Heywood Cases in Unidimensional Factor Models and Item Response Models for Binary Data

Selena Wang¹, Paul De Boeck², and Marcel Yotebieng³

Abstract
Heywood cases are known from linear factor analysis literature as variables with communalities larger than 1.00, and in present day factor models, the problem also shows in negative residual variances. For binary data, factor models for ordinal data can be applied with either delta parameterization or theta parametrization. The former is more common than the latter and can yield Heywood cases when limited information estimation is used. The same problem shows up as non convergence cases in theta parameterized factor models and as extremely large discriminations in item response theory (IRT) models. In this study, we explain why the same problem appears in different forms depending on the method of analysis. We first discuss this issue using equations and then illustrate our conclusions using a small simulation study, where all three methods, delta and theta parameterized ordinal factor models (with estimation based on polychoric correlations and thresholds) and an IRT model (with full information estimation), are used to analyze the same datasets. The results generalize across WLS, WLSMV, and ULS estimators for the factor models for ordinal data. Finally, we analyze real data with the same three approaches. The results of the simulation study and the analysis of real data confirm the theoretical conclusions.

Keywords
Heywood cases, item response model, factor model

Introduction
Heywood cases are variables with a proportion explained variance (i.e., the communality) larger than 1.00 in factor analytic models, an anomaly for evident reasons. Harman and Fukuda (1966)

¹Yale University, New Haven, CT, USA
²The Ohio State University, Columbus, OH, USA
³Albert Einstein College of Medicine, Bronx, NY, USA

Corresponding Author:
Selena Wang, Department of Biostatistics, Yale University, 1827 Neil Ave, New Haven, CT 06513, USA.
Email: selena.wang@yale.edu
named this problem after Heywood (Heywood, 1931), its first describer, and since then “Heywood cases” has become a common term in applied and methodological research. In factor models of today, with covariance matrix-based estimation, Heywood cases are identified by the appearance of negative residual variances. Heywood cases are far from exceptional (Chen et al., 2001; Jöreskog & Lawley, 1968; Martin & McDonald, 1975), and when they occur, the corresponding factor solutions are considered improper solutions (e.g., Jöreskog & Lawley, 1968). While Heywood cases are not uncommon for factor models, linear factor models or factor models for ordinal data, they do not occur in item response (IRT) models. In addition to factor models, IRT models constitute a prominent family of latent variable models for categorical variables, and the two are known to be connected (Kamata & Bauer, 2008; Takane & de Leeuw, 1987). It is therefore natural to wonder why factor model estimation runs into Heywood cases, but IRT model estimation does not. As will be explained, the issue that leads to Heywood cases in linear factor models shows in a different way when models for ordinal and binary data are analyzed, depending on the factor model parameterization and on whether an IRT model is used. In practice, we will focus on binary data, the simplest case of categorical data that can be treated as ordinal. To explain how a Heywood case shows in different ways depending on the type of analysis, we first briefly explain factor models and IRT models for ordinal (and binary) data.

For ordered-category variables, including binary variables, IRT models, as well as factor models can be and are being used. The two-parameter logistic (2PL) unidimensional and multidimensional IRT models are common models for binary data. The probit version of these IRT models can be estimated using factor models for ordinal data based on polychoric correlations (i.e., tetrachoric correlations for binary variables) and thresholds. In the following, factor models for ordinal data will be labeled as “ordinal factor models”, which is a contraction term.

There are roughly speaking three major differences between factor models and IRT models: (1) most factor models are linear models for continuous and normally distributed variables. The linear form of the model is a defining feature, and the normal distribution is a common but not necessary assumption (e.g., not necessary for least square methods). If the distributional assumption is correct, then the variance-covariance matrix is a sufficient statistic, and the chi-square statistic and goodness-of-fit indices based on it can be interpreted without much problem. For ordered-category variables, including binary variables, one can rely on ordinal factor models. IRT models are always for categorical variables. Evidently, neither ordinal factor models nor IRT are linear models for the observed variables, and no distributional assumptions are made for these variables. See the next point for the assumptions that are made instead. (2) For ordinal factor models, an underlying and normally distributed continuous item response variable is postulated, with thresholds to obtain ordinal response categories. This is in correspondence with the assumption of polychoric correlations, called tetrachoric correlations if the response categories are binary. These correlations are then factor-analyzed. There are two parameterizations available: the delta parameterization and the theta parameterization (Asparouhov & Muthen, 2020; Paek et al., 2018), to be explained in the next Section. IRT models are based on a probit link or a logit link for the expected response value of binary choices between pairs of subsets of the ordinal response categories. The expected values of the binary outcomes (i.e., probabilities) are expressed as a function of a linear combination of the latent variable(s) and item parameters. The function in question to obtain the expected value from the linear combination is the inverse probit or logit link. If there are more than two response categories, different approaches are available to define the pairs of subsets of response categories. Partial credit modeling and graded response modeling are the most popular approaches, and sequential modeling is a less commonly used approach. Each of these approaches defines binary choices between subsets of response categories. For example, for three ordered response categories, the partial credit model makes use of adjacent categories (1 vs 2, and 2 vs 3), the graded
response model makes use of cumulative categories (1 vs 2&3, 1&2 vs 3), and the sequential model makes use of continuation categories (1 vs 2&3, 2 vs 3).

In order to understand how Heywood cases appear differently in factor models and IRT models, we provide a brief discussion about the relationship between IRT and ordinal factor models, first described by Takane and de Leeuw (1987). Note that it is not the purpose of this paper to connect IRT models with factor models as such; rather, the connection is a necessary element to our goal of understanding Heywood cases in unidimensional factor models and item response models. The availability of three different IRT modeling types for data with three or more ordered categories is a complication compared with ordinal factor models. Therefore, we will focus on binary data, for which the distinction is non-existent. (3) Estimation of factor models is commonly based on the variance-covariance matrix. More specifically for ordinal factor models, it is based on the polychoric correlation matrix (tetrachoric correlations for binary data) and thresholds. For linear factor models, the variance-covariance matrix is a sufficient statistic (when normality applies); for ordinal factor models, the polychoric correlation matrix does not entail all information in the data, and therefore the estimation is called a limited information approach. Although a full information approach with maximum likelihood is available for ordinal factor models (Bock et al., 1988), it is not commonly used. Meanwhile, for IRT models, estimation is usually based on the full data, instead of pairwise relationships between variables. For an overview and discussion of these estimation methods, see Bolt (2005), among others.

We recognize that the distinction between IRT models and factor models is not as sharp as it has been, for example, with newer estimation approaches such as full information maximum likelihood and Bayesian approaches, the distinction is more blurred. Our main motivation for this study on Heywood cases is to improve the understanding of certain issues one may encounter when estimating ordinal factor models and IRT models using popular software. A better understanding of the issues may also help for the newer estimation methods, for example, to identify and remedy problems.

Aim and overview

The aim of our study is not to connect and compare IRT models and ordinal factor models with different parameterizations as such or more in general, but to illustrate how and why they differ specifically in their consequences when fitted to a kind of data structure that would lead to Heywood cases in linear factor models. We will describe and explain anomalies in the results when a Heywood case is expected for the data under consideration. We focus on binary data, but the same principles apply to polytomous ordered-category data fitted to an ordinal factor model or to any of the three types of IRT models. Because different options are available for how ordered-category data with three or more categories can be modeled with IRT and because the associated consequence of a Heywood case using one approach does not necessary show for the other, we decided to focus on binary data to avoid distraction from the core message.

In the next section, we describe the different approaches to answer the question of what would happen if a Heywood case is encountered, and a delta parameterized or theta parameterized ordinal factor model is fitted to the data, or an IRT model is fitted to the data. In the sections after that, we illustrate our conclusions with a small simulation study and a real dataset; the data in both illustrative sections are binary. Finally, we discuss implications of our findings.
Three Approaches and Their Heywood Case Issues

Delta Parameterization

The observed values of ordered-category variables in an ordinal factor model are assumed to stem from underlying continuous variables, denoted as $V$ in the following, also called latent response variables, with a normal distribution and thresholds for discretization (into ordered-category response options). For binary variables, just one threshold is needed to differentiate between the observed values of 0 and 1. For identification reasons, a parameterization choice needs to be made for the latent response variables. Delta parameterization means that the total variance of the latent response variable is set to 1.00. A Heywood case means that more than 100% of the variable’s variance is explained by the (latent) factor(s), which implies that the residual variance is negative. For a model with one factor, it means that the (standardized) loading is larger than 1.00.

The equation for the variance of the latent response variable with delta parameterization, $\sigma^2_{V_i}$, is:

$$\sigma^2_{V_i} = \lambda^2_{\delta i} + \sigma^2_{\delta i}$$

with $\sigma^2_{V_i} = 1$ (in the left-hand side of the equation) as the variance of the latent response variable $i$ in the delta parameterized model, $\lambda^2_{\delta i}$ as its squared loading, and $\sigma^2_{\delta i}$ as its residual variance. To avoid having the total variance $\sigma^2_{V_i}$ exceeding 1.00, the model estimates the residual variance as negative if $\lambda_{\delta i} > 1$.

Theta Parameterization

For theta parameterization, the ordered-category variables in an ordinal factor model are also assumed to stem from the continuous latent response variables, but for this type of parameterization, the residual variance, instead of the total variance, is set to 1.00. The equation for the variance of the latent response variable with theta parameterization, $\sigma^2_{Y_i}$, is:

$$\sigma^2_{Y_i} = \lambda^2_{\theta i} + 1$$

with $\sigma^2_{Y_i}$ as the variance of variable $i$ in the theta parameterized model, $\lambda^2_{\theta i}$ as its squared loading and $\sigma^2_{\theta i} = 1$ as its residual variance (the second term on the right-hand side of the equation). The theta loading $\lambda_{\theta i}$ can be translated into the delta parameterized factor loading $\lambda_{\delta i}$ as follows:

$$\lambda_{\delta i} = \sqrt{\lambda^2_{\theta i} / (1 + \lambda^2_{\theta i})}$$

Under the square root is the ratio of explained variance $\lambda^2_{\theta i}$, given that the factor variance is 1.00) versus the total variance $(1 + \lambda^2_{\theta i})$. It follows that:

$$\lambda_{\delta i} = \sqrt{\lambda^2_{\theta i} / (1 - \lambda^2_{\theta i})}$$.

Equation (3) shows that a theta loading when translated into a delta loading yields a delta loading smaller than 1.00 (with a limit value of 1.00), and equation (4) shows that the translation of a delta loading into a theta loading is problematic if $\lambda^2_{\delta i} \geq 1$. A Heywood case with delta parameterization (i.e., $\lambda_{\delta i} > 1$) would lead to a negative value under the square root in equation (4). For a factor to explain 100% of the variance of a variable (the boundary condition of Heywood cases), the delta parameterized factor loading, $\lambda_{\delta i}$ would need to be 1.00, and the theta parameterized loading, $\lambda_{\theta i}$ would be infinitely large because the proportion explained variance is $\lambda^2_{\theta i} / (1 + \lambda^2_{\delta i})$. When $\lambda_{\delta i} = 1.00$, the proportion explained variance is 1.00, making $\lambda_{\delta i}$ infinitely large (see
equation (4)). Iterations on the way to extremely high proportions of explained variance would necessarily lead to convergence issues because even large increases in an already large theta loading would translate into extremely small and barely noticeable increases in proportions of explained variance as well as into barely noticeable decreases of the discrepancy between the polychoric correlations and the model-based correlations. This is equivalent with an optimization process that reaches a plateau and cannot find the highest point on the plateau, or equivalently, the lowest point in a flat valley. Therefore, when applying the same data to factor models with delta parameterization and with theta parameterization, we expect a Heywood case in the delta parameterized factor model to co-occur with (1) a nonconvergence case in the theta parameterized factor model and (2) extremely high theta parameterized loadings when the iterations are stopped without convergence.

We believe that some of the nonconvergence cases encountered in Paek et al. (2018) in theta parameterized factor models may show as Heywood cases if the same data were fitted to delta parameterized models instead. Paek et al. simulated binary IRT data and analyzed these data with (1) theta parameterized factor models using three different limited information least squares estimators in Mplus (Muthen & Muthen, 1998-2017): fully weighted least squares (WLS), diagonally weighted least squares (WLSMV), and unweighted least squares (ULS) and (2) logistic and probit IRT models using full information Maximum Likelihood (ML) estimation. The WLS estimator is also known as the ADF approach, fully weighted least squares with no distributional assumption (Browne, 1984), mostly used for observed continuous variables, and the diagonally weighted least squares is implemented if WLSMV is selected as the estimator in Mplus and in the cfa( ) function from lavaan (Rosseel, 2012). Paek et al. (2018) aimed to compare parameter recovery and standard errors, but the authors also found a substantial nonconvergence rate with WLS, much more than with WLSMV and ULS, while nonconvergence did not occur with the IRT approach. Although Paek et al. did not use delta parameterization and did not consider Heywood cases, we suspect that some of the datasets showing convergence issues with theta parameterization would show Heywood cases in an analysis with delta parameterization.

**IRT models**

To compare IRT models with ordinal factor models, a latent response formulation of IRT models is presented here. For the unidimensional normal ogive model (the probit link model), assuming a standardized latent variable, equation (2) (theta parameterization) applies, with the $\lambda$ parameter as the item discrimination. Therefore, the parameterization is the same as for the theta parameterized factor models. For the corresponding and more popular logistic model (the logit link model), the situation is more complex. A simple additive variance decomposition cannot be made because the underlying continuous response variable would be the sum of two terms with different distributions. Per item, the first term is commonly but not necessarily assumed to be normally distributed (the item discrimination times the latent variable plus the item intercept\(^1\)), but the second (i.e., the residual term) is logistically distributed, with a standard logistic variance of 3.29. To make the logistic model approximately comparable with the normal ogive model, a scale factor of 1.7 is used\(^2\). Independent of these ramifications, the item parameters are expressed relative to the square root of the residual variance (the standard deviation of the residual), which is fixed, and—importantly for our study—it can never be negative, neither in the probit version nor in the logit version of the model.

To estimate IRT models, a marginal maximum likelihood (MML) approach based on full data information is commonly used, rather than limited information statistics such as the polychoric correlations. Because the optimization is not formulated in terms of the polychoric correlation matrix, the previous convergence issues, stemming from extremely small improvements in
recovering the polychoric correlations, do not occur. This explains why Paek et al. (2018) never run into convergence issues using full information IRT based estimation despite that IRT models, the same as theta parameterized factor models, assume fixed residual variances.

When data showing Heywood cases in delta parameterized factor models are fitted to IRT models, we expect extremely high discrimination values (i.e., the “loadings” in an IRT model). An extremely high discrimination implies an extremely high proportion explained variance. For example, 10 is an extremely high discrimination. To illustrate the extremity, let us assume that the standardized latent variable has a value of 0.00 and an item difficulty of 0.00, which corresponds to a probability of 0.50 of observing a 1-response to a binary item. The corresponding probability would increase to \( \exp(-7.62\times10^{-24}) \) for an increase of one unit of the latent variable value if the item discrimination were 10. For Heywood cases, the proportion explained variance (of the latent response variable) exceeds 1.00, which is not possible when the residual variance is fixed, as is the case in theta parameterized factor models and in IRT models. However, we do not expect this type of convergence issues in an IRT model because pairwise relationships between item responses in an IRT model are not the target of the optimization. The relationships between item responses are only one aspect of the full information used for estimating IRT models.

Simulation-Based Illustration

Method

To illustrate the differences of the three approaches (delta and theta parameterized factor models with limited information estimation and IRT models with full information estimation), we conducted a small simulation study with just 100 datasets, each with a sample size of 200, \( N = 200 \). Only a small simulation study is needed because the purpose is to illustrate the differences, not to find an answer to a research question. The data were generated in two steps: obtaining normally distributed continuous data in the first step and dichotomizing the continuous data into binary data in the second step.

First, we generated 100 datasets, each with four normally distributed continuous variables with means of 0.00, variances of 1.00 and a covariance matrix (in this case, a correlation matrix due to the unit variance) as shown in Table 1. The matrix in Table 1 is the population covariance matrix for the continuous variables. The data were generated with the mvrnorm() function from the MASS package in R (Venables & Ripley, 2002) without enforcing the estimated covariance matrix to be identical to the theoretical (true) covariance matrix (accomplished by using the option empirical = FALSE in the mvrnorm() function). The theoretical covariance matrix in Table 1 was chosen in such a way that there would be a substantial probability for the generated data to show Heywood cases after dichotomization.

An important condition we specified for the theoretical covariance matrix was that continuous data generated with the mvrnorm() function with the empirical = TRUE option would lead to a Heywood case when a one-dimensional linear factor model is fitted to the data. In this way, a

|     | Y1 | Y2 | Y3 | Y4 |
|-----|----|----|----|----|
| Y1  | 1.00 | | | |
| Y2  | 0.10 | 1.00 | | |
| Y3  | 0.48 | 0.76 | 1.00 | |
| Y4  | 0.48 | 0.48 | 0.48 | 1.00 |
dataset was generated in perfect agreement with the population covariance matrix without specifying the factor model parameters. Using the default ML estimator in the cfa () function from lavaan 0.6–7 (Rosseel, 2012) to estimate a linear factor model with one factor, the residual variance of the 3rd variable (Y3) had indeed a negative value, which gave us confidence in the proposed covariance matrix. The goodness-of-fit of the model was poor (RMSEA = 0.519, CFI = 0.721, SRMR = 0.139), as one could expect from the very small correlation between Y1 and Y2.

Next, for the 100 continuous datasets generated with the empirical = FALSE option, a threshold of 0.00 was used to dichotomize the data, the same threshold for all variables in all datasets. In this way, 100 datasets consisting of binary variables were obtained. This is an indirect pragmatic way to generate data with a substantial probability of Heywood cases, without having to define a Heywood case factor model with a negative residual variance. We wanted to avoid using such a model to generate data from because a negative residual variance would make it an improper model.

We expect (1) a substantial proportion of Heywood cases out of all fitted ordinal factor models with delta parameterization, (2) a substantial proportion of nonconvergent cases out of all fitted ordinal factor models with theta parameterization, and (3) a substantial proportion of IRT models showing extremely large discrimination parameter estimates. To be clear, we use the term “Heywood case” exclusively for converged models with negative residual variances, even though the output in cases of nonconvergence may also show negative residual variances. Based on the theoretical covariance matrix, we expect variable Y3 to be the problematic variable for most datasets showing: a negative residual variance with delta parameterization, an extremely high loading when the iterations are stopped due to nonconvergence with theta parameterization, and an outlying discrimination parameter value with IRT.

The 100 dichotomized datasets were analyzed with the cfa () function of lavaan 0.6–7 using the limited information estimators including WLS (i.e., ADF), WLSMV (diagonally weighted least squares), and ULS (unweighted least squares), and each estimator was used in both delta and theta parameterized ordinal factor models. The reason to use three estimators is to investigate whether our findings are estimator specific. While WLS is primarily used for continuous variables, ULS is not, but is shown to perform well for ordinal factor models (Forero & Maydeu-Olivares, 2009; Yang-Wallentin et al., 2010; Paek et al., 2018). It is unclear from the literature whether the estimators have an effect on the occurrence of Heywood case related issues. The data were also fitted to an IRT model with the grm() function from the ltm package (Rizopoulos, 2006). The frequency of Heywood cases and the range of the resulting negative residual variances were determined for each estimator with delta parameterization. For the WLSMV estimator, the default estimator in lavaan, a more detailed description of the negative values is given. As will be noticed, the results with the ULS estimator are highly similar. The frequency of nonconvergence cases is reported per estimator with theta parameterization, and the range of extreme loadings when the iterations are stopped under nonconvergence are determined. Finally, for the IRT approach, the frequency of nonconvergence cases and an overview of estimated discriminations are reported.

**Results**

**Delta Parameterization**

Using WLSMV and ULS as estimators, the fitted ordinal factor models with delta parameterization never showed convergence issues, but they did show Heywood cases for 34 and 33 out of the 100 datasets, respectively. The program reported warning messages indicating that negative variances were observed. In contrast, using WLS as estimator, the fitted ordinal factor models with delta
parameterization reported Heywood cases in 51 out of the 100 datasets and failed to converge in 4 out of the 100 generated datasets. When the models failed to converge, extremely negative values of the residual variances (−2626745, −694,577, −4,376,004 and −1,309,322) were observed in the output. In these cases, the `cfa()` function reported warning messages about convergence issues, but not about negative residual variances. These models were counted as nonconvergent cases, but not Heywood cases. For most of the Heywood cases, variable Y3 was the problem showing up in 47 out of the 51 Heywood cases with WLS estimators, 34 out of 34 with WLSMV and 33 out of 33 with ULS. In only 4 Heywood cases, with WLS estimators, the residual variances of Y2 were negative. Figure 1 shows the histogram of the 34 negative residual variances using the WLSMV estimator with delta parameterization. Most negative values are small. The range of negative values was similar for WLS and ULS, but the frequency of negative values was larger for WLS and about the same for ULS.

**Theta Parameterization**

Of the 100 datasets generated, nonconvergence was found for 54 datasets with WLS estimators, 34 datasets with WLSMV and 33 datasets with ULS, totaling 121 cases. Of the 121 cases of nonconvergence, the reported loadings were extremely large (positive or negative) for at least one predictor (always for Y3), and in 4 cases, extreme loadings were reported for 2 variables (Y2 and Y3). In a few rare cases, extreme loadings for variable Y3 were found in converged models: 2/46 with WLS, 2/66 with WLSMV, and 2/67 with ULS, and none was found for Y2. Finally, for all datasets with negative residual variances or nonconvergence when using delta parameterization, convergence issues were encountered when the same data were fitted to theta parameterized models. This result illustrates that the negative residual variances in delta parameterized models co-occur with nonconvergent cases in theta parameterized models.

![Histogram of negative residual variances](image)

**Figure 1.** Histogram of negative residual variances obtained using the WLSMV estimator (the default in lavaan) with delta parameterization.
Although less important than the finding that the anomalies occur with all three estimators, it was also found that the WLS estimator reported the most problematic results in both delta and theta parameterized models compared to the other two limited information estimators, WLSMV and ULS, in line with the findings by Paek et al. (2018). The WLS estimator is known to be based on a complex algorithm that works best for extremely large datasets and continuous variables. Our results also show that in comparison with WLS, both the WLSMV and the ULS estimators seem to be more robust to possible negative variance problems in delta parameterized models and to convergence issues in theta parameterized models.

**IRT Model Estimation**

All IRT models fitted to the generated data converged though some showed extremely large discrimination parameter estimates. Discrimination estimates of Y3 ranged from 2.26 to 18.77, and discrimination estimates of Y2 ranged from 1.24 to 13.89. For 58 out of 100 datasets, the estimated Y3 discrimination is larger than 10, and only for 1 other dataset, the estimated Y2 discrimination is larger than 10. None of the datasets showed both estimated discriminations larger than 10. *Figure 2* shows a histogram of the most extreme discrimination parameter estimates per dataset. The discrimination estimates for the 34 Heywood cases obtained with delta parameterization and WLSMV ranged from 10.18 to 17.85; they are among the highest 58 values in *Figure 2*.

To summarize, the results of the simulation study illustrate that the same data structure can lead to very different estimation results depending on how the model is formulated. Data structures that yield negative residual variance estimates using delta parameterized factor models may lead to nonconvergence using theta parameterized factor models, and to extreme discrimination parameter estimates using IRT models. The kind of estimation anomalies obtained with factor model

![Figure 2. Histogram of the largest discrimination parameters per dataset.](image)
estimations is not specific to the estimator that is used and can therefore be generalized although the extent to which the predicted anomalies show does seem to depend somewhat on the estimator. WLS seems to generate more anomalies, which is not surprising given the binary nature of the data.

**Application-Based Illustration**

The data for the application are a subset from a larger dataset (Yotebieng et al., 2016; Yotebieng et al., 2017), concerning five PHQ9 items with \( N = 200 \), randomly selected from a larger sample of respondents. The purpose of analyzing these data is to illustrate the differences among delta and theta parameterized ordinal factor models and IRT models with respect to the issues discussed earlier. The PHQ9 is a self-reported depression questionnaire intended to be one-dimensional (Spitzer et al., 1999), and in this study, we included items 1, 2, 3, 7, and 8 referencing anhedonia, feeling depressed, sleep problems, concentration problems, and movement problems, respectively. To be in line with the simulation that focuses on binary data in the analysis of the parameterizations, we have dichotomized the item responses. Dichotomized responses were obtained by recoding 0 and 1 as 0, and 2 and 3 as 1. The data were then fitted to delta and theta parameterized one-factor ordinal factor models using lavaan with the default WLSMV estimator and to an IRT model with MML using ltm. The purpose of the analyses is purely illustrative. We do not otherwise recommend a dichotomization, as it implies a loss of information.

The results of the delta parameterized factor model are shown in **Table 2**. The model had a good fit, with RMSEA = 0.00 and SRMR = 0.044, but the residual variance of item 2 (feeling depressed) was negative, which means that a Heywood case was encountered.

The theta parameterized model failed to converge, as expected. The loading of item 2 was reported as 14,850.40 when the iterations were stopped. Similarly high loadings are usually found when delta parameterized models failed to converge, but for these data, the delta parameterized model did converge, albeit with a negative residual variance estimate. The application suggests that the same dataset showing a Heywood case in the delta parameterized model would show nonconvergence in the theta parameterized model, for reasons explained earlier.

Finally, we fitted an IRT model to the same data using the ltm package in R. The results are shown in **Table 3**. The model converged, but the problematic item 2 was reported with a discrimination of 13.917, an extreme value on the logit scale.

The PHQ9 application illustrates how estimating the same model using different parameterizations can yield quite different results with respect to improper and non-converged solutions, and yet these results are threaded by the same underlying problem. Using delta and theta parameterized factor models and IRT models, we showed three different ways the same problem is manifested in model estimation.

| Item | Standardized Loading | Residual Variance |
|------|----------------------|-------------------|
| Item 1 | 0.795 | 0.368 |
| Item 2 | 1.011 | −0.021 |
| Item 3 | 0.744 | 0.446 |
| Item 7 | 0.644 | 0.585 |
| Item 8 | 0.672 | 0.548 |
To resolve the estimation issue in the delta and theta parameterized factor models for the present PHQ9 dataset, most researchers would consider dropping the problematic items from the scale. Alternatively, one can add correlated residuals for items 1 and 2 and for items 2 and 8 in the model specification. We chose these two pairs of items because they had the largest modification indices. It is more complex to deal with remaining dependencies in an IRT model though local dependencies can be detected using well-functioning methods (e.g., Edwards et al., 2018). How to extend an IRT model to accommodate local dependencies is not yet clear, but see Braeken et al. (2007) for a copula approach, and Thissen, Steinberg, and Mooney (1989), Hoskens and De Boeck (1997) and Ip (2002) for other options. The fact that the inclusion of local dependencies solves the problem (as suggested based on modification indices) is an interesting finding. The dependences may either stem from the true covariance matrix or they may originate from sampling variation and thus from the unreliability of observed covariances. The latter is of course less likely the larger the sample size is.

Discussion

The theoretical discussion and the illustrations show that the same problem appears in different forms depending on the method for analysis. Delta and theta parameterized ordinal factor models and IRT models produce different anomalies resulting from the same problem: a negative residual variance (in delta parameterized ordinal factor model), nonconvergence (in theta parameterized ordinal factor model) and largely outlying discrimination parameter estimates (in an IRT model). When using IRT models and if the results show an extremely large discrimination estimate, we recommend a further analysis. Because it is known that local dependencies can lead to large discrimination values (Edwards et al., 2018; Ip, 2000; Tuerlinckx & De Boeck, 2001), one may use local dependence statistics to investigate the problem. Although it may sound as an unusual recommendation, we believe it may also be of interest for diagnostic reasons to apply a delta parameterized ordinal factor model and verify whether a negative residual variance shows up.

If the theta parameterized ordinal factor model is the first choice, and one encounters nonconvergence, one can move to a delta parameterized model instead to explore the sign of residual variances. There are other possible causes of nonconvergence, but a Heywood case is one of them, so that a delta parameterized analysis may aid a diagnosis. We believe that it is helpful for users if they can interpret the issues they encounter with a chosen parameterization.

An evident general recommendation is to work with a large sample size, to reduce the probability of Heywood cases due to unreliable covariances. If the true correlation matrix does not imply a Heywood case (given the estimation model), a large sample size helps to avoid Heywood cases. However, if the true correlation matrix does imply a Heywood case (given the estimation model), a large sample size would increase the probability of finding a Heywood case. However,
to compare the three methods of analysis for the purpose of how a Heywood issue shows the sample size is less important. We have worked with \( N = 200 \).

A limitation of our study is that it does not offer a more fundamental explanation for Heywood cases, and therefore a more detailed solution for the problem. Our purpose was to explain and report the different appearances of the same problem under the three different approaches. The causes of Heywood cases and nonconvergence issues (in theta parameterized models) are beyond the scope of our study. For causes and solutions of Heywood cases in linear factor models, we refer readers, among others, to Chen et al. (2001). A further investigation is required to find out how well the issues with linear factor models apply to models for ordered-category data, including binary data. It would also be valuable to investigate other (and new) methods than the three we investigated to establish how Heywood cases show. Nonetheless, we hope that our comparative study of three commonly used methods for the analysis of binary data contributes to the interpretation of outlying discrimination values encountered in IRT models, negative residual variances in delta parameterized factor models, and nonconvergence issues encountered in theta parameterized factor models.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the NIHCD (R01HD075171).

Author’s Note
The data for the application are from a study funded by grant R01HD075171 from NIHCD.

ORCID iD
Selena Wang https://orcid.org/0000-0002-4358-5382

Notes
1. Three different forms are possible. We adopt the form that shows a close resemblance to the general factor model shown in equation (2). This form is also adopted by Kamata and Bauer (2008) and Takane and de Leeuw (1987).
2. See Savalei (2006) for a discussion of the logistic approximation to the normal.
3. The WLSMV becomes the default when at least one endogenous variable is specified as ordinal.
4. The data are from a study funded by grant R01HD075171 from NIHCD.

References
Asparouhov, T., & Muthen, B. (2020). IRT in Mplus. https://www.statmodel.com/download/MplusIRT.pdf
Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. Applied Psychological Measurement, 12(3), 261–280. https://doi.org/10.1177/014662168801200305
Bolt, D. M. (2005). Limited- and full-information estimation of item response theory models. In A. Maydeu-Olivares, & J. J. McArdle (Eds.), Contemporary psychometrics (pp. 27–71): Lawrence Erlbaum.
Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copulas for residual dependencies. Psychometrika, 72(3), 393–411. https://doi.org/10.1007/s11336-007-9005-4
Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*(1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*(4), 468–508. https://doi.org/10.1177/0049124101029004003

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods, 23*(1), 138–149. https://doi.org/10.1037/met0000121

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods, 14*(3), 275–299. https://doi.org/10.1037/a0015825

Harman, H. H., & Fukuda, Y. (1966). Resolution of the Heywood case in the minres solution. *Psychometrika, 31*(4), 563–571. https://doi.org/10.1007/bf02289525

Heywood, H. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 134*(824), 486–501.

Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods, 2*(3), 261–275. https://doi.org/10.1037/1082-989x.2.3.261

Ip, E. H. (2000). Adjusting for information in inflation due to local dependency in moderately large item clusters. *Psychometrika, 65*(1), 73–91. https://doi.org/10.1007/bf02294187

Ip, E. H. (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika, 67*(3), 367–386. https://doi.org/10.1007/bf02294990

Jöreskog, K. G., & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 21*(1), 85–96. https://doi.org/10.1111/j.2044-8317.1968.tb00399.x

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal, 15*(1), 136–153. https://doi.org/10.1080/10705510701758406

Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika, 40*(4), 505–517. https://doi.org/10.1007/bf02291552

Muthen, L. K., & Muthen, B. O. (1998-2017). *Mplus user’s guide* (8th ed.): Muthen & Muthen.

Paek, I., Cui, M., Öztürk Gübes¸, N., & Yang, Y. (2018). Estimation of an IRT model by Mplus for dichotomously scored responses under different estimation methods. *Educational and Psychological Measurement, 78*(4), 569–588. https://doi.org/10.1177/0013164417715738

Rizopoulos, D. (2006). Itm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25. http://www.jstatsoft.org/v17/i05/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Savalei, V. (2006). Logistic approximation to the normal: The KL rationale. *Psychometrika, 71*, 763–767. https://doi.org/10.1007/s11336-004-1237-y

Spitzer, R. L., Kroenke, K., & Williams, J. B.W. (1999). Patient health questionnaire study group. Validity and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA, 282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*(3), 393–408. https://doi.org/10.1007/bf02294363

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement, 26*(3), 247–260. https://doi.org/10.1111/j.1745-3984.1989.tb00331.x
Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods, *6(2), 181–195. https://doi.org/10.1037/1082-989x.6.2.181

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.): Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Yang-Wallentin, F., Joreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, *17(3), 392–423. https://doi.org/10.1080/10705511.2010.489003

Yotebieng, K. A., Fokong, K., & Yotebieng, M. (2017). Depression, retention in care, and uptake of PMTCT service in Kinshasa, the Democratic Republic of Congo: A prospective cohort. *AIDS Care, *29(3), 285–289. https://doi.org/10.1080/09540121.2016.1255708

Yotebieng, M., Thirumurthy, H., Mohalaracco, K. E., Kawende, B., Chalachala, J. L., Wenzi, L. K., & Behets, F. (2016). Conditional cash transfers and uptake of and retention in prevention of mother-to-child HIV transmission care: A randomized controlled trial. *Lancet HIV, *3(2), e85–e93.