theoretical basis *and* standardize their reporting habits using the GRoLTS, then we believe latent growth trajectory modeling can take the next step and become one of the most transparent and replicable areas of applied statistics.

References

Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267): Budapest: Akademiai Kiado.

Andersen, S. B., Karstoft, K. I., Bertelsen, M., & Madsen, T. (2014). Latent trajectories of trauma symptoms and resilience: the 3-year longitudinal prospective USPER study of Danish veterans deployed in Afghanistan. *J Clin Psychiatry, 75*(9), 1001-1008. doi:10.4088/JCP.13m08914

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Nosek, B. A. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108-119.

Asendorpf, J. B., Van De Schoot, R., Denissen, J. J., & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies The benefits of multiple imputation. *International Journal of Behavioral Development*, 0165025414542713.

Asparouhov, T., & Muthén, B. (2007). Wald test of mean equality for potential latent class predictors in mixture modeling. *Retrieved November, 9*, 2009.

Asparouhov, T., & Muthén, B. (2013). Auxiliary variables in mixture modeling: A 3-step approach using Mplus (Mplus web notes: No. 15, Version 6).

Aydin, B., Leite, W. L., & Algina, J. (2014). The consequences of ignoring variability in measurement occasions within data collection waves in latent growth models. *Multivariate Behavioral Research, 49*(2), 149-160.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*(1), 272-311.

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 20-31.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*(4), 757-786.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods, 8*, 338–363.

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of Latent Trajectory Classes: Much Ado About Nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003).

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods, 9*(1), 3.

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.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Bates, T. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3-27.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*, 195–212.

Connell, A. M., & Frye, A. A. (2006a). Growth mixture modelling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. *Infant and Child Development, 15*(6), 609-621.

Connell, A. M., & Frye, A. A. (2006b). Response to commentaries on target paper,'Growth Mixture Modeling in Developmental Psychology'. *Infant and Child Development, 15*(6), 639-642.

Coulombe, P., Selig, J. P., & Delaney, H. D. (2015). Ignoring individual differences in times of assessment in growth curve modeling. *International Journal of Behavioral Development*, 0165025415577684.

Croudace, T. J., Jarvelin, M.-R., Wadsworth, M. E., & Jones, P. B. (2003). Developmental typology of trajectories to nighttime bladder control: epidemiologic application of longitudinal latent class analysis. *American Journal of Epidemiology, 157*(9), 834-842.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus bayesian estimation. *Psychological Methods, 18*(2), 186-219. doi:10.1037/a0031609

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application*: Routledge Academic.

Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology, 20*(1), 1-26.

Erosheva, E. A., Matsueda, R. L., & Telesca, D. (2014). Breaking bad: two decades of life-course data analysis in criminology, developmental psychology, and beyond.

Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology, 45*(3), 652.

Finch, W. H., & Bronk, K. C. (2011). Conducting Confirmatory Latent Class Analysis Using M plus. *Structural Equation Modeling, 18*(1), 132-151.

Galatzer-Levy, I. R., Ankri, Y., Freedman, S., Israeli-Shalev, Y., Roitman, P., Gilad, M., & Shalev, A. Y. (2013). Early PTSD Symptom Trajectories: Persistence, Recovery, and Response to Treatment: Results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *PLoS ONE, 8*(8).

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology, 73*(2), 229.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in M plus and SAS. *Structural Equation Modeling, 16*(4), 676-701.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: an overview of methodological issues. *Psychology and aging, 18*(4), 639.

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

Hoeksma, J. B., & Kelderman, H. (2006). On growth curves and mixture models. *Infant and Child Development, 15*(6), 627-634.

Jedidi, K., Ramaswamy, V., & DeSarbo, W. S. (1993). A maximum likelihood method for latent class regression involving a censored dependent variable. *Psychometrika, 58*(3), 375-394.

Jeffries, N. O. (2003). A note on 'Testing the number of components in a normal mixture'. *Biometrika, 90*(4), 991-994.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research, 29*(3), 374-393.

Jung, T., & Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kaplan, D., & Depaoli, S. (2011). Two studies of specification error in models for categorical latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(3), 397-418.

Klijn, S. L., Weijenberg, M. P., Lemmens, P., van den Brandt, P. A., & Passos, V. L. (2015). Introducing the fit-criteria assessment plot–A visualisation tool to assist class enumeration in group-based trajectory modelling. *Statistical methods in medical research*, 0962280215598665.

Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear-linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods, 20*(2), 259-275. doi:10.1037/met0000034

Kreuter, F., & Muthén, B. O. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology, 24*(1), 1-31.

La Greca, A. M., Lai, B. S., Llabre, M. M., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (2013). Children's Postdisaster Trajectories of PTS Symptoms: Predicting Chronic Distress. *Child and Youth Care Forum*, 1-19.

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

Lommen, M. J. J., Van De Schoot, R., & Engelhard, I. M. (2014). The experience of traumatic events disrupts the stability of a posttraumatic stress scale. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01304

McLachlan, G., & Peel, D. (2004). *Finite mixture models*: John Wiley & Sons.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*(1), 23.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., . . . Imbens, G. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30-31.

Muthen, B. (2001). Latent variable mixture modeling. *New developments and techniques in structural equation modeling*, 1-33.

Muthén, B. O. (2003). Statistical and Substantive Checking in Growth Mixture Modeling. *Psychological Methods, 8*, 369-377.

Muthén, B. O. (2006). The potential of growth mixture modelling. *Infant and Child Development, 15*(6), 623-625.

Muthén, B. O., & Muthén, L. K. (2000a). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

Muthén, B. O., & Muthén, L. K. (2000b). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891.

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

Muthén, L. K., & Muthén, B. O. (2013). *Mplus: Statistical analysis with latent variables: User's guide*: Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: a semi-parametric group based approach. *Psychological Methods, 4*, 139-157.

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

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model*. *Criminology, 31*(3), 327-362.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods, 6*(1), 18.

Nosek, B., Alter, G., Banks, G., Borsboom, D., Bowman, S., Breckler, S., . . . Christensen, G. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science (New York, NY), 348*(6242), 1422.

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

Orcutt, H. K., Bonanno, G. A., Hannan, S. M., & Miron, L. R. (2014). Prospective trajectories of posttraumatic stress in college women following a campus mass shooting. *Journal of Traumatic Stress, 27*(3), 249-256.

Palardy, G. J., & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics, 35*(5), 532-565.

Peeters, M., Zondervan-Zwijnenburg, M., Vink, G., & van de Schoot, R. (2015). How to handle missing data: A comparison of different approaches. *European Journal of Developmental Psychology, 12*(4), 377-394.

Peutere, L., Vahtera, J., Kivimäki, M., Pentti, J., & Virtanen, P. (2015). Job Contract at Birth of the First Child as a Predictor of Women's Labor Market Attachment: Trajectory Analyses over 11 Years. *Nordic Journal of Working Life Studies, 5*(1), 9.

Pietrzak, R. H., Feder, A., Singh, R., Schechter, C. B., Bromet, E. J., Katz, C. L., . . . Southwick, S. M. (2014). Trajectories of PTSD risk and resilience in World Trade Center responders: An 8-year prospective cohort study. *Psychological Medicine, 44*(1), 205-219.

Piquero, A. R. (2008). Taking stock of developmental trajectories of criminal activity over the life course *The long view of crime: A synthesis of longitudinal research* (pp. 23-78): Springer.

Plint, A. C., Moher, D., Morrison, A., & Schulz, K. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia, 185*(5), 263.

Proust-Lima, C., Philipps, V., & Liquet, B. (2015). Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *arXiv preprint arXiv:1503.00890*.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM manual.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565-576.

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage.

Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003).

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the american statistical association, 82*(398), 543-546.

Rubin, D. B., & Little, R. J. (2002). Statistical analysis with missing data. *Hoboken, NJ: J Wiley & Sons*.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177. doi:10.1037//1082-989X.7.2.147

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-465.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*: Oxford university press.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: a practical guide to their development and use*: Oxford university press.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 640-657.

Twisk, J., & Hoekstra, T. (2012). Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *Journal of clinical epidemiology, 65*(10), 1078-1087. doi:http://dx.doi.org/10.1016/j.jclinepi.2012.04.010

Van Buuren, S., & oudshoorn, C. G. M. (2005). *Multivariate Imputation by Chained Equations* (PG/VGZ/00.038). Retrieved from Leiden:

Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial "Measurement Invariance". *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01064

Vermunt, J. K. (2010a). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis, 18*(4), 450-469.

Vermunt, J. K. (2010b). Longitudinal research using mixture models *Longitudinal research with latent variables* (pp. 119-152): Springer.

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).

Vermunt, J. K., & Magidson, J. (2013). Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax. *Statistical Innovations Inc., Belmont*.

Wang, C.-P., H., B. C., & Bandeen-Roche, K. (2005a). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the american statistical association, 100*(471), 1054-1076.

Wang, C.-P., H., B. C., & Bandeen-Roche, K. (2005b). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the american statistical association, 100*, 1054-1076.

Wicherts, J. M. (2013). Science revolves around the data. *Journal of Open Psychology Data, 1*(1).

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE, 6*(11), e26828.

Zuur, A., Ieno, E. N., & Smith, G. M. (2007). *Analysing ecological data*: Springer Science & Business Media.

Table 1. Final list of items of the GRoLTS -checklist: Guidelines for Reporting on Latent Trajectory Studies.

	Checklist item	Reported?
1.	Is the metric of time used in the statistical model reported?	Yes / No
2.	Is information presented about the mean and variance of time within a wave?	Yes / No
3a.	Is the missing data mechanism reported?	Yes / No
3b.	Is a description provided of what variables are related to attrition/missing data?	Yes / No
3c.	Is a description provided of how missing data in the analyses was dealt with?	Yes / No
4.	Is information about the distribution of the observed variables included?	Yes / No
5.	Is the software mentioned	Yes / No
6a.	Are alternative specifications of within-class heterogeneity considered (e.g. LGCA versus LGMM) and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?	Yes / No
6b.	Are alternative specifications of the between-class differences in variance-covariance matrix structure considered and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?	Yes / No
7.	Are alternative shape/functional forms of the trajectories described?	Yes / No
8.	If covariates have been used, can analyses still be replicated?	Yes / No
9.	Is information reported about the number of random start values and final iterations included?	Yes / No
10.	Are the model comparison (and selection) tools described from a statistical perspective?	Yes / No
11.	Are the total number of fitted models reported, including a 1-class solution?	Yes / No
12.	Are the number of cases per class reported for each model (absolute sample size, or proportion)?	Yes / No
13.	If classification of cases in a trajectory is the goal, is entropy reported?	Yes / No
14a.	Is a plot included with the estimated mean trajectories of the final solution?	Yes / No
14b.	Are plots included with the estimated mean trajectories for each model?	Yes / No
14c.	Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?	Yes / No
15.	Are characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc)?	Yes / No
16.	Are the syntax files available (either in the appendix, supplementary materials, or from the authors)?	Yes / No

Table 2. Example table with hypothetical results. Note that the values reported in this table are hypothetically derived for illustrative purposes

Nr. Of classes	**AIC**	**BIC**	**BLRT**	**VLMR**	**Entropy**	**Sample size per class based on most likely class membership**
1	8459	8512	-	-	-	220
2	8350	8532	p <.001	p <.001	.601	119/21
3 [a]	8247	8598	p <.001	p <.001	.702	106/94/20
4 [a b]	8143	8456	p <.001	$p = .06$	.804	101/90/22/7
4 [a c]	8140	8318	p <.001	$p = .18$	.901	100/80/35/5
5 [a]	8420	8548	p <.001	$p = .29$	.909	55/45/81/34/5
6 [a d]	-	-	-	-	-	-

[a] number of random starts increased to 1000 with 40 optimization phases due to convergence issues.
[b] Non-positive definite matrix because of negative slope variance
[c] variance slope fixed to zero
[d] no convergence

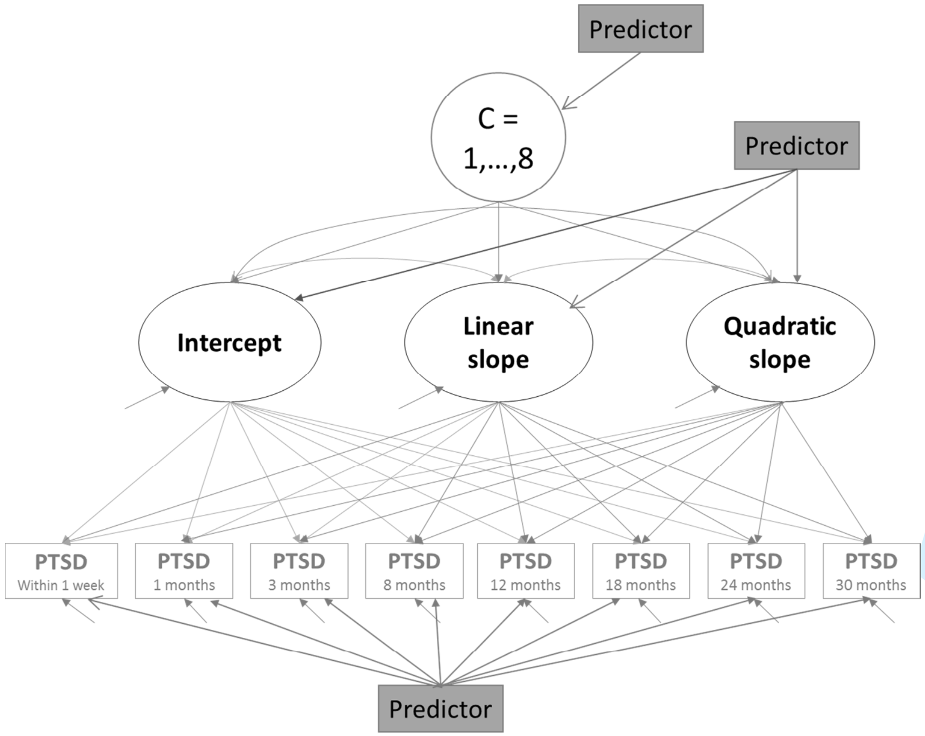

Figure 1. An example of a latent trajectory model with 1 to 8 classes (C=1,…,8), eight overserved variables (post-traumatic stress disorder), three growth parameters (intercept, slope quadratic term) and three places where covariates can be added to the model.

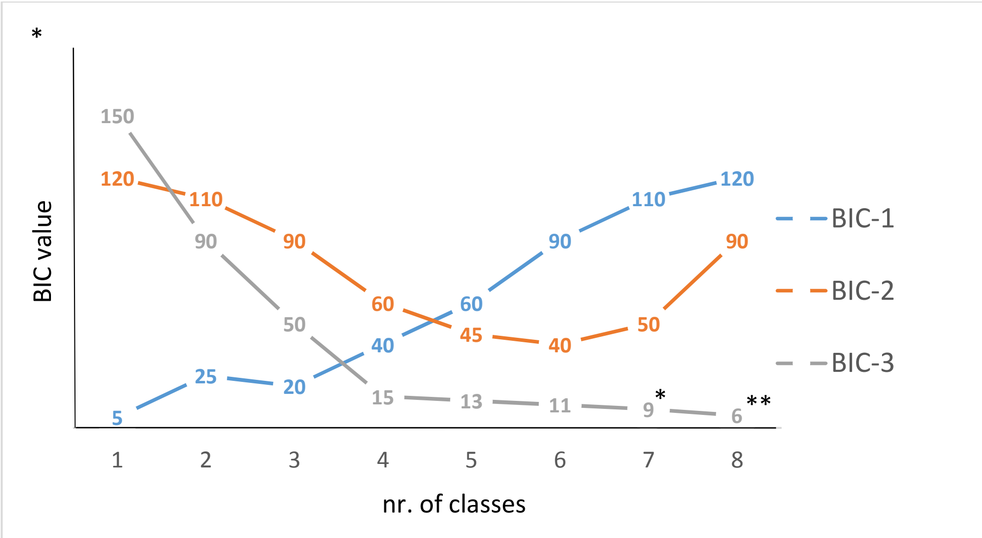

Figure 2. BIC values for three hypothetical model results (BIC-1, BIC-2, BIC-3). Note that the model with one asterisk indicates the number of random starts was increased to 1,000 and the model with two asterisks did not reach convergence.

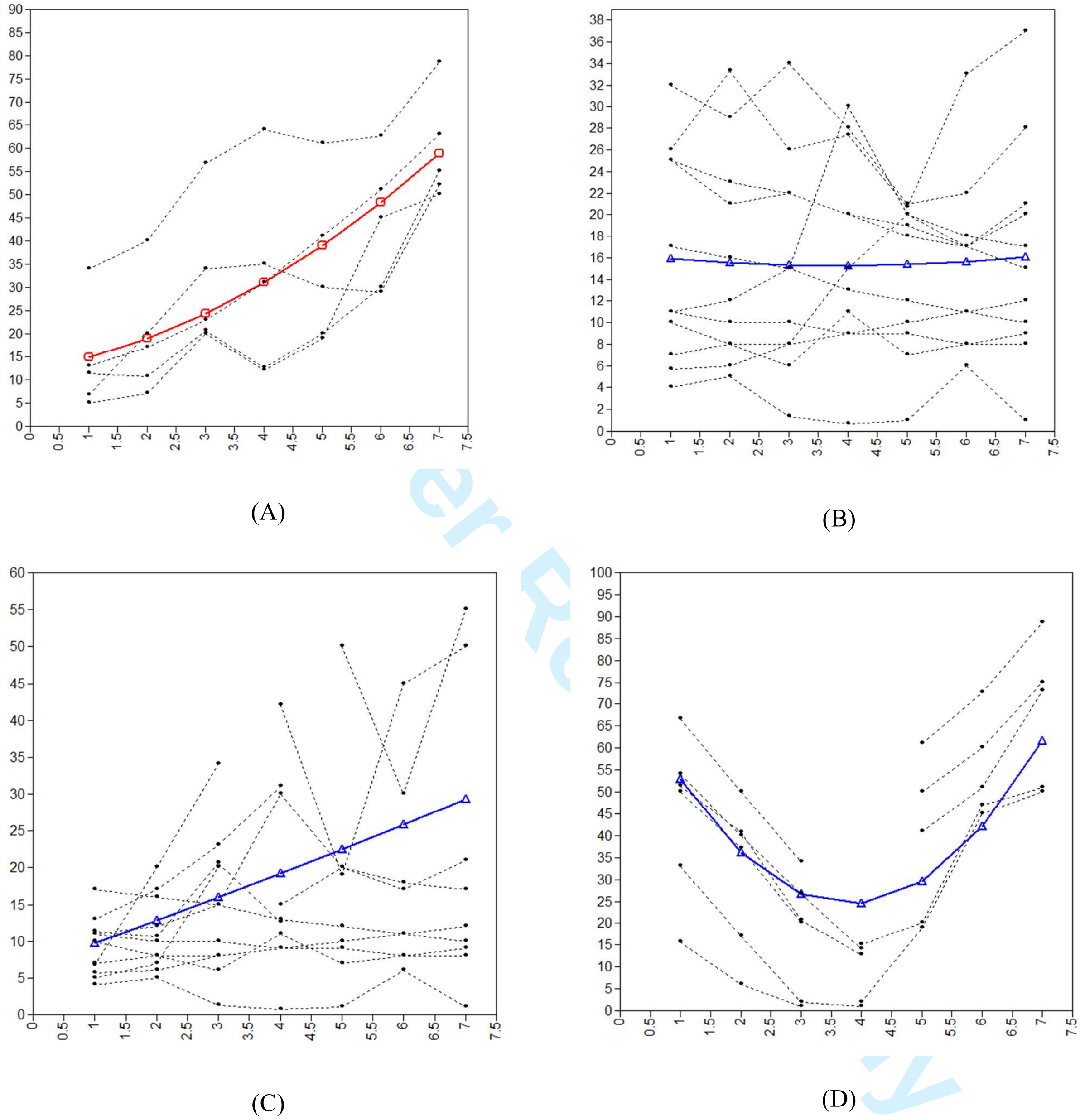

Figure 3. Plots of estimated means with individual trajectories based on hypothetical data (but inspired by empirical results).

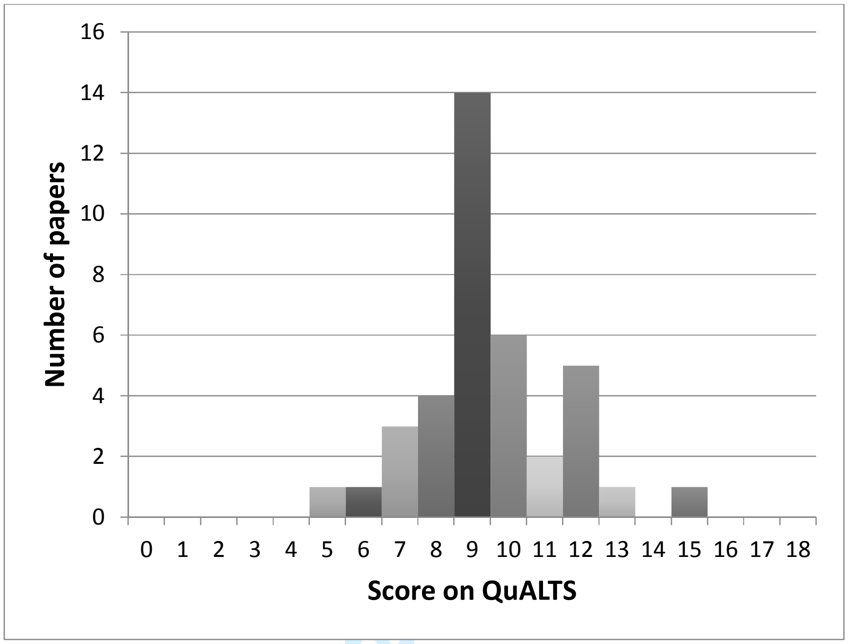

Figure 4. Total GRoLTS score plotted for 38 studies applying latent trajectory analyses examining development of posttraumatic stress symptoms after a traumatic event.

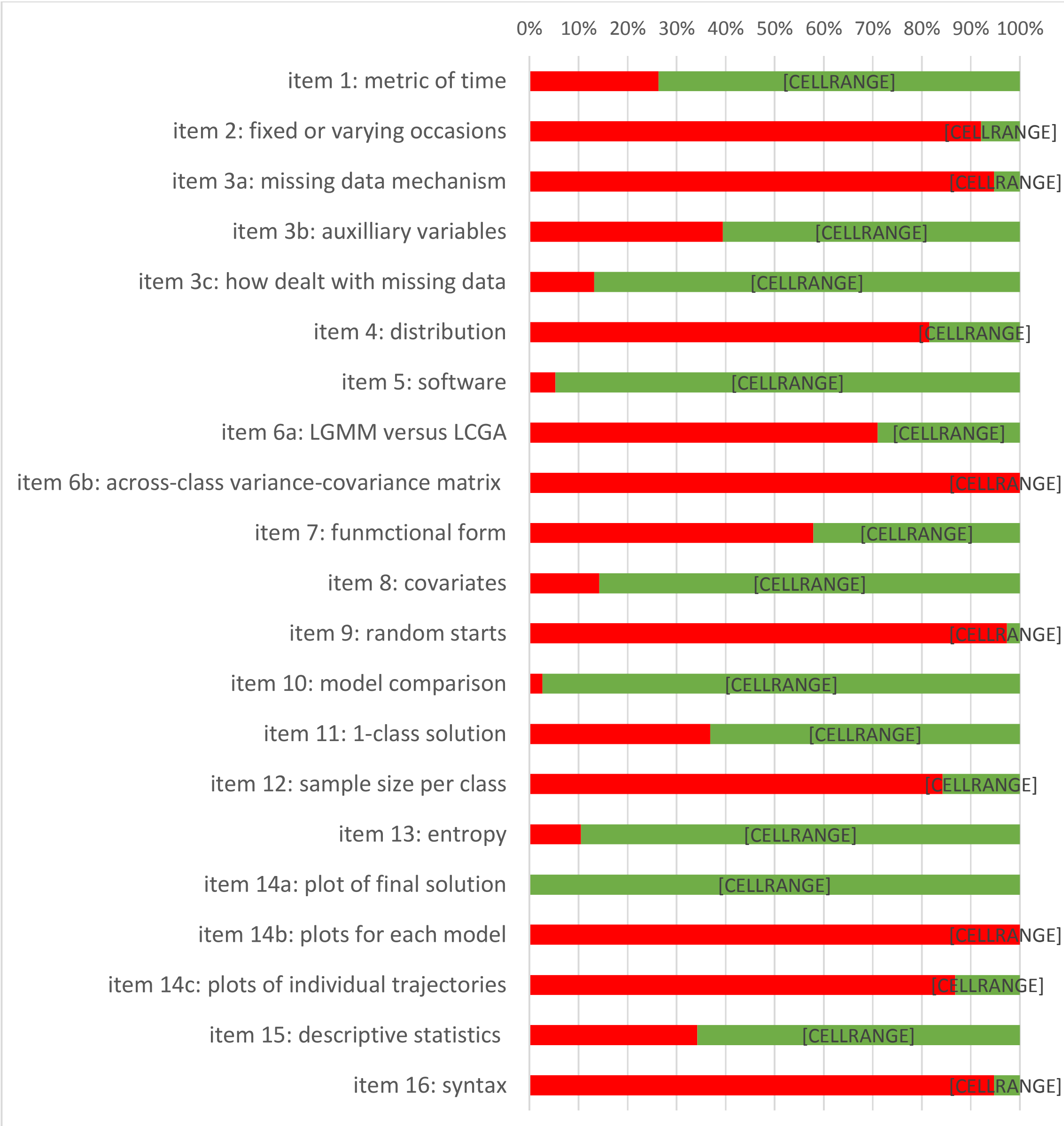

Figure 5. Papers (%) fulfilling individual GRoLTS items. Note: total number of papers examined for systematic review is N=38, exact percentages given in green bar.