SEM differs substantively from traditional generalized linear modeling applications in that model-data discrepancies (i.e., residuals) are not captured by evaluating differences between observed and predicted values of the response variables (, but rather differences between elements of the observed and model-implied covariance matrixes (**S-**). The traditional methods of evaluating model adequacy also differ. Generalized linear modeling textbooks emphasize residual diagnostic procedures to evaluate the tenability of model assumptions (e.g., normally distributed errors, homoskedasticity, etc.) and evaluation of model predictive performance using plots of predicted vs. observed values in addition to the overall *F* statistic and R2 metrics (Fox, 2008). Whereas there is greater emphasis in SEM than standard regression applications in using visualizations (e.g., path diagrams or directed acyclic graphs) to communicate model specification, there is less emphasis on visualization post-estimation, during the model evaluation stage.

SEM practitioners generally rely on global fit tests and approximate fit indices rather than visual inspection. Amidst the contentious debate over the best practices for evaluating the adequacy of SEMs, there appears to be a consensus that no one test or metric is sufficient in determining whether a model represents the data-generating mechanisms adequately. Model evaluation requires more than scanning statistics summarizing the extent to which one’s model is consistent with the data or better than a competing model. A significant χ2 test alerts modelers that the overall model-data discrepancy is greater than chance expectation using an *a priori* cutoff value but does not help identify the faulty component(s) of the model. Modification indices can help researchers identify post-hoc model constraints that, if freed, would lessen model-data discrepancies; however, the fact that an estimated parameter improves fit does not mean that it provides a better representation of the data-generating mechanisms.

A number of scholars have emphasized the importance of supplementing global tests and indices of model adequacy with local fit evaluation procedures (Asparouhov & Muthén, 2017; Goodboy & Kline, 2017; Thoemmes et al., 2018; Tomarken & Waller, 2005).

**Visualization Strategies to Evaluate SEMs**

Poor model fit in SEM can be the result of misspecification of the causal structure and/or violation of model assumptions. A number of scholars have used visual strategies to evaluate the tenability of model assumptions, diagnose causal misspecifications, and select the best model from a group of competitors. Importantly, evidence of violations of modeling assumptions (e.g., non-normally distributed case residuals) could be a signal of causal misspecification.

**Residual covariances/correlations.** A common approach to evaluating local model fit is to plot the distribution of the residual covariances/correlations (e.g., using stem-and-leaf plots or histograms) to aid in identifying specific model implications that are most inconsistent with the observed data (Bollen, 1989; Bollen & Arminger, 1991). Residual covariances may signal problems that are otherwise masked in global fit indices. Distributional plots of residual covariances are particularly helpful for identifying aberrant residual covariances in complex models in which the residual covariance matrix has many elements.

**Individual case residuals.** Global fit indices and residual covariances are important for detecting model misspecification, but because they are based on aggregate statistics, they do not identify aberrant cases that may influence model parameter estimates and fit. Bollen and Arminger (1991) demonstrated methods of calculating raw and standardized individual case residuals (ICRs) representing differences between observed and model-predicted case values for outcome variables. They used stem-and-leaf, index, and histogram plots to help locate outlying and influential observations. This technique requires estimation of factor scores (e.g., using regression-based or Bartlett methods). Pek and MacCallum (2011) demonstrated how case diagnostic procedures commonly used in GLMs (e.g., Mahalanobis distances, generalized Cook’s *D*, and *DFBETAs*) can be applied to SEMs to detect influential cases. Similarly, Flora and colleagues applied these diagnostic procedures and others specifically to factor analysis models (Flora et al., 2012). Visualization procedures showing case influence on model fit (e.g., likelihood differences) and parameter estimations (e.g., generalized Cook’s *D*) have been implemented in open-source R packages, including fa.outlier (Chalmers & Flora, 2015) and influence.SEM (Pastore & Altoe, 2018). These packages use both statistical tests and graphs to aid in identifying cases with strong influence on model fit and parameter estimates.

Muthén and Asporouhov (2017) demonstrated how ICRs can be used to detect specific model misspecifications. First, they showed that plots of estimated factor scores for a latent outcome variable against observed predictor variables can be used to detect unspecified nonlinear effects of the predictor on the latent outcome. Second, they used residual scatterplots to detect violations of local independence in a latent factor model. When two reflective indicators (Y1 and Y2) of a latent factor (η) had an unmodeled common cause, they showed that plotting the indictors’ residuals against one another could help identify the non-zero residual covariance, prompting modelers to consider an alternate model. Finally, they demonstrated in a latent factor model how plotting predicted values for a reflective indicator () against the observed indicator values (*Y*) could uncover unmodeled heterogeneity that could be better captured using a mixture model.

Raykov and Penev (2014) demonstrated how ICRs can be used to aid in model selection in the context of latent growth curve modeling. When comparing linear and quadratic growth curve models for the same data, for example, they showed that a scatterplot of the ICRs for the quadratic model vs. ICRs for a linear model can help the modeler see which model best minimizes ICRs, leading to selection of the appropriate model.

demonstrated visual diagnostic methods—reminiscent of those used in linear regression models—in detecting certain types of model misspecifications. First, A benefit of these visual diagnostic strategies relative to commonly used tests of model fit and approximate fit indices is that the plots may provide guidance in determining the nature of the model misspecifications and how they might be remediated (Asparouhov & Muthén, 2017).

There are notable limitations, however, to the Asporouhov and Muthén (2017) approach. First, when using factor score estimates for model diagnostic purposes, the result may differ depending on the method of factor score estimation owing to factor indeterminacy (Grice, 2001; Rigdon et al., 2019; Steiger, 1996). At high levels of indeterminacy, two sets of factor score estimates could differ markedly—and even be negatively correlated (Grice, 2001). Thus, estimating the degree of factor indeterminacy would be prudent prior to relying on the factor score estimates and potentially evaluating whether similar visual patterns hold across multiple methods of calculating factor score estimates. Second, as the authors note in their article, in their method of obtaining predicted values for an outcome variable (), the latent factor scores () contributes to the prediction even though already contains information about *Y*. However, this dependence between *Y* and becomes less problematic when the latent factor is measured with increasing numbers of highly reliable indicators (Asparouhov & Muthén, 2017).