AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

2022, Vol. 27, No. 2, 156–176
https://doi.org/10.1037/met0000384

Heywood You Go Away! Examining Causes, Effects, and Treatments for Heywood Cases in Exploratory Factor Analysis

Allison W. Cooperman and Niels G. Waller
Department of Psychology, University of Minnesota, Twin Cities

Abstract

Exploratory factor analysis (EFA) is a popular method for elucidating the latent structure of data. Unfortunately, EFA models can sometimes produce improper solutions with nonsensical results. For example, improper EFA solutions can include one or more Heywood cases, where common factors account for 100% or more of an observed variable's variance. To better understand these senseless estimates, we conducted four Monte Carlo studies that illuminate the (a) causes, (b) consequences, and (c) effective treatments for Heywood cases in EFA models. Studies 1 and 2 showed that numerous model and data characteristics are associated with Heywood cases, such as small sample sizes, poorly defined factors with low factor score determinacy values, and factor overextraction. In Study 3, we examined the consequences of Heywood cases for EFA model interpretation and found that Heywood cases increase factor loading variances and upwardly bias factor score determinacy values. Study 4 compared the model recovery of several EFA algorithms that were designed to avoid Heywood cases. Our results indicated that, among the algorithms compared, regularized common factor analysis (Jung & Takane, 2008) was the most reliable method for avoiding Heywood cases and producing EFA parameter estimates with small mean squared errors. We discuss best practices for conducting EFA with data sets that might yield Heywood cases.

Translational Abstract

Most psychological traits (e.g., depression, personality) cannot be directly measured. Thus, models of causal structure in the social sciences routinely include latent variables. Exploratory factor analysis (EFA) is a popular technique in the psychological sciences for illuminating the underlying structure of a set of measured variables. Yet like all statistical models, EFA models can produce nonsensical results. For example, estimated EFA models can include Heywood cases, where common factors are estimated to account for 100% or more of an observed variable's variance. We conducted a series of Monte Carlo simulation studies to illuminate the (a) causes, (b) consequences, and (c) effective treatments for Heywood cases in EFA models. Numerous model and data characteristics were differentially prone to produce Heywood cases, including sample size, factor loading sizes, the factor extraction method, and incorrect dimensionality specifications. Our results showed that regularized common factor analysis is an effective extraction method for avoiding Heywood cases, particularly in situations when other treatment methods (e.g., adding more subjects or variables) are not plausible. We discuss best practices for the use of EFA techniques under conditions where Heywood cases are more common. Implications for reproducibility and model interpretation are also reviewed.

Keywords: Heywood cases, improper solutions, exploratory factor analysis, Monte Carlo simulation

Supplemental materials: https://doi.org/10.1037/met0000384.supp

A well-known problem in the social sciences is that many psychological traits cannot be directly measured (Cronbach & Meehl, 1955; Loevinger, 1957). Thus, biobehavioral models of causal structure routinely include latent variables (Bollen, 1989; Loehlin, 2004) that must be estimated from observed data. Among the several methods that have been developed for this purpose, exploratory factor analysis (EFA) is one of the most common methods for highlighting the latent structure of psychological data (Fabrigar et al., 1999; Ford et al., 1986; Goretzko et al., 2019; Henson & Roberts, 2006; Mulaik, 2010; Preacher & MacCallum, 2003; Thurstone, 1947). For example, a Google Scholar search (conducted on May 7, 2020) for "exploratory factor analysis" returned over 30,000 related articles from the past five years. One reason for the continuing popularity of EFA is its ability to elucidate parsimonious models for

This article was published Online First July 1, 2021.

Preliminary results were presented in a poster at the University of Minnesota's Institute for Research in Statistics and its Applications Conference on Causal Inference and Data Science in May 2019 and at the International Meeting of the Psychometric Society in July 2020.

Correspondence concerning this article should be addressed to Niels G. Waller, Department of Psychology, University of Minnesota, Twin Cities, Elliott Hall, 75 East River Road, Minneapolis, MN 55455, United States. Email: nwaller@umn.edu

high-dimensional data sets (Bentler & Mooijaart, 1989; Fabrigar & Wegener, 2011). That is, when the model fits, factor analysis allows researchers to represent the correlational structure of a potentially large number of observed variables with a smaller number of more fundamental latent variables (Fabrigar et al., 1999; Lattin et al., 2003; Mulaik, 2010).

Unfortunately, EFA can produce nonsensical results under many scenarios. For instance, with some data sets, EFA can produce a unique factor variance that is zero or negative—the former case being psychometrically implausible and the latter case being mathematically senseless. When this occurs, the solution is said to contain one or more *Heywood cases* (Heywood, 1931).[1] Unfortunately, Heywood cases arise often in EFA applications (Costello & Osborne, 2005; De Winter & Dodou, 2012; Fabrigar et al., 1999; Kolenikov & Bollen, 2012; Preacher & MacCallum, 2002; Velicer & Jackson, 1990).[2] Despite previous work in this area (e.g., Briggs & MacCallum, 2003; De Winter et al., 2009; De Winter & Dodou, 2012; Kano, 1998; MacCallum et al., 1999; Van Driel, 1978; Velicer & Fava, 1998), particularly in confirmatory factor analysis (CFA; Anderson & Gerbing, 1984; Boomsma, 1985; Chen et al., 2001; Dillon et al., 1987; Rindskopf, 1984; Sato, 1987), a paucity of studies have been directly designed to uncover the causes and consequences of Heywood cases in EFA (although see Dillon et al., 1987; Gerbing & Anderson, 1987). As a result, the etiology of Heywood cases and their effects on EFA model interpretation are still not fully understood.

In this article, we summarize the results of four Monte Carlo studies that illuminate the (a) causes of, (b) consequences of, and (c) effective treatments for Heywood cases in EFA. In Study 1, we examined eight putative causes of Heywood cases in EFA models (e.g., small sample sizes, small indicator-to-factor ratios, the presence of various levels of model approximation error). Next, in Study 2, we conducted targeted simulations to more fully explore three specific (and understudied) causes of Heywood cases: (a) factor score determinacy (FSD; Guttman, 1955); (b) correlated common factors; and (c) matrix smoothing algorithms (Bentler & Yuan, 2011; Higham, 2002; Knol & Berger, 1991). In Study 3, we focused on the effects of Heywood cases on EFA model recovery, paying particular attention to potential biases in dimensionality assessment via parallel analysis, factor loading recovery, and FSD values. Finally, in Study 4, we compared various model estimation algorithms (principal axes, constrained least-squares, and regularized common factor analysis; Jung & Takane, 2008) in their ability to recover known model parameters in data sets that produce Heywood cases under at least one estimation method. In aggregate, we analyzed over 4,000,000 data sets. All simulations were conducted in R (R Core Team, 2019), and all of our program code is included in the online supplemental materials. We end our article by proposing best practices for applying EFA to data sets that might yield Heywood cases.

We first provide a brief review of the underlying mechanics of EFA, arguably the most common latent variable model for examining the dimensionality among a set of items or scales (Fabrigar et al., 1999; Goretzko et al., 2019; Henson & Roberts, 2006). In this article, we focus our attention on the linear factor analysis model. A fundamental premise of factor analysis is that observations on p manifest variables arise from linear combinations of k "common" factors and p "unique" factors ($k \ll$ p). In this model, the common factors represent the hypothesized latent traits that are presumed to influence two or more observed variables. Each unique factor influences only a single observed variable and accounts for both the unreliable variance due to measurement error, and the reliable variance that is specific to that variable.[3] Mathematically, we can represent the common factor analysis model as follows (Lattin et al., 2003, p. 142; Mulaik, 2010, pp. 135–136; Thurstone, 1947):

$$\mathbf{Y} = \mathbf{X}\mathbf{\Lambda}' + \mathbf{E}\mathbf{\Psi}, \tag{1}$$

where $\mathbf{Y}$ is a $n \times p$ matrix of standardized observed data, $\mathbf{X}$ is a $n \times k$ matrix of common factor scores, $\mathbf{\Lambda}$ is a $p \times k$ matrix of common-factor pattern loadings, $\mathbf{E}$ is a $n \times p$ matrix of unique factor scores, and $\mathbf{\Psi}$ is a $p \times p$ diagonal matrix of unique-factor loadings. When the model holds, a model-implied dispersion matrix ($\mathbf{P}_{\mathrm{YY}}$) for the observed variables can be represented as (Mulaik, 2010, pp. 136–137; Thurstone, 1947)

$$\mathbf{P}_{\mathrm{YY}} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}^2, \tag{2}$$

where $\mathbf{\Phi}$ is a $k \times k$ matrix of common factor correlations, and $\mathbf{\Psi}^2$ is a $p \times p$ diagonal matrix of unique factor variances. When estimating the parameters of the EFA model with standardized data, we seek to match $\mathbf{P}_{\mathrm{YY}}$ with our observed correlation matrix ($\mathbf{R}_{\mathrm{YY}}$; Thurstone, 1947).

As previously noted, in some data sets, EFA can produce results that are psychometrically implausible (implying error-free measurements) or mathematically senseless (negative variances). As shown in Equation 2, an observed variable's variance (which is 1.00 in the fully standardized model) can be decomposed into two components: (a) the variance due to the common factors (termed the communality, h_i^2, and found on the diagonal of $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}'$); and (b) the variance due to the associated unique factor (u_i^2, found on the diagonal of $\mathbf{\Psi}^2$). In a proper solution with standardized data (Chen et al., 2001; Van Driel, 1978), both variance components are positive and sum to 1.00 (Mulaik, 2010). However, when a Heywood case occurs, an estimated communality can equal or exceed 1.00, which implies that the associated unique factor variance is zero or negative, respectively (Van Driel, 1978).

To illustrate the notion of a Heywood case, we analyzed a data set that was first described by Harman (1960) for this purpose. Table 1 displays the correlation matrix for these data (Harman, 1960, p. 125). With these correlations, we used the fungible R package (Waller, 2019) to estimate a one-factor model with two common

[1] There are different definitions of a Heywood case in the factor analysis literature. Some researchers define a Heywood case as a negative unique factor variance (Harman, 1960, p. 125; Mulaik, 2010, p. 389). Others differentiate between a Heywood case—a null unique factor variance, and an ultra-Heywood case—a negative unique factor variance (Revelle, 2018; SAS Institute Inc., 2015). In this article, we define a Heywood case as a unique factor variance that is either zero or negative.

[2] Although it is difficult to quantify the frequency of Heywood cases in applied research, a Google Scholar search for the term with factor analysis returned over 600 results from the past 5 years.

[3] Although the psychometric EFA model and the mathematical principal components analysis (PCA) model are sometimes mistaken for one another, note that the PCA model does not distinguish between reliable and unreliable sources of observed score variance (Velicer & Jackson, 1990; Widaman, 1993).

Table 1
Two Solutions to Harman's (1960) Data Set Showing a Heywood Case

Correlation matrix					PAF		ML	
					λ	h^2	λ	h^2
1					1.05	1.10	1.00	1.00
0.945	1				0.90	0.81	0.95	0.89
0.840	0.720	1			0.80	0.64	0.84	0.71
0.735	0.630	0.560	1		0.70	0.49	0.73	0.54
0.630	0.540	0.480	0.420	1	0.60	0.36	0.63	0.40

Note. Correlation matrix originally reported in Harman (1960, p. 125). PAF = principal-axis factoring; ML = maximum likelihood estimation; λ = factor-pattern loadings; h^2 = indicator communalities.

factor-extraction algorithms: iterative principal axis factors (PAF) and maximum likelihood (ML). The results from these analyses demonstrate that both extraction methods produced a Heywood case on Variable 1. Using PAF, the estimated communality was 1.10, a value that implies a unique factor variance of –.10. This solution matches the results that were originally reported by Harman (1960). Recall that the factor loadings in a one-factor model represent correlations between the observed variables and the single common factor. Thus, a loading of 1.10 falls outside of its acceptable range ($-1.00 \leq r \leq 1.00$). Notice also that the ML solution produced a Heywood case that accounted for exactly 100% of the observed variance for the first variable (implying that this variable was measured without error). For this example, the solution differences between PAF and ML are largely a function of constraints imposed by the underlying estimation procedure (e.g., the ML function in fungible requires upper bounds of 1.00 on the communalities, thus guaranteeing that a Heywood case will never exceed this value). Despite these estimation differences, Table 1 presents EFA solutions that are clearly problematic for drawing robust, or even plausible, inferences about the underlying factor structure. To better understand these results, we designed four interrelated Monte Carlo studies to garner a more detailed understanding of the differential causes of Heywood cases, and to better inform researchers regarding how to handle such improper solutions if and when they occur.

Study 1: A Systematic Study of Eight Causes of Heywood Cases in EFA

Previous research has highlighted data characteristics and model choices that are associated with Heywood cases in EFA. For instance, some research suggests that Heywood cases occur more often in smaller sample sizes (Costello & Osborne, 2005; De Winter & Dodou, 2012; Jung & Takane, 2008; Velicer & Fava, 1998; Zhang et al., 2010), when factors have fewer salient markers (De Winter & Dodou, 2012, 2016; MacCallum et al., 1999; Velicer & Fava, 1998), when population factor loadings are small in absolute value (e.g., .20–.40; De Winter & Dodou, 2016; MacCallum et al., 1999; Velicer & Fava, 1998), and when a population factor loading is large in absolute value (Van Driel, 1978). Additionally, Heywood cases are reportedly more common in samples with outliers (Kano, 1998; see also Bollen, 1987), when researchers extract too many common factors (De Winter & Dodou, 2012, 2016; Velicer & Fava, 1998; see also Sato, 1987; Ximénez, 2006), and when the observed data violate EFA's linearity assumption (Van Driel, 1978; see also Rindskopf, 1984). Further evidence suggests that factor extraction methods are differentially prone to produce Heywood cases. For example, Heywood cases have been reported to occur more often when using maximum likelihood (ML) estimation rather than unweighted least squares (ULS; Briggs & MacCallum, 2003; Jung & Takane, 2008) or principal axis factoring (PAF; De Winter & Dodou, 2012, 2016; although see Gorsuch, 1974; Snook & Gorsuch, 1989). Heywood cases may also arise from convergence issues during factor extraction (Kano, 1998). Many of these potential causes are described in the CFA literature (Anderson & Gerbing, 1984; Boomsma, 1985; Ding et al., 1995; Gagne & Hancock, 2006; Jackson et al., 2013; Marsh et al., 1998; McArdle, 1990; Rindskopf, 1984; Ximénez, 2006) because Heywood cases also occur in CFA models with non-negligible frequency (Chen et al., 2001).

Although the aforementioned summary of Heywood case causes may give the impression that little work remains in this area, much of what we know about Heywood cases is drawn from studies with primary focuses elsewhere (e.g., Briggs & MacCallum, 2003; De Winter & Dodou, 2012; Velicer & Fava, 1998). Additionally, some studies have treated Heywood cases in a piecemeal manner rather than using a fully integrated design that could elucidate potential interactions among Heywood causes (e.g., De Winter & Dodou, 2012). In light of these issues, our aim in Study 1 was to more comprehensively investigate Heywood case prevalence rates[4] using a fully-crossed factorial design that varied eight model and data characteristics that are known to produce—or that we hypothesized would produce—Heywood cases in EFA models.

Method

Study 1 was a fully-crossed Monte Carlo simulation that included eight design factors: (a) sample size (N = 50, 150, 300, 500); (b) number of common factors (k = 2, 3, 4); (c) number of indicators per factor ($p{:}k$ = 3:1, 5:1, 8:1); (d) salient loading pattern (constant high loadings, sequential loadings, one strong loading, one weak factor, two weak factors, constant low loadings); (e) cross-loading presence (either no cross-loadings or one cross-loading of size .20 added to each common factor); (f) model approximation error (no, low, or moderate levels of error); (g) model specification (defined as either the extraction of the correct number of common factors from the data-generating model, overfactoring by one factor, or underfactoring by one factor); and (h) factor extraction method (PAF vs. ML). These eight design factors were selected because they are among the most commonly-cited sources

[4] In this study, we use the phrase "Heywood case prevalence rates" to indicate the proportion of factor solutions with at least one Heywood case.

of Heywood cases in factor analysis models (e.g., Chen et al., 2001; De Winter & Dodou, 2012; Dillon et al., 1987; Kano, 1998; Velicer & Jackson, 1990).

Table 2 lists the six salient loading patterns that we used in Study 1 to investigate the effects of different factor loading sizes on Heywood case prevalence rates. As shown in this table, the constant high (A) and constant low loadings (F) conditions included salient factor loadings of .80 and .30, respectively. The sequential loadings (B) condition had similar patterns of salient loadings between factors, but the loading magnitudes were not fixed to common values. Rather, the loadings ranged from .80 to .30 and descended in equal increments in a stepwise manner. The one strong loading (C) condition included salient loadings of .50 for all indicators except one, where the loading was set to .80. The one weak factor (D) condition had salient loadings of .80 on all factors except the last, which had salient loadings of .30. Similarly, the two weak factors (E) condition had salient loadings of .80 on all factors except the last two, which had salient loadings of .30.

Model approximation error, a relatively understudied topic in the Heywood case literature (see Briggs & MacCallum, 2003; De Winter & Dodou, 2012) refers to the fact that the specified common factors are unlikely to represent all latent influences on indicator covariation (MacCallum, 2003; MacCallum & Tucker, 1991; Tucker et al., 1969). Rather, in the presence of model approximation error, numerous minor factors also contribute to indicator covariation. To account for these minor influences in the data-generation stage of our simulations, we introduced model approximation error via the following equation (Briggs & MacCallum, 2003; MacCallum, 2003; MacCallum & Tucker, 1991):

$$\tilde{\mathbf{P}}_{\mathrm{YY}} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}^2 + \mathbf{WW}'. \tag{3}$$

In this equation, $\tilde{\mathbf{P}}_{\mathrm{YY}}$ is the model-implied dispersion matrix for the observed variables with added model error, $\mathbf{W}$ is a $p \times 150$ matrix of minor factor loadings, and all other parameters are as previously defined in Equation 2. Across models, the proportion of model approximation error was varied by controlling the size of the population model's root mean square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger & Lind, 1980) value. Specifically, following the guidelines proposed in the model fit literature (Browne & Cudeck, 1992; Marsh et al., 2004), models with low levels of approximation error were chosen to have RMSEA values between .00 and .05. Similarly, models with moderate levels of error produced RMSEA values between .07 and .089. To generate appropriate population models, we reallocated a proportion of unique factor variance (for each variable) to the minor factors in $\mathbf{W}$. Following the method proposed by Tucker et al. (1969), we fixed the number of minor factors at 150 and varied ϵ between .02 and .20 (see (Briggs & MacCallum, 2003, p. 32) for a description of the ϵ parameter). To maintain the number of major common factors in conditions with model approximation error, $\mathbf{W}$ was constructed such that each column had (a) all factor loadings less than or equal to $|0.30|$ and (b) no more than two factor loadings equal to $|0.30|$ (Waller, 2019).

In total, we investigated 1,296 population models. For each model, we generated 3,000 sample data sets (drawn from a model-implied multivariate normal distribution), each of which was analyzed using both PAF and ML factor extraction. In 1,000 data sets, we extracted the known, data-generating number of factors (ignoring the minor factors described in Equation 3). In another 1,000 data sets, we extracted one too many common factors, and in the remaining 1,000 data sets we extracted one too few common factors. Only data sets that produced converged PAF factor solutions (using a convergence threshold of .0001 and a maximum of 15,000 iterations) were included in our analyses. In other words, we continued to generate data sets until we had 1,000 suitable samples for each condition.[5] For each factor extraction method and factor model, we calculated the proportion of solutions that produced at least one Heywood case. In this context, a Heywood case was defined as a communality greater than or equal to 1.00 with PAF extraction, or a communality greater than or equal to .998 with ML extraction (cf. De Winter & Dodou, 2012). Note that because the ML factor analysis function in R (factanal; R Core Team, 2019) requires that the estimated unique factor matrix be invertible, we had to use this slightly modified definition of a Heywood case when using ML. The ML results reported below could differ slightly under alternative communality bounds.

All analyses were conducted using R statistical software (R Core Team, 2019). Population factor models and corresponding sample data sets were generated with the simFA function (available in the fungible library; Waller, 2019), which is designed to simulate user-specified factor analysis models. All factor models were estimated using functions available in the fungible library (Waller, 2019) or with the factanal function (R Core Team, 2019).

Results

We first computed an analysis of variance (ANOVA) on the Heywood case prevalence rates using our eight design factors and their two-way interactions. The full results from this analysis are included in the online supplemental materials. Here we focus on the design factors with the greatest effect sizes (any $\eta^2 \geq .02$),[6] because with the large sample sizes in our study virtually all main effects and two-way interactions were statistically significant. Using this criterion, five design factors were associated with meaningful effects: number of indicators per factor (η^2 = .27), model specification (η^2 = .20), sample size (η^2 = .09), factor extraction method (η^2 = .08), and salient loading pattern (η^2 = .06). Moreover, the effect sizes for three two-way interactions also exceeded our threshold: indicators per Factor $\times$ Model Specification (η^2 = .05), indicators per Factor $\times$ Salient Loading pattern (η^2 = .03), and Sample Size $\times$ Factor Extraction Method (η^2 = .03).

Based on the above findings, we created tables of Heywood case prevalence rates in which we marginalized over the three design factors that did not reach practical significance according to our criterion (i.e., number of common factors in the data-generating model, level of model approximation error, and presence of cross-loadings). Table 3 reports the Heywood case prevalence rates from the PAF solutions and Table 4 reports the corresponding rates

[5] Because Heywood cases are often found in factor solutions that have failed to converge (Kano, 1998), we tried to separate any confounding influence from convergence issues by subsetting only converged solutions.

[6] Note that we are using the classical η^2 rather than partial η^2. See Cohen (1973), Levine and Hullett (2002), and Richardson (2011) for a review of the difference between these two statistics.

Table 2
Example Population Salient Loading Patterns Used for the Monte Carlo Simulations

A. Constant high loadings			B. Sequential loadings			C. One strong loading		
94[†]	94	94	86	86	86	86	86	86
80	0	0	80	0	0	80	0	0
80	0	0	63	0	0	50	0	0
80	0	0	47	0	0	50	0	0
80	0	0	30	0	0	50	0	0
0	80	0	0	80	0	0	80	0
0	80	0	0	63	0	0	50	0
0	80	0	0	47	0	0	50	0
0	80	0	0	30	0	0	50	0
0	0	80	0	0	80	0	0	80
0	0	80	0	0	63	0	0	50
0	0	80	0	0	47	0	0	50
0	0	80	0	0	30	0	0	50
D. One weak factor			**E. Two weak factors**			**F. Constant low loadings[‡]**		
94	94	53	94	53	53	53	53	53
80	0	0	80	0	0	30	0	0
80	0	0	80	0	0	30	0	0
80	0	0	80	0	0	30	0	0
80	0	0	80	0	0	30	0	0
0	80	0	0	30	0	0	30	0
0	80	0	0	30	0	0	30	0
0	80	0	0	30	0	0	30	0
0	80	0	0	30	0	0	30	0
0	0	30	0	0	30	0	0	30
0	0	30	0	0	30	0	0	30
0	0	30	0	0	30	0	0	30
0	0	30	0	0	30	0	0	30

Note. Example salient loading patterns are shown with three common factors and four indicators per factor (the number of common factors and indicators per factor were manipulated in Study 1, but were held constant at these levels in Study 2). All loadings and factor score determinacy values multiplied by 100.
[†] Indicates factor score determinacy value. [‡] In Studies 2–4, this salient loading pattern was renamed "Three Weak Factors."

from the ML solutions. These results reveal that Heywood cases are more likely to be observed in models with fewer indicators per factor, with more factors than included in the data-generating model (i.e., overfactoring), when sample sizes are small, when using ML factor extraction, and when factor loading sizes are relatively small. Regarding this latter finding, our results show that Heywood cases are more likely to occur when models include higher proportions of low-communality variables. For example, averaging over the other seven design factors, the probability of observing a Heywood case in our simulations exceeded .20 for the constant low loadings (F), two weak factors (E), and sequential loadings (B) factor patterns. In Study 2, we discuss whether these results are due to the factor loading sizes per se, or whether they are better understood as a function of FSD, which represents the correlation between the true and the estimated factor scores (Guttman, 1955; Steiger, 1994).

Notice in Tables 3 and 4 that one of the strongest influences on Heywood case prevalence rates concerns the issue of model specification. Specifically, our data showed that when EFA models were misspecified by extracting more than the data-generating number of common factors (ignoring the issue of model approximation error), the likelihood of encountering a Heywood case rose drastically. For example, consider the constant high loadings data in Table 4 for the $p = 3$ and $N = 50$ condition. Notice that for these ML solutions, only 11.1% produced one or more Heywood cases when the data-generating number of factors were extracted, but that 79.5% of the solutions included Heywood cases when one too many factors were extracted. The PAF solutions exhibited a similar, but less extreme trend (3.0% vs. 53.3%). Moreover, our aforementioned ANOVA analysis showed that the vitiating effect of a small $p : k$ ratio was strongest in the overfactored models. For instance, Figure S1 in the online supplemental materials shows that the differences in Heywood case prevalence rates between overfactoring and extracting the data-generating number of major common factors were $\{0.29, 0.20, 0.09\}$ for $p : k$ ratios of $\{3 : 1, 5 : 1, 8 : 1\}$.

Our simulation results also corroborate previously-reported findings (De Winter & Dodou, 2012, 2016) that Heywood cases are more likely to occur when using ML extraction versus PAF. In particular, as seen in Figure S2 in the online supplemental materials, we found that the difference in Heywood case prevalence rates between ML and PAF increased with smaller sample sizes. After marginalizing over all other design factors, the differences in the Heywood case prevalence rates for these methods equalled $\{.29, .14, .09, .08\}$ for sample sizes of $N \in \{50, 150, 300, 500\}$. These results suggest that one way to minimize the possibility of encountering a Heywood case in smaller sample sizes is not to use ML estimation. In Study 4, we discuss alternative factor extraction methods that may be better suited for research with small samples.

Table 3
Heywood Case Prevalence Rates Across Model and Data Characteristics When Using Principal Axis Factoring

			Salient loading pattern					
Model specification	$p:k$	N	A	B	C	D	E	F
Correct	3 : 1	50	3.0	51.7	41.1	30.6	44.2	51.9
		150	0.0	41.7	24.8	21.8	36.1	45.9
		300	0.0	33.3	16.2	15.1	28.8	40.6
		500	0.0	27.7	12.1	10.7	22.3	33.6
	5 : 1	50	0.0	6.1	6.2	8.9	14.6	16.1
		150	0.0	0.7	0.9	3.1	6.5	8.2
		300	0.0	0.1	0.1	1.0	2.4	3.4
		500	0.0	0.0	0.0	0.6	1.0	1.6
	8 : 1	50	0.0	0.2	0.3	1.2	1.7	1.7
		150	0.0	0.0	0.0	0.1	0.2	0.3
		300	0.0	0.0	0.0	0.0	0.0	0.0
		500	0.0	0.0	0.0	0.0	0.0	0.0
Under-Factor	3 : 1	50	1.2	23.6	15.9	2.6	16.9	26.6
		150	0.0	17.2	8.2	0.0	10.3	19.6
		300	0.0	13.1	5.6	0.0	7.6	16.1
		500	0.0	10.9	4.1	0.0	4.9	12.6
	5 : 1	50	0.0	1.7	1.4	0.0	3.5	5.4
		150	0.0	0.2	0.1	0.0	0.9	1.9
		300	0.0	0.0	0.0	0.0	0.3	0.6
		500	0.0	0.0	0.0	0.0	0.1	0.3
	8 : 1	50	0.0	0.0	0.0	0.0	0.2	0.2
		150	0.0	0.0	0.0	0.0	0.0	0.0
		300	0.0	0.0	0.0	0.0	0.0	0.0
		500	0.0	0.0	0.0	0.0	0.0	0.0
Over-Factor	3 : 1	50	53.3	75.4	70.1	65.3	70.4	74.4
		150	35.5	64.5	54.6	49.7	59.2	65.7
		300	26.9	54.9	42.6	38.1	49.3	59.9
		500	20.9	46.7	35.7	30.5	41.1	52.5
	5 : 1	50	18.6	22.5	22.7	28.2	32.1	33.5
		150	11.8	12.1	12.6	23.8	22.0	22.7
		300	8.5	9.6	9.7	19.5	18.1	17.6
		500	7.1	8.2	8.0	15.9	15.3	14.5
	8 : 1	50	2.8	2.8	2.8	7.9	6.6	5.8
		150	1.5	1.6	1.2	5.5	3.1	2.6
		300	1.3	1.9	1.2	5.1	2.4	1.6
		500	1.2	2.0	1.4	5.4	2.4	1.6

Note. p = number of indicators; k = number of common factors; N = sample size. Example factor loading patterns are displayed in Table 2. A Heywood case occurred if at least one communality was greater than or equal to 1.00. Heywood proportions marginalized over (a) the number of common factors in the population, (b) level of model approximation error, and (c) presence of cross-loadings. All proportions multiplied by 100.

Study 2: Exploring Understudied Causes of Heywood Cases in EFA

Study 2 was designed to shed additional light on three relatively understudied causes of Heywood cases in EFA. First, we tested whether lower population FSD values exacerbate Heywood case prevalence rates. FSD is defined as the correlation between true and (least-squares) estimated factor scores (Guttman, 1955), and arises because an infinite set of mathematically (but not substantively) exchangeable factor scores will perfectly fit a given factor solution. A higher degree of FSD indicates that all members in this infinite set are more similar to one another (Grice, 2001; Guttman, 1955; Mulaik, 2010; Steiger, 1994). Our review of the literature and our Study 1 simulation results indicated that factor loading strength, the indicator-to-factor ratio, and salient loading patterns—all correlates of FSD—influence Heywood case prevalence rates (Anderson & Gerbing, 1984; Boomsma, 1985; De Winter et al., 2009; De Winter & Dodou, 2012; Gerbing & Anderson, 1987; Jung & Takane, 2008; MacCallum et al., 1999). For example, there is evidence that adding factor indicators (of any magnitude) to a previously weak factor can reduce the likelihood of encountering a Heywood case (De Winter et al., 2009; Van Driel, 1978). However, no study to our knowledge has considered whether (low) FSD is a direct cause of Heywood cases.

Furthermore, like previous simulations exploring Heywood case prevalence rates in EFA (e.g., Briggs & MacCallum, 2003), our Study 1 design focused exclusively on orthogonal population EFA models. However, oblique models (i.e., with correlated common factors) are frequently used in the social sciences (Fabrigar et al., 1999) and are often recommended over orthogonal models in EFA (Browne, 2001). We thus expanded the generalizability of our Study 1 results by comparing Heywood case prevalence rates when systematically increasing common factor correlation (ϕ_{ij}) magnitudes.

Our final set of simulations in Study 2 concerns nonpositive definite (NPD) correlation matrices and matrix smoothing. To our

Table 4
Heywood Case Prevalence Rates Across Model and Data Characteristics When Using Maximum Likelihood Factoring

			Salient loading pattern					
Model specification	$p : k$	N	A	B	C	D	E	F
Correct	3 : 1	50	11.1	79.8	70.6	58.0	74.3	81.0
		150	0.0	65.5	43.7	37.5	57.9	70.9
		300	0.0	52.6	27.0	26.1	45.7	62.5
		500	0.0	43.1	20.1	18.4	36.3	54.4
	5 : 1	50	0.0	41.1	40.9	39.3	63.7	77.4
		150	0.0	5.7	6.3	13.7	28.5	44.6
		300	0.0	1.3	1.5	4.5	11.1	22.0
		500	0.0	0.4	0.5	1.6	4.3	10.8
	8 : 1	50	0.0	6.7	13.1	21.9	44.5	61.7
		150	0.0	0.0	0.1	2.4	7.7	15.6
		300	0.0	0.0	0.0	0.1	0.9	3.2
		500	0.0	0.0	0.0	0.0	0.3	1.1
Under-Factor	3 : 1	50	6.6	62.2	49.0	10.3	48.2	65.0
		150	0.1	47.5	26.5	0.2	26.8	48.7
		300	0.0	36.5	16.2	0.0	17.5	38.5
		500	0.0	30.5	12.5	0.0	11.6	31.1
	5 : 1	50	0.0	23.3	21.7	0.1	32.2	54.8
		150	0.0	3.5	3.2	0.0	8.6	22.7
		300	0.0	0.8	0.6	0.0	2.5	8.5
		500	0.0	0.4	0.2	0.0	0.6	3.6
	8 : 1	50	0.0	2.5	4.9	0.0	17.2	36.3
		150	0.0	0.0	0.0	0.0	1.4	5.1
		300	0.0	0.0	0.0	0.0	0.2	0.9
		500	0.0	0.0	0.0	0.0	0.0	0.3
Over-Factor	3 : 1	50	79.5	90.9	88.5	85.9	88.8	90.4
		150	63.1	84.8	79.3	73.9	80.4	85.3
		300	54.8	78.3	70.3	65.7	72.9	81.8
		500	50.1	72.6	64.5	59.8	67.7	77.8
	5 : 1	50	55.4	74.6	75.0	73.6	83.3	89.9
		150	30.3	38.0	38.6	43.8	53.9	67.2
		300	20.9	24.3	24.9	29.5	35.6	48.4
		500	16.7	18.8	19.4	22.5	26.8	37.2
	8 : 1	50	31.4	42.1	47.2	52.4	69.3	81.5
		150	11.7	13.2	13.3	18.1	26.7	37.0
		300	7.5	8.1	8.0	9.9	12.7	16.4
		500	6.3	6.8	6.7	7.4	9.1	10.7

Note. p = number of indicators; k = number of common factors; N = sample size. Example factor loading patterns are displayed in Table 2. A Heywood case occurred if at least one communality was greater than or equal to .998. Heywood proportions marginalized over (a) the number of common factors in the population, (b) level of model approximation error, and (c) presence of cross-loadings. All proportions multiplied by 100.

knowledge, this topic has not received much, if any, attention in the extant Heywood case literature. Although some researchers have proposed NPD correlation matrices as potential causes of Heywood cases in structural equation models (Kolenikov & Bollen, 2012; Wothke, 1993), we are unaware of any studies that have investigated whether matrix smoothing (e.g., Bentler & Yuan, 2011; Higham, 2002; Knol & Berger, 1991) influences this relationship, particularly within the context of EFA.

Method

In total, Study 2 included three independent simulations with model features informed by the trends uncovered in Study 1. All population models included three common factors using the six salient loading patterns presented in Table 2. Note that for the remaining studies, we renamed the constant low loadings pattern to three weak factors. We did not incorporate cross-loadings or vary the number of common factors in the population models, as these design factors demonstrated negligible effects on Heywood case prevalence rates in Study 1. Moreover, for the Study 2 simulations, all sample data sets were limited to 100 subjects to ensure that each simulation included a sufficient number of data sets that would generate one or more Heywood cases.

Our first simulation focused on the extent to which FSD influences Heywood case prevalence rates. To investigate this issue, we systematically varied the FSD values of factors in different population models by incrementally adding a small constant (of equal size) to each factor loading of a designated factor pattern. Specifically, we began our simulations with three one-factor population models, each of which included four factor indicators. Using previously-defined terminology, the three models were: constant low loadings ($\boldsymbol{\lambda}_{\text{ConstLow}} = \{0.35, 0.35, 0.35, 0.35\}$), sequential loadings ($\boldsymbol{\lambda}_{\text{Seq}} = \{0.45, 0.35, 0.30, 0.25\}$), and one stronger loading ($\boldsymbol{\lambda}_{\text{OneStronger}} = \{0.45, 0.30, 0.30, 0.30\}$). These loading patterns were specifically selected to have similar FSD values for the population-level factors, which were calculated via Guttman's (1955) formula,

$$\widehat{\boldsymbol{\rho}} = \sqrt{\mathrm{diag}(\mathbf{F}_S^T\mathbf{R}_{\mathrm{YY}}^{-1}\mathbf{F}_S)} \tag{4}$$

where $\widehat{\boldsymbol{\rho}}$ is a $k \times 1$ vector of determinacy values for k common factors, $\mathbf{F}_S$ is the population factor structure matrix, and $\mathbf{R}_{\mathrm{YY}}$ is the population correlation matrix for the factor indicators.[7] Each common factor for the three loading patterns began with an FSD of approximately .60. To increase the FSD value for the population-level factor, we then added a small constant value of .03 to each of the four factor loadings. This step was repeated nine times. For example, the constant low factor loading pattern at the ninth iteration was {.62, .62, .62, .62}. Thus, we examined 10 increasing levels of FSD for each one-factor, four-indicator population model. The factor loading patterns and constant value were chosen such that the factors in the three population models would have approximately equal FSD values at each iteration.

Next, we explored the likelihood of encountering a Heywood case in oblique population factor models. We studied three factor correlation sizes ($\phi_{ij} = \{0.0, 0.3, 0.8\}$ where all factor correlations were equal in a given model) with each of the aforementioned six salient loading patterns (see Table 2). We selected these ϕ_{ij} values to encapsulate the range of factor correlations that are typically seen in psychological research (e.g., the correlations between common factors of intelligence can exceed .70; Canivez & Watkins, 2010; Ryan et al., 2003). To further enhance the generalizability of our findings, we compared Heywood case prevalence rates across increasing factor correlation sizes in two sets of models: (a) models with no model approximation error, and (b) models with low levels of model approximation error (Briggs & MacCallum, 2003; Tucker et al., 1969). As in Study 1, low model approximation error was defined as a population model with a RMSEA value between .00 and .05 (Browne & Cudeck, 1992; Marsh et al., 2004). To generate these models, for each variable, we reallocated 10% of the unique factor variance to model approximation error. As in Study 1, we used a $\mathbf{W}$ matrix with 150 minor factors and the aforementioned constraints on minor factor loading size (the ϵ parameter was set to .20 in these analyses).

Our final set of simulations in Study 2 explored the influence of matrix smoothing algorithms on Heywood case prevalence rates. To investigate this topic, we generated sample data corresponding to one of the six salient loading patterns. In each sample data matrix, each variable had approximately 25% missing data (specified to be missing completely at random; see Enders, 2010). Next, we computed NPD correlation matrices from these incomplete data sets using pairwise deletion (Wothke, 1993). To obtain sufficiently large sets of NPD matrices without using unrealistically high missingness rates, we generated sample data from models with six indicators per factor and either 10% or 30% of the unique factor variance reallocated to model approximation error (Tucker et al., 1969). The data were generated such that each population model of interest produced the same number of NPD correlation matrices. Next, we smoothed each NPD matrix by each of three matrix smoothing algorithms: (a) the alternating projection algorithm (APA; Higham, 2002); (b) the Bentler and Yuan (BY; 2011) algorithm; and (C) the Knol and Berger (KB; 1991) algorithm. We also included a "no smoothing" condition to obtain baseline data for the Heywood case frequencies.

For each of the aforementioned models across the three simulation studies, we generated 1,000 data sets (based on a multivariate normal distribution) and extracted the model-implied number of common factors (ignoring the presence of minor factors, if applicable) using iterative PAF. As in Study 1, only converged factor solutions were retained for further analyses. For each sample, we recorded whether at least one Heywood case was present (for these analyses, a Heywood case was defined as an estimated communality greater than or equal to 1.00). All simulations used the simFA function (Waller, 2019) in R (R Core Team, 2019) to generate the population factor models and sample data sets. The matrix smoothing and factor analysis routines were then conducted using the R fungible library (Waller, 2019).

Results

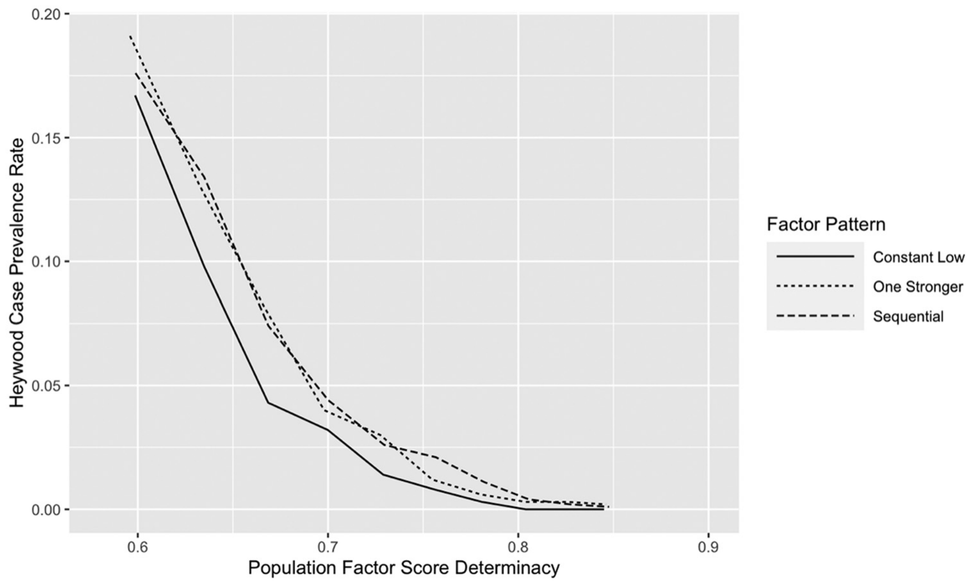
Our first set of simulations showed that the FSD value for the population-level common factor is inversely related to the likelihood of observing a Heywood case. As seen in Figure 1, when FSD values are high, the likelihood of observing a Heywood case is low. Relatedly, when FSD values are relatively low—for example, FSD $\leq$ .70—the likelihood of observing one or more Heywood cases is non-negligible. Our results also suggest that when a common factor is not well-defined (i.e., FSD $<$ .75), the factor loading pattern may moderate the relationship between FSD and Heywood case prevalence rates. For instance, notice that when the constant low and sequential loadings models both had FSD values of approximately .65, the latter model produced slightly more factor solutions with Heywood cases. However, it is important to note that the differences in Heywood case prevalence rates among the factor pattern conditions never exceeded 5%. Although admittedly limited, our findings provide the first evidence that FSD values in the data-generating model play a causal role in the production of Heywood cases in EFA models. These results point to the importance of working with well-defined factors with high determinacy values if one wishes to avoid improper estimates in EFA models.

We next examined the effects of correlated common factors on Heywood case prevalence rates. The results from these analyses are summarized in Table 5. Corroborating previous research in exploratory (De Winter et al., 2009; De Winter & Dodou, 2012) and confirmatory factor analysis (Rindskopf, 1984), our results indicate that Heywood cases were more likely to occur among models with stronger interfactor correlations. For example, in the one weak factor (D) condition, the proportion of solutions with at least one Heywood case increased from 18.3% to 30% as the factor correlation magnitude (ϕ_{ij}) increased from .00 to .80. Importantly, our results also indicate that the population factor loading pattern moderated the relationship between factor correlation size and Heywood case prevalence rates. Notice that irrespective of ϕ_{ij}, Heywood cases were more likely to occur among models with higher proportions of low-communality indicators (e.g., Patterns D–F). Table 5 also highlights that the proportional increases in Heywood case prevalence rates differed among the factor pattern conditions. As ϕ_{ij} increased from .00 to .80 among models without approximation error, the Heywood case prevalence rates increased

[7] Equation 4 assumes that $\mathbf{F}_S$ is a matrix of correlations. When the elements of $\mathbf{F}_S$ are not correlations (e.g., when the solution contains one or more Heywood cases), this formula yields meaningless FSD estimates (which we will see in Studies 3 and 4).

Figure 1

Observed Heywood Case Prevalence Rates at Varying Levels of Factor Score Determinacy



by 23.1% and 19.7% for the sequential (B) and one strong loading (C) patterns, respectively. Alternatively, the proportional increase among Patterns D–F was at most 11.7%. Integrating these trends, our simulation suggests that differences in Heywood case prevalence rates among the examined factor patterns decrease as the magnitude of ϕ_{ij} increases in the population model.

Table 5 also indicates that model approximation error may moderate the relationship between the magnitude of the interfactor correlation and the likelihood of a encountering a Heywood case. Comparing the right- and left-hand sides of this table, the proportional increases in Heywood case prevalence rates were generally smaller when models included model approximation error (defined as population EFA models with RMSEA values between .00 and .05; Browne & Cudeck, 1992; Marsh et al., 2004). For example, for the two weak factors (E) condition, the Heywood case prevalence rates increased from 28.5 to 40.2 among models without approximation error (a difference of 11.7). On the other hand, the Heywood case prevalence rates increased from 30.0 to 36.9 among models with approximation error (a difference of 6.9). Additionally, our findings suggest that across all magnitudes of ϕ_{ij}, incorporating model approximation error reduced the average Heywood case prevalence rates in conditions with weak factors (Patterns D–F).

Our final set of simulations in Study 2 focused on Heywood case prevalence rates in factor models computed from (possibly smoothed) NPD correlation matrices that were generated from incomplete data. Table 6 presents the main findings from these analyses. We found that when compared with the baseline no-smoothing conditions, the Heywood case prevalence rates generally decreased (and never increased) after applying any of the three matrix smoothing algorithms. Moreover, we also found that the effects of smoothing varied by method. Our data showed that among the three matrix smoothers, the BY algorithm (Bentler & Yuan, 2011) provided the greatest protection against Heywood cases. For example, in data sets under the low model error (.1) conditions, the average Heywood case prevalence rate for the BY algorithm was approximately half the size of that for the other smoothing algorithms (including the no-smoothing condition). Similarly, for the moderate model error conditions, the Heywood case prevalence rate for the BY algorithm was approximately 35%–40% as large as the other algorithms.

From these results, one might conclude that the BY algorithm is an obvious choice for smoothing an NPD correlation matrix prior to conducting an exploratory factor analysis. However, that conclusion fails to consider how each smoothing algorithm affects the accuracy of the estimated factor loadings. To explore this additional issue, we aligned[8] each of the estimated factor patterns (rotated with direct Quartimin using the GPArotation R library; Bernaards & Jennrich, 2005) with its (error free) population counterpart. We next computed root mean square error (RMSE) values between the (aligned) sample factor patterns and the population factor patterns. In summary, we found no meaningful differences in the RMSE values among the three smoothing algorithms. Specifically, across the three methods, the average RMSE values never differed by more than .001 (the full findings from this analysis are presented in Table S2 of the online supplemental materials). Thus, although we found evidence that the BY algorithm reduces the probability of encountering a Heywood case when factor analyzing an NPD correlation matrix, this benefit did not translate into improved factor loading recovery for our examined models.

[8] Alignment was carried out with the faAlign function from the R fungible package (Waller, 2019). This function minimizes the root mean squared deviation between a population and sample pattern matrix.

Table 5
Sample Heywood Case Prevalence Rates Across Increasing Factor Correlation Sizes

	No model approximation error			Low model approximation error		
Loading pattern	$\Phi_{ij} = 0.0$	$\Phi_{ij} = 0.3$	$\Phi_{ij} = 0.8$	$\Phi_{ij} = 0.0$	$\Phi_{ij} = 0.3$	$\Phi_{ij} = 0.8$
A	0.0	0.0	0.3	0.0	0.0	0.4
B	7.7	9.8	30.8	10.7	11.8	29.1
C	7.2	8.2	26.9	10.1	10.4	25.7
D	18.3	21.5	30.0	16.1	20.1	24.1
E	28.5	34.2	40.2	30.0	30.4	36.9
F	36.8	37.8	40.9	33.7	32.8	33.3

Note. ϕ_{ij} indicates the common correlation among the three major factors in the population model. Model error introduced by allocating 10% of the unique factor variance to model approximation error (i.e., 150 minor factors not captured by the three major common factors). Population models with error generated to have root mean square error of approximation values between .00 and .05. All population models generated with four indicators per factor and 100 subjects per sample data matrix. Six salient loading patterns: A = constant high loadings; B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Factor analysis was conducted using principal axis factoring. All proportions multiplied by 100.

Study 3: The Effects of Heywood Cases on EFA Model Recovery

In our previous two studies, we reviewed several model and data characteristics that are associated with Heywood cases in exploratory factor analysis (e.g., factor score determinacy, ML parameter estimation, smoothing algorithms for nonpositive definite matrices). In many cases, researchers can spot if a Heywood case is present in their analyses, and in some cases, software packages (e.g., psych in R; Revelle, 2018) warn users when communalities are greater than one. However, sometimes researchers fail to identify Heywood cases or are unable to remove them by adding salient indicators, increasing sample size, or by using a different factor extraction method. These scenarios give rise to the following question: How do Heywood cases influence our interpretations of EFA output?

Prior research in this area has mainly considered the effects of Heywood cases on estimated factor loadings. For instance, in the context of CFA, several researchers have noted that Heywood cases can bias factor loadings in complicated ways (Boomsma, 1985; Chen et al., 2001; Gerbing & Anderson, 1987; Marsh et al., 1998). Moreover, some evidence suggests that the degree and sign of the loading bias depends on the factor loading's location relative to other loading(s) associated with the Heywood case (Gerbing & Anderson, 1987; Marsh et al., 1998). In an EFA context, other researchers (e.g., De Winter et al., 2009; MacCallum et al., 1999; MacCallum et al., 2001; Preacher & MacCallum, 2002) have failed to find substantial differences in factor loading recovery between solutions with and without Heywood cases (however, see Briggs & Mac-Callum, 2003). To build upon this work, and to help resolve these discrepant findings, we examined whether and how Heywood cases affect various aspects of EFA model recovery in addition to factor loading bias. In particular, we considered how Heywood cases potentially bias the estimated (a) dimensionality of the EFA model, (b) factor loadings and correlations, and (c) factor score determinacy (FSD) values.

Table 6
Heywood Case Prevalence Rates When Factor Analyzing Smoothed and Unsmoothed Nonpositive Definite Correlation Matrices Resulting From Missing Data

		Smoothing method			
Model error	Loading pattern	No smoothing	APA	BY	KB
0.1	A	0.0	0.0	0.0	0.0
	B	9.2	8.0	1.3	7.5
	C	13.7	11.9	1.8	11.2
	D	8.9	8.9	8.0	8.6
	E	18.8	16.2	11.6	16.2
	F	28.0	24.8	13.2	24.8
	Column average	13.1	11.6	6.0	11.4
0.3	A	0.0	0.0	0.0	0.0
	B	9.8	8.5	2.0	8.0
	C	11.8	10.7	2.7	10.4
	D	5.4	4.6	2.7	4.2
	E	7.5	5.8	3.4	5.6
	F	16.3	14.7	6.6	14.5
	Column average	8.5	7.4	2.9	7.1

Note. APA = alternating projection algorithm; BY = Bentler & Yuan; KB = Knol & Berger. Model approximation error introduced by reallocating a specified proportion of unique factor variance to error. All population models generated with six indicators per factor and 100 subjects per sample data matrix. Six salient loading patterns: A = constant high loadings; B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Missing data values were generated from a "missing completely at random" (MCAR) model where the probability of a missing value for each variable was .25. Factor extraction was conducted using principal axis factoring. All proportions multiplied by 100.

Method

The aforementioned results on causes of Heywood case in EFA informed the design of our Study 3 population factor models. For instance, we did not incorporate cross-loadings, a design factor which showed negligible effects in Study 1, and we again used the factor pattern loadings presented in Table 2. However, we excluded the constant high loadings (A) pattern because it produced few Heywood cases in our previous simulations. All models had three common factors and four salient indicators per factor, and each sample data set had 150 subjects. Although oblique factor models were found to exacerbate Heywood case prevalence rates in Study 2, all population models in Study 3 were orthogonal to focus the scope of our analyses. Furthermore, all sample data sets were factor-analyzed with principal-axis factoring (PAF), rotated with direct Quartimin (Bernaards & Jennrich, 2005), and aligned to the corresponding population model using an alignment algorithm that minimizes a least squares function of the target and aligned factor loadings matrices (see the faAlign function in the R fungible package; Waller, 2019).

To compare EFA solutions with and without a Heywood case, we generated four solution sets, termed (a) "Heywood solutions without error," (b) "Heywood solutions with low error," (c) "non-Heywood solutions without error," and (d) "non-Heywood solutions with low error." Each solution set had 1,000 sample data matrices (generated from multivariate normal distributions to fit the given population model). In the two Heywood solution sets, all sample data matrices produced at least one Heywood case when factor-analyzed with PAF. Alternatively, in the two non-Heywood solution sets, no sample data matrices produced a Heywood case. In the solutions with low levels of error, population models were generated such that 10% of each unique factor variance was reallocated to model approximation error (Briggs & MacCallum, 2003; Tucker et al., 1969). As in Studies 1 and 2, models with low levels of approximation error produced RMSEA values between .00 and .05 (Browne & Cudeck, 1992; Marsh et al., 2004). Moreover, only samples that produced converged PAF solutions were retained for further analysis.

Our first analysis in Study 3 focused on dimensionality bias of the common factor space within Heywood and non-Heywood solutions. We used a traditional parallel analysis (Horn, 1965; Lim & Jahng, 2019) to estimate the essential dimensionality of the sample data matrices. Specifically, for each sample correlation matrix, we simulated 1,000 sets of ordered eigenvalues from random data matrices that were produced by permuting the scores (separately for each column) of the sample data matrices (this method retains the original data [univariate] distributions but destroys the correlational structure).[9] The suggested number of common factors was defined as the number of consecutive observed eigenvalues greater than the 95th quantile of the corresponding simulated eigenvalue distributions. Note that, according to this definition, if the first, second, and fourth observed eigenvalues exceeded the quantiles of the simulated distributions, but not the third observed eigenvalue, then the parallel analysis retained only two common factors. We calculated dimensionality bias by subtracting the number of common factors in the known population factor model from the number of factors suggested by the parallel analysis.

Next, to measure factor loading recovery, we calculated RMSEs between each (aligned) estimated factor pattern and the corresponding population factor pattern. Specifically, we calculated two sets of RMSE values for each sample solution: (a) for the full factor-pattern matrix ("full-model") and (b) for each factor loading ("element-wise"). Next, we subtracted each population factor loading from its associated estimate to compute element-wise factor loading bias. To compute the bias of the estimated factor correlations, we similarly subtracted each population factor correlation matrix—which was a 3×3 identity matrix for all of our models—from the corresponding estimated factor correlation matrix. Finally, to compute FSD bias, we estimated FSD values for each common factor (see Equation 4), and subtracted the population FSD values from their sample estimates.[10]

As before, we used R statistical software (R Core Team, 2019) for all analyses. We used the fungible package (Waller, 2019) to generate population EFA models and corresponding sample data sets, as well as for all factor analysis routines. The fungible package implements the GPArotation package (Bernaards & Jennrich, 2005) for factor rotation algorithms.

Results

In our first set of analyses, we examined whether parallel analysis provides accurate dimensionality recommendations in data sets that produce one or more Heywood cases when factor analyzed by PAF. Table 7 presents the average estimated dimensionality and corresponding bias for the data sets associated with (or without) Heywood cases. Three trends from these results merit comment. First, within each salient loading pattern and level of model approximation error, data sets that produced or did not produce Heywood cases demonstrated bias in the same direction. In other words, if the estimated dimensionality was biased downward in the Heywood solution set, it was also biased downward in the non-Heywood solution set. This trend was most prominent in the three salient loading patterns with at least one weak factor: one weak factor (D), two weak factors (E), and three weak factors (F). Our results thus align with those from Lim and Jahng (2019), where parallel analysis produced less accurate dimensionality estimates when models included higher proportions of low-communality indicators.

The second trend shown in Table 7 relates to differences in estimated dimensionality between data matrices comprising the Heywood and non-Heywood solution sets. Namely, our results show that across both levels of model approximation error, the downward bias in dimensionality estimates was slightly larger for the data matrices from the Heywood solution set compared to those from the non-Heywood solution set. For instance, in the two weak factors condition, parallel analysis more often suggested extracting a single component in the Heywood solution set than in the non-Heywood solution set. Although the bias differences were small (e.g., the bias differences between the associated Heywood and Non-Heywood solutions sets were at most .37), this trend was consistent for the one, two, and three weak factors condition. To better understand this trend, we compared the distributions of the first

[9] See Horn (1965), Hayton et al. (2004), and Reise et al. (2000) for more comprehensive reviews of the parallel analysis method.

[10] Note that when estimating FSD for a sample-level common factor, $\mathbf{R}_{YY}$ in Equation 4 is replaced by the sample correlation matrix.

Table 7
Estimated Dimensionality From Parallel Analysis in Data Sets That Do or Do Not Produce Heywood Cases

	Heywood solutions				Non-Heywood solutions			
Loading pattern	*M* (*SD*)	Range	Bias	RMSE	*M* (*SD*)	Range	Bias	RMSE
No model error								
B	3.03 (0.17)	2–4	0.03	0.18	3.01 (0.10)	2–4	0.01	0.10
C	3.00 (0.08)	2–4	0.00	0.08	3.00 (0.03)	2–3	0.00	0.03
D	2.07 (0.26)	2–3	−0.93	0.96	2.44 (0.50)	2–3	−0.56	0.75
E	1.54 (0.60)	1–4	−1.46	1.58	1.86 (0.78)	1–4	−1.14	1.38
F	0.50 (0.80)	0–4	−2.50	2.62	0.74 (1.00)	0–4	−2.26	2.47
Low model error								
B	3.09 (0.30)	2−5	0.09	0.31	3.04 (0.21)	3−4	0.04	0.21
C	3.03 (0.18)	2−4	0.02	0.18	3.01 (0.10)	2−4	0.01	0.10
D	2.15 (0.36)	2−4	−0.85	0.92	2.47 (0.50)	2−3	−0.53	0.73
E	1.71 (0.66)	1−4	−1.29	1.45	2.09 (0.78)	1−4	−0.91	1.20
F	1.02 (1.03)	0−5	−1.98	2.23	1.36 (1.25)	0−5	−1.64	2.06

Note. *M* = mean; *SD* = standard deviation; RMSE = root mean square error. Parallel analysis calculated by simulating eigenvalues after randomly permuting the columns of the raw data matrix. Dimensionality estimated with the number of observed eigenvalues greater than the 95% percentile of the distribution of simulated eigenvalues. Bias calculated by subtracting the number of common factors in the population (i.e., three factors) from the number of factors suggested by parallel analysis, and averaged across all sample matrices (1,000 in each solution set). Model error introduced by allocating 10% of the unique factor variance to model approximation error (i.e., 150 minor factors not captured by the three major common factors). Population models with error generated to have root mean square error of approximation values between .00 and .05. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. All population models had four indicators per common factor and 150 subjects in each sample data matrix.

three observed eigenvalues for the Heywood and non-Heywood correlation matrices. We also examined the distributions of the 95th quantile for the set of permuted eigenvalues computed in each parallel analysis. Figure S3 presents histograms for these distributions from the two weak factors condition.

Scanning these histograms, we found negligible differences in the 95th quantile for the three comparison eigenvalue distributions between the Heywood and non-Heywood solution sets. However, more distinct differences emerged in the observed eigenvalue distributions. First, notice in the second column of Figure S3 that the first eigenvalue in either the so-called Heywood or non-Heywood correlation matrices was consistently larger than the average 95th quantile of the corresponding permuted eigenvalue distribution. However, the distribution of the first eigenvalues for the Heywood matrices demonstrated higher kurtosis and greater left skewness, indicating that the first eigenvalue for the Heywood matrices were more frequently larger than the first eigenvalue for the non-Heywood matrices. Given that the eigenvalues for a particular correlation matrix must sum to a constant, it follows that the second and third observed eigenvalues tended to be smaller in the Heywood correlation matrices as compared with the non-Heywood matrices. Indeed, descriptive statistics revealed larger right skewness among the distributions of the second and third eigenvalues in the Heywood matrices. Additionally, correlations among the three observed eigenvalues in the Heywood and non-Heywood correlation matrices also indicated a slightly larger negative relationship between the first and third eigenvalues in the Heywood matrices ($r_{13} = -.101$) compared to the non-Heywood matrices ($r_{13} = -.064$; the correlations between the first and second eigenvalues were negligible in magnitude). Integrating these findings, we posit that these distributional shifts in observed eigenvalues explain why parallel analysis more frequently suggested retaining only one component in the Heywood solutions with two weak factors as compared to the non-Heywood solutions. Still, we stress that both the differences in observed eigenvalues and the biases in parallel analysis were relatively small, and necessitate further exploration. At this stage, our data suggest that inaccurate parallel analysis results are more strongly a function of the population factor loading pattern rather than whether a Heywood case is produced in the estimated factor model.

The third major trend evident in our dimensionality analyses concerns the inclusion of model approximation error (Briggs & MacCallum, 2003; Tucker et al., 1969). Looking only at data sets with model approximation error, the bottom half of Table 7 shows that solutions with weak factors and Heywood cases also demonstrated worsened dimensionality recovery. Comparing among data sets with and without model approximation error, notice that for the sequential (B) and one strong loading (C) factor patterns, the parallel analysis provided slightly less accurate dimensionality estimates when model error was included. For example, the RMSE for the sequential loadings condition in the Heywood solution set increased from .18 to .31 when 10% of the unique factor variance was reallocated to approximation error. Alternatively, for the remaining three factor patterns (D–F), dimensionality recovery was slightly improved in models with approximation error. Aligning with the findings from Study 2, our results highlight the ways in which model approximation error can moderate the recovery of EFA solutions in the presence of Heywood cases.

We next explored factor loading and factor correlation recovery in data sets that produce Heywood cases. Interestingly, although the inclusion of model approximation error influenced the dimensionality findings reported above, the remaining results in Study 3 showed negligible differences between the two solution sets with and without model approximation error. Therefore, below we present the recovery results only for models without model approximation error. The corresponding results for models with model approximation error are available in the online supplemental materials (Tables S3–S10).

First comparing full-model factor loading recovery, we found that average RMSE values were consistently higher for EFA solutions with a Heywood case compared to solutions without a Heywood case. As summarized in Table S3 in the online supplemental materials, the three models with weak factors produced larger differences in average RMSE values ($RMSE_{\Delta}$; one weak factor $RMSE_{\Delta} = 0.13$; two weak factors $RMSE_{\Delta} = 0.11$; three weak factors $RMSE_{\Delta} = 0.09$) than the other two loading patterns (sequential loadings $RMSE_{\Delta} = 0.01$; one strong loading $RMSE_{\Delta} = 0.01$). This finding supports research from Briggs and MacCallum (2003), who reported worsened factor loading recovery for EFA models with weak factors.

When examining factor loading recovery for just the salient (i.e., nonzero) population loadings, our results again showed that the average RMSE values were higher for solutions with a Heywood case compared with solutions without a Heywood case. Our results also showed that the difference between the full-model and salient-loading RMSE values were larger for the estimated models in the Heywood solution set. For example, when subsetting the salient loadings in the Heywood solution set for the three weak factors (F) condition, the average RMSE increased from .27 to .41. Additionally, in both the sequential (B) and one strong loading (C) conditions, the average RMSE increased from .09 to .12. Among the non-Heywood solutions, across all population-level factor loading patterns (B-F), the average RMSE increased by at most .03 when comparing the salient loadings to all loadings.

We next examined patterns of factor loading recovery by analyzing element-wise RMSE values for each factor-pattern matrix. To collate these results, we calculated the RMSE values for each factor loading separately for the Heywood and non-Heywood solution sets. We then computed the difference between these two values for each element in the factor-pattern matrix. These results are presented in Table 8, where a positive value indicates a higher RMSE value for the Heywood solution set. Scanning these results, we can see that the largest differences in RMSE values occurred on the salient population loadings (i.e., loadings $\geq |.30|$ in the population). In some cases, the RMSE values differed by up to .50, indicating that Heywood cases resulted in significantly worsened factor loading recovery. Interestingly, the factor-pattern elements with nonsalient (i.e., zero) population loadings showed no substantial recovery differences between the Heywood and non-Heywood solutions.

Because RMSE values reflect both estimator variance and squared bias, we explored which of these components more strongly influenced the RMSE values for the estimated factor patterns in the Heywood solution set. Table 9 presents the RMSE and bias value (in parentheses) for each factor loading. Again, notice that the salient population factor loadings produced larger RMSE values than the null population factor loadings. Moreover, the RMSE values were often larger for small salient population factor loadings (e.g., $\lambda = .3$) as compared with large loadings (e.g., $\lambda = .8$). The bias magnitudes were generally small (for example, bias $\times 100 < 10$), although we found a few elevated values across the elements (for example, see f_3 for salient loading, Pattern D). Looking closer at these results, we determined that the lion's share of the RMSE values was due to the sampling variances in the estimated factor loadings (for example, in Table 10, the largest $\text{bias}^2/\text{MSE} = 0.24$, where MSE denotes mean squared error). For reference, we also calculated bias values for the non-Heywood solution set (see Table S6 in the online supplemental materials). Overall, in the non-Heywood solutions the loadings bias was negligible and never exceeded $|0.03|$. Moreover, across the five salient loading patterns, the average bias (within a pattern) among factor loadings with salient population values was at most $|0.02|$. Among the non-Heywood solution set, the largest $\text{bias}^2/\text{MSE} = 0.03$.

Another notable trend concerning the bias in estimated factor loadings can be seen in the sequential loadings and one strong loading conditions results in Table 9. Namely, we found that when the loading for the first salient indicator of a factor was biased upward (presumably due to Heywood cases), the other salient loadings on that factor were biased downward. This trend supports findings from past CFA researchers (Dillon et al., 1987; Gerbing & Anderson, 1987) on the behavior of factor loadings in the presence of Heywood cases.

Table 8

Differences in Element-Wise Root Mean Square Error Between Heywood and Non-Heywood Solutions

B			C			D			E			F		
f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3
8	−1	−1	**9**	−1	−1	**1**	0	0	**1**	−1	−1	**21**	−1	−1
4	0	0	**2**	0	0	**2**	0	1	**0**	−2	−1	**22**	−1	−1
2	0	0	**3**	0	0	**1**	0	0	**0**	−2	0	**21**	−1	−2
0	−1	−1	**3**	0	0	**1**	0	−1	**0**	−1	−2	**23**	1	0
−1	**9**	−1	−1	**9**	−1	0	**2**	3	0	**31**	0	−1	**19**	1
0	**3**	0	0	**2**	0	0	**−1**	−4	0	**28**	1	0	**24**	−2
0	**3**	0	0	**2**	−1	0	**0**	−4	0	**32**	0	−1	**18**	0
−1	**0**	0	1	**2**	0	0	**0**	−4	1	**30**	1	2	**18**	0
−1	−1	**9**	−1	−1	**8**	0	0	**47**	0	−2	**25**	1	2	**21**
0	0	**4**	0	0	**2**	0	0	**47**	0	−1	**31**	1	−2	**20**
0	−1	**3**	0	0	**2**	0	0	**47**	0	0	**26**	0	0	**20**
0	−1	**1**	0	0	**2**	0	0	**50**	0	−1	**28**	1	−1	**23**

Note. f_1 = First common factor. Element-wise root mean square error (RMSE) computed across 1,000 Heywood and non-Heywood solutions, and the non-Heywood RMSE values were subtracted from the average Heywood RMSE values. Factor analysis was conducted using principal axis factoring and Quartimin rotation. All population models had four indicators per common factor, 150 subjects in each sample data matrix, and did not include model approximation error. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Salient (i.e., nonzero) population loadings are bolded. All differences multiplied by 100.

Table 9
Root Mean Square Error and Bias for Factor Solutions With at Least One Heywood Case

B			C			D			E			F		
f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3	f_1	f_2	f_3
17 (7)	3 (0)	3 (0)	**17 (8)**	3 (0)	3 (0)	**5 (0)**	5 (0)	9 (0)	**5 (0)**	6 (0)	7 (0)	**42 (1)**	15 (0)	15 (0)
12 (−4)	7 (0)	7 (0)	**11 (−4)**	8 (0)	8 (0)	**6 (0)**	5 (0)	9 (0)	**4 (0)**	6 (0)	6 (0)	**43 (2)**	15 (0)	15 (−1)
10 (−4)	8 (0)	8 (1)	**11 (−4)**	8 (0)	8 (0)	**5 (0)**	5 (0)	8 (0)	**5 (0)**	6 (0)	7 (0)	**42 (1)**	15 (0)	14 (0)
9 (−2)	9 (0)	9 (0)	**11 (−4)**	8 (0)	8 (0)	**6 (0)**	5 (0)	7 (0)	**4 (0)**	6 (0)	6 (0)	**44 (2)**	16 (0)	15 (0)
3 (0)	**18 (8)**	3 (0)	3 (0)	**16 (7)**	3 (0)	5 (0)	**6 (−1)**	11 (−1)	8 (0)	**51 (4)**	16 (0)	14 (0)	**41 (1)**	15 (0)
7 (0)	**12 (−4)**	7 (0)	8 (0)	**11 (−4)**	8 (0)	5 (0)	**4 (0)**	4 (0)	8 (0)	**49 (6)**	16 (0)	15 (0)	**45 (4)**	13 (0)
8 (0)	**11 (−3)**	8 (0)	8 (0)	**10 (−3)**	8 (0)	5 (0)	**4 (0)**	3 (0)	8 (0)	**53 (7)**	16 (0)	15 (0)	**40 (0)**	16 (0)
9 (0)	**10 (−2)**	9 (1)	9 (0)	**11 (−4)**	8 (0)	5 (0)	**4 (0)**	4 (0)	8 (0)	**51 (5)**	16 (0)	17 (0)	**40 (1)**	16 (−1)
3 (0)	3 (0)	**18 (9)**	3 (0)	3 (0)	**16 (8)**	8 (0)	8 (0)	**66 (11)**	8 (0)	15 (0)	**46 (4)**	16 (−1)	17 (0)	**43 (3)**
7 (0)	7 (0)	**12 (−5)**	8 (0)	8 (0)	**10 (−3)**	7 (0)	8 (0)	**66 (12)**	8 (0)	15 (0)	**52 (8)**	17 (0)	15 (1)	**41 (1)**
8 (0)	8 (0)	**11 (−4)**	8 (0)	8 (0)	**11 (−4)**	8 (1)	8 (0)	**66 (12)**	8 (0)	15 (0)	**46 (3)**	16 (−1)	16 (0)	**41 (0)**
9 (0)	9 (0)	**10 (−3)**	8 (0)	8 (0)	**11 (−3)**	8 (0)	8 (0)	**69 (16)**	8 (0)	15 (0)	**48 (5)**	16 (0)	14 (0)	**45 (4)**

Note. f_1 = First common factor. RMSE values and factor loading bias (in parentheses) computed across 1,000 solutions with at least one Heywood case. Factor analysis was conducted using principal axis factoring and Quartimin rotation. All population models had four indicators per common factor, 150 subjects in each sample data matrix, and did not include model approximation error. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Salient (i.e., nonzero) population loadings are bolded. All differences multiplied by 100.

To better understand why the salient population loadings reflected the largest RMSE values (as seen in Tables 8 and 9), for each salient loading pattern we calculated the proportion of times (across 1,000 trials) that a Heywood case occurred on each factor indicator. As shown in Table S8 of the online supplemental materials, we found that the Heywood cases occurred on predictable factor indicators in each sample factor solution. For instance, Heywood cases occurred most frequently on indicators with a salient (i.e., nonzero) population loading on one of the weak factors in the one, two, or three weak factors conditions. In the one strong loading condition, Heywood cases always occurred on the indicator associated with the largest population loading of each factor. In the sequential loadings condition, Heywood cases almost always occurred on the indicator with the largest population factor loading (.80). Note that when a Heywood case occurs on a given indicator, the estimated factor loading (on its most salient factor) is generally more extreme than its population value. Integrating these findings with those mentioned above, our results suggest that Heywood

Table 10
Factor Score Determinacy Bias Between Heywood and Non-Heywood Solutions

Factor	Population factor score determinacy	Heywood solutions: Bias	Heywood solutions: Proportional bias	Non-Heywood solutions: Bias	Non-Heywood solutions: Proportional bias
Pattern B					
f_1	0.86	0.08	0.09	0.01	0.02
f_2	0.86	0.08	0.09	0.01	0.02
f_3	0.86	0.09	0.10	0.01	0.02
Pattern C					
f_1	0.86	0.08	0.09	0.01	0.01
f_2	0.86	0.07	0.09	0.01	0.01
f_3	0.86	0.08	0.09	0.01	0.01
Pattern D					
f_1	0.94	0.00	0.00	0.00	0.00
f_2	0.94	0.00	0.00	0.00	0.00
f_3	0.53	1.04	1.95	0.16	0.31
Pattern E					
f_1	0.94	0.00	0.00	0.00	0.01
f_2	0.53	0.62	1.18	0.18	0.33
f_3	0.53	0.58	1.09	0.17	0.32
Pattern F					
f_1	0.53	0.47	0.88	0.18	0.34
f_2	0.53	0.45	0.84	0.18	0.33
f_3	0.53	0.46	0.87	0.18	0.34

Note. f_1 = First common factor. Factor score determinacy calculated as the correlation between true and estimated factor scores on a given factor. Bias (calculated by subtracting the population factor determinacy from the sample determinacy) computed across 1,000 trials. Proportional bias calculated by dividing the bias by the corresponding population determinacy value. All 1,000 data sets used for the Heywood solution set produced at least one Heywood case when factor analyzed with principal axis factoring. All population models had four indicators per common factor, 150 subjects in each sample data matrix, and did not include model approximation error. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Factor analysis was conducted with Quartimin rotation.

cases produce biased factor loadings and increase factor loading variability. Moreover, as shown in the RMSE results, this increased variation primarily affects factor loadings with salient population values. Regarding the estimated factor correlations, we found that factor correlation bias never exceeded .01 across the five salient loading patterns.

We now return to the question of how Heywood cases affect FSD values. Recall from Study 2 that we addressed this question at the population level (see Figure 1 above). We now look at how Heywood cases affect estimated FSD values at the sample level. Table 10 presents the FSD bias for each factor in the Heywood and non-Heywood solutions for the five salient loading Patterns B through F. When analyzing these data, we examined both bias (the average difference between the estimated sample FSD values and the corresponding population FSD values) and proportional bias (the bias divided by the population FSD value). These results indicate that the elevated factor loadings that were associated with a Heywood case upwardly biased the estimated FSD values. For example, for factors with weak population indicators (salient loading Patterns D, E, and F), the bias and proportional bias of the estimated FSD values always exceeded .40 and .80, respectively. Even though we found evidence for small positive FSD bias for weak factors in the non-Heywood solutions, these bias magnitudes never exceeded 40% of their corresponding bias values in the Heywood solution set. We believe that these results are the first to suggest that Heywood cases can bias estimated FSD. In the Discussion section, we consider how these biased determinacy values can influence model misinterpretation and model reproducibility.

Table 10 also highlights an important finding for EFA solutions that include at least one Heywood case. Namely, computing FSD values in Heywood solutions can yield senseless results (for example, FSD > 1). As previously noted, Guttman's (1955) equation (see Equation 4) for calculating FSD assumes that the factor structure matrix contains correlations between the observed variables and latent factors. However, when an EFA solution includes a Heywood case, the elements of the factor structure matrix can no longer be interpreted as correlations. Therefore, in the presence of Heywood cases, FSD estimates should not be interpreted.

Study 4: Evaluating Potential Treatments for Heywood Cases

Psychologists who use EFA to better understand their data arguably want evidence that their factor solution is robust and accurate. Our previous study indicated that Heywood cases can hinder these desiderata by biasing factor loadings and factor score determinacy values. What, then, should investigators do if they encounter a Heywood case? Researchers have proposed numerous treatments to remove Heywood cases when they occur, including constraining unique factor variances to non-negative values (Chen et al., 2001; Dillon et al., 1987; Gerbing & Anderson, 1987; Kano, 1998; Kolenikov & Bollen, 2012; Velicer & Jackson, 1990), removing or adding variables (Kano, 1998; Sato, 1987; Van Driel, 1978), respecifying the model by altering the estimation method or number of extracted factors (Dillon et al., 1987; Gerbing & Anderson, 1987; Kano, 1998; Mislevy, 1986; Velicer & Fava, 1998), or by applying Bayesian estimation (Martin & McDonald, 1975). Unfortunately, there are numerous problems associated with many of these proposals. For example, bounding a unique factor variance to zero (arguably the most common treatment for Heywood cases; Velicer & Jackson, 1990) precludes maximum likelihood estimation in many software packages and implies the psychometrically-implausible situation of an error-free variable (Dillon et al., 1987; Kolenikov & Bollen, 2012). Additionally, adding variables (or increasing sample size) may not be a feasible option given time or budget constraints.

In light of these issues, Jung and Takane (2008) developed regularized common factor analysis (RCFA) to constrain estimated factor loading variability and to avoid Heywood cases. This method, which is based on the shrinkage properties of penalized regression (for example, see Friedman et al., 2001), sets unique factor variances proportional to a set of initial values (most commonly, all ones or the anti-image variances). In this way, the factor extraction method need estimate only a single proportionality constant when estimating the unique factor variances (Jung & Takane, 2008, p. 4). Other researchers have proposed similar penalized factor analysis methods to control estimated factor loading variability, such as adding a small positive constant to the diagonal of a possibly nonpositive definite covariance matrix (Finkbeiner & Tucker, 1982; Yuan et al., 2011; Yuan & Chan, 2008).

An attractive feature of RCFA is that it is a general method that can be applied to different factor extraction methods (for example, unweighted or generalized least squares, maximum likelihood; Jung & Takane, 2008). Moreover, previous work (Jung, 2013; Jung & Lee, 2011; Jung & Takane, 2008) has demonstrated favorable model recovery with RCFA relative to nonregularized factor extraction methods. For instance, RCFA seems particularly well-suited for factor analysis in small samples (Jung, 2013; Jung & Lee, 2011), conditions that are likely to produce Heywood cases (for example, Costello & Osborne, 2005; Jung & Takane, 2008; Velicer & Fava, 1998).

We designed Study 4 to evaluate RCFA's factor loading recovery capabilities across the five population models that were described in Study 3. Specifically, to gauge the performance of this approach against more traditional methods, we evaluated factor loading recovery using four factor extraction methods: two forms of RCFA (regularized ULS and ML), PAF, and constrained least-squares (CLS) factor analysis. For each method, we also evaluated bias in the estimated FSD values.

Method

In Study 4, we used the same four Heywood and non-Heywood solution sets that were used in Study 3. Recall that the Heywood solution sets included 1,000 data samples that produced at least one Heywood case when extracting three PAF factors, whereas the non-Heywood solution sets included 1,000 data samples that produced zero Heywood cases. Additionally, two of the solution sets included models with low levels of model approximation error (Briggs & MacCallum, 2003; MacCallum & Tucker, 1991; Tucker et al., 1969), defined as population models with RMSEA values between .00 and .05 (Browne & Cudeck, 1992; Marsh et al., 2004). As in Study 3, all population models were orthogonal, with three common factors and four salient indicators per factor. Moreover, each sample (generated from multivariate normal data) had 150 subjects.[11] All sample factor patterns were rotated with direct

[11] Once again, sample size was kept relatively small in these simulations to increase the probability of observing one or more Heywood cases in the EFA solutions.

Quartimin and subsequently aligned to their population counterpart using the faAlign function (Waller, 2019).

As noted previously, we evaluated the model recovery performance of four factor extraction methods using the samples in the Heywood solution sets. The CLS factor algorithm (Waller, 2019) applies a penalty function to bound all communality estimates to the [0, 1] interval. We used the fareg function (also in the fungible R package; Waller, 2019) to implement the unweighted least squares (RCFA-ULS) and maximum likelihood (RCFA-ML) variants of the regularized common factor analysis method with initial anti-image unique factor variance estimates (Jung & Takane, 2008). To generate baseline data, we also factor-analyzed the non-Heywood solution sets in three ways: (a) extracting PAF factors, (b) applying RCFA-ULS, and (c) applying RCFA-ML.

To compare model recovery performance among these methods, we calculated RMSE values between each sample factor-pattern matrix and its corresponding population model. We also calculated estimated FSD values (see Equation 4) and bias for all factors estimated by the aforementioned factor extraction algorithms. All simulations were completed in R (R Core Team, 2019), using the simFA function from the fungible library (Waller, 2019) to generate the population models and corresponding data sets.

Results

As in Study 3, we found trivial differences in results between solution sets with and without model approximation error. For simplicity, we present here the results for models without approximation error (the corresponding tables for models with low levels of approximation error are available in Tables S11–S12 of the online supplemental materials). Table 11 reports the RMSE values for factor pattern recovery for each extraction method. These values were computed by averaging the full-model RMSE values across the 1,000 solutions generated by each method. A notable finding from this table is that the RCFA methods consistently produced the lowest average RMSE values among the four examined extraction methods. Importantly, our results indicate that the average RMSE values for the RCFA methods applied to the Heywood solution sets closely approximated those from the solutions of the non-Heywood solution sets. For example, the average RMSE values for RCFA differed by no more than .03 when applied to solutions that either did or did not produce a Heywood case. Comparing RCFA and PAF, our results suggest that at the studied sample size, RCFA (with either a ULS or ML discrepancy function) recovers population factor loadings in so-called Heywood data sets as well as PAF in comparable non-Heywood data sets. Notice also that the differences between the full-model and salient loading RMSEs were smallest in the RCFA conditions, partly because RCFA cannot produce Heywood cases.

In our next set of analyses, we compared the estimated FSD values that were generated from the four factor extraction methods. Recall that we previously found that Heywood cases substantially biased the estimated factor determinacy values for the factor on which a Heywood case was present in the sample. Extending these findings, Table 12 presents the FSD bias values for each common factor across the five salient loading patterns. Notice in this table that the data in Columns 1 and 4 (which were originally reported in Table 10) again demonstrate that Heywood cases in PAF factors upwardly bias the estimated FSD values (by as much as 1.04 in the one weak factor condition). The FSD bias was consistently largest for weak factors in multiple factor solutions (e.g., salient loading Patterns D, E, and F). Also notice that, for these problematic cases, implementing a constrained least-squares algorithm reduces the FSD bias by approximately 50%. Moreover, implementing RCFA reduces the bias to approximately 10% to 30% of its original value, and was associated with even lower FSD bias than that found in the non-Heywood PAF solutions. Applying RCFA to solutions with at least one Heywood case produced comparable FSD bias as when applying this extraction method to non-Heywood solutions. Interestingly, for our data, there were no substantial differences between the RCFA-ULS and RCFA-ML solutions in terms of factor loading RMSE and FSD bias.

General Discussion on the Causes, Effects, and Treatments for Heywood Cases in EFA

Although EFA remains a popular psychometric method in the psychological sciences (Fabrigar et al., 1999; Henson & Roberts, 2006; Preacher & MacCallum, 2003), it can produce improper solutions with nonsensical values. A well-known symptom of improper EFA solutions is a Heywood case, wherein the model accounts for 100% or more of an observed variable's variance.

Table 11

Factor Loading Recovery (RMSE) in Heywood and Non-Heywood Solution Sets Using Four Factor Extraction Methods

	Non-Heywood						Heywood							
	PAF		RCFA-ULS		RCFA-ML		PAF		CLS		RCFA-ULS		RCFA-ML	
Loading pattern	All	Salient	All	Salient	All	Salient	All	Salient	All	Salient	All	Salient	All	Salient
B	0.08	0.08	0.08	0.09	0.08	0.09	0.09	0.12	0.09	0.11	0.08	0.09	0.08	0.09
C	0.08	0.08	0.08	0.09	0.08	0.09	0.09	0.12	0.09	0.11	0.08	0.08	0.08	0.08
D	0.09	0.11	0.07	0.08	0.07	0.08	0.22	0.38	0.14	0.23	0.08	0.11	0.08	0.11
E	0.13	0.16	0.11	0.12	0.12	0.13	0.24	0.39	0.17	0.26	0.12	0.14	0.13	0.14
F	0.18	0.21	0.16	0.16	0.16	0.16	0.27	0.41	0.21	0.28	0.16	0.17	0.17	0.18

Note. PAF = principal-axis factoring; RCFA-ULS = regularized common factor analysis with unweighted least-squares; RCFA-ML = regularized common factor analysis with maximum likelihood estimation; CLS = constrained least-squares. Factor loading recovery calculated as the average root mean square errors (RMSE) across 1,000 trials. RMSE calculated either for all loadings in the factor-pattern matrix ("All") or only for the twelve salient (i.e., nonzero) population loadings ("Salient"). All population models had four indicators per common factor, 150 subjects in each sample data matrix, and did not include model approximation error. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Factor analysis was conducted with Quartimin rotation.

Table 12
Factor Score Determinacy Bias in Heywood and Non-Heywood Solution Sets Using Four Factor Extraction Methods

	Non-Heywood			Heywood			
		RCFA				RCFA	
Factor	PAF	ULS	ML	PAF	CLS	ULS	ML
Pattern B							
f_1	0.01	−0.03	−0.01	0.08	0.05	−0.03	−0.01
f_2	0.01	−0.03	−0.01	0.08	0.06	−0.03	−0.01
f_3	0.01	−0.03	−0.01	0.09	0.06	−0.03	−0.01
Pattern C							
f_1	0.01	−0.02	−0.01	0.08	0.06	−0.03	−0.01
f_2	0.01	−0.02	−0.01	0.07	0.05	−0.03	−0.01
f_3	0.01	−0.03	−0.01	0.09	0.06	−0.03	−0.01
Pattern D							
f_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
f_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
f_3	0.16	0.12	0.14	1.04	0.46	0.09	0.10
Pattern E							
f_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
f_2	0.18	0.12	0.13	0.62	0.32	0.11	0.12
f_3	0.17	0.12	0.13	0.58	0.31	0.11	0.12
Pattern F							
f_1	0.18	0.12	0.13	0.47	0.27	0.12	0.12
f_2	0.18	0.12	0.13	0.45	0.27	0.12	0.13
f_3	0.18	0.12	0.13	0.46	0.27	0.12	0.13

Note. RCFA = regularized common factor analysis; PAF = principal-axis factoring; ULS = unweighted least-squares; ML = maximum likelihood estimation; CLS = constrained least-squares. Factor determinacy calculated as the correlation between true and estimated factor scores on a given factor. Bias (calculated by subtracting the population factor determinacy from the sample determinacy) computed across 1,000 trials. All population models had four indicators per common factor, 150 subjects in each sample data matrix, and did not include model approximation error. Five salient loading patterns: B = sequential loadings; C = one strong loading; D = one weak factor; E = two weak factors; F = three weak factors. Factor analysis was conducted with Quartimin rotation.

These improper solutions occur at non-negligible frequencies in both simulation-based and empirical psychological research (Costello & Osborne, 2005; De Winter & Dodou, 2012; Kolenikov & Bollen, 2012; Preacher & MacCallum, 2002), and pose problems for drawing robust inferences regarding underlying factor structures in exploratory work. Although numerous studies have investigated the data and model characteristics that produce Heywood cases in CFA analyses, fewer studies have directly and comprehensively examined Heywood cases in an EFA context. In the present work, we integrated the relevant EFA and CFA literatures to execute four Monte Carlo studies that examined the (a) causes of, (b) effects of, and (c) solutions for Heywood cases in EFA models.

To our knowledge, the first of our four studies is the largest and most comprehensive Monte Carlo study conducted to date that has compared the causes of Heywood cases in EFA models. Like others before us, we found that Heywood cases are more likely to arise in small samples (Costello & Osborne, 2005; De Winter & Dodou, 2012, Jung & Takane, 2008; Velicer & Fava, 1998; Zhang et al., 2010), with fewer salient indicators per factor (De Winter & Dodou, 2012, 2016; MacCallum et al., 1999; Velicer & Fava, 1998), and when implementing maximum likelihood (ML) estimation (De Winter & Dodou, 2012, 2016). Moreover, we found that regardless of the number of common factors in the data-generating model, Heywood cases are more prevalent when estimating one too many factors in the sample solution, corroborating Velicer and Fava (1998) finding that "the occurrence of a [Heywood] case can be viewed as an indication of overextraction" (p. 246). Furthermore, our analysis of the two-way interactions among our design factors suggested that the negative effects of small indicator-to-factor ratios on Heywood case prevalence rates are amplified when overfactoring. In aggregate, these results highlight the importance of working with well-determined questionnaires (i.e., questionnaires with many indicators per dimension, cf. De Winter et al., 2009) and sufficiently large samples to reduce the likelihood of Heywood cases in EFA solutions.

In Study 2, we delved deeper into three potential causes of Heywood cases: (a) factor score determinacy (FSD), (b) correlated common factors, and (c) nonpositive definite (NPD) correlation matrices and matrix smoothing. Expanding upon our earlier results, these simulations suggested that Heywood cases are more likely to occur when population common factors are not well-determined and the factor intercorrelation magnitudes are high. Our simulations also indicated that among the three common matrix smoothing algorithms that we studied, the Bentler and Yuan (2011) algorithm was the most effective at reducing the likelihood of Heywood cases when working with NPD correlation matrices.

Overall, the results from Studies 1 and 2 highlight the importance of working with well-determined factors in EFA research. Numerous design factors in Study 1, including the factor loading pattern and the indicator-to-factor ratio, are known determinants of FSD (Grice, 2001). Study 2 subsequently provided the first set of evidence that the nontrivial effects of these design factors on Heywood case prevalence rates can be conceptualized through the lens of population-level FSD. Given the strong role that FSD played in our results, we propose that future researchers routinely consider this component when researching EFA model recovery. We also hope that our findings on FSD will encourage new work on improper solutions in CFA. Moreover, we encourage psychological researchers to routinely build measures that will enhance the FSD of the common factors. Our studies corroborated previous research (e.g., De Winter & Dodou, 2012; Velicer & Fava, 1998) in that that factor patterns with higher proportions of low-communality factor indicators (e.g., .30, a magnitude that is arguably at the threshold for what researchers would consider interpretable; Fabrigar et al., 1999) consistently produced higher proportions of improper EFA solutions. The development and dissemination of robust scales with strong factor loadings will help ameliorate the likelihood of encountering Heywood cases in psychological data.

We next turn to Study 3, where we examined how Heywood cases can potentially bias researchers' model interpretations. A main conclusion from these simulations is that Heywood cases make it more difficult to accurately estimate factor loadings. Although we did not find large differences in factor loading bias between solutions with or without Heywood cases, we did find larger differences in the RMSE values. These results were particularly magnified for the salient loadings of the independent cluster factor patterns (i.e., simple structure with no cross-loadings; McDonald, 1985) that we examined. We suspect that these increased RMSE values are due to an inconvenient truth of the EFA model. Namely, all rotated factor solutions (including the unrotated solution) from an EFA will generate the same model-implied dispersion matrix (e.g., a correlation matrix). Thus, when one loading of a factor is overestimated (due to a Heywood case on that indicator), the remaining salient loadings on that factor are often underestimated to maintain the model-implied correlations. Although the estimated loadings may not be biased in the technical sense of the term (i.e., when averaged across many samples), in any given sample the factor loadings may be poorly estimated due to the Heywood cases.

We believe that our Study 3 findings have important implications for the evaluation of EFA models. Researchers who do not routinely review their solution's communalities run the risk of interpreting inaccurate factor loadings. Relatedly, because Heywood cases upwardly bias estimates of factor score determinacy, failure to identify Heywood cases when they occur may also facilitate misinterpretations of factor strength (and the implied utility of estimated factor scores). Moreover, our findings have clear interpretations for whether one should include or exclude improper solutions in EFA Monte Carlo studies (MacCallum et al., 2001; MacCallum et al., 1999; Zhang et al., 2010) or when generating bootstrapped estimates of factor parameters (Zhang, 2014; Zhang et al., 2010). In both scenarios, including Heywood cases will tend to inflate the RMSE values of the estimated parameters, thereby inflating bootstrapped standard errors and confidence intervals when performing statistical analyses.

Our final study focused on methods to avoid Heywood cases during factor extraction. Among the four factor extraction algorithms that we evaluated, Jung and Takane's (2008) method for RCFA emerged as an excellent option for both preventing Heywood cases and accurately recovering EFA model parameters. Given these encouraging findings, we recommend that researchers routinely include RCFA in future EFA simulation studies so that more is known about the empirical performance of this promising factor extraction method. However, we stress that implementing "treatment" methods, such as bounding unique factor variances or applying RCFA, does not necessarily remedy the underlying data issues that can engender Heywood cases. Rather, the occurrence of a Heywood case should motivate researchers to carefully review their model structure and assumptions prior to drawing inferences from their results.

To further enhance the generalizability of our simulation results, we incorporated model approximation error (MacCallum, 2003; MacCallum & Tucker, 1991; Tucker et al., 1969) into the design of all four studies. In social science research, it is more realistic to assume that the major common factors do not fully represent psychological processes. Rather, there are likely a myriad of additional minor latent factors, not captured by the specified EFA model, that also contribute to the manifestation of such processes (MacCallum, 2003). Using a well-known technique for modeling this approximation error in EFA (Briggs & MacCallum, 2003; Tucker et al., 1969), we found that model approximation error moderated certain trends in our results, such as the relationship between the factor correlation magnitude and Heywood case prevalence rates in Study 2. Importantly, our analyses examined only models with relatively low levels of approximation error. It is possible that incorporating moderate to high error levels would have further highlighted the influence of model approximation error on improper EFA solutions, particularly in Studies 3 and 4 (where we found negligible differences in results between solution sets with and without model approximation error). In future work, we plan to expand upon these results by systematically studying Heywood case prevalence rates across a larger range of model approximation error levels.

Although much was learned from our simulations, our four Monte Carlo studies were not without limitations. Importantly, although our population models for Studies 2–4 were informed by the trends uncovered in our large-scale Study 1 simulation, certain model features (e.g., factor extraction method, model specification) were held constant. As such, our findings should not be generalized to situations not covered in our simulation designs. We thus recommend that future investigations of Heywood cases examine (a) models with more common factors, (b) models with different numbers of indicators per factor, (c) population factor models with cross loadings, (d) models estimated by different factor extraction algorithms, (e) models with different factor correlation patterns, (f) data drawn from non-normal and noncontinuous distributions, and (g) data sets with different levels or types of model approximation error (Lai, 2019). Additionally, we note that when generating collections of EFA solutions to compute Heywood case prevalence rates, we only analyzed solutions that converged using PAF. As a reviewer noted, this method may bias our results due to the confounding of factor extraction nonconvergence and Heywood cases (see Kano, 1998).

In summary, our four Monte Carlo studies have drawn attention to the well-known, but often underdiscussed, fact that Heywood cases arise often in EFA research (Fabrigar et al., 1999; Kano, 1998; Velicer & Fava, 1998; Velicer & Jackson, 1990). By comprehensively reviewing the causes and effects of Heywood cases, we have shown that Heywood cases can facilitate inaccurate inferences about latent variables, thereby weakening the foundations of reproducible research. We have also offered several recommendations for avoiding or removing Heywood cases when they arise. We hope that these recommendations will help researchers to better elucidate the underlying dimensional structure of their data.

References

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 155–173. https://doi.org/10.1007/BF02294170

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin*, *106*(2), 315–317. https://doi.org/10.1037/0033-2909.106.2.315

Bentler, P. M., & Yuan, K.-H. (2011). Positive definiteness via off-diagonal scaling of a symmetric indefinite matrix. *Psychometrika*, *76*(1), 119–123. https://doi.org/10.1007/s11336-010-9191-3

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Bollen, K. A. (1987). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods & Research*, *15*(4), 375–384. https://doi.org/10.1177/0049124187015004002

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229–242. https://doi.org/10.1007/BF02294248

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, *38*(1), 25–56. https://doi.org/10.1207/S15327906MBR3801_2

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258.

Canivez, G. L., & Watkins, M. W. (2010). Investigation of the factor structure of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS–IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*(4), 827–836. https://doi.org/10.1037/a0020429

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

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*(1), 107–112. https://doi.org/10.1177/001316447303300111

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1–9. https://doi.org/10.4135/9781412995627

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

De Winter, J. C., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*(4), 695–710. https://doi.org/10.1080/02664763.2011.610445

De Winter, J. C., & Dodou, D. (2016). Common factor analysis versus principal component analysis: A comparison of loadings by means of simulations. *Communications in Statistics-Simulation and Computation*, *45*(1), 299–321. https://doi.org/10.1080/03610918.2013.862274

De Winter, J. C., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147–181. https://doi.org/10.1080/00273170902794206

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, *101*(1), 126–135. https://doi.org/10.1037/0033-2909.101.1.126

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(2), 119–143. https://doi.org/10.1080/10705519509540000

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Finkbeiner, C., & Tucker, L. (1982). Approximate uniqueness estimates for singular correlation matrices. *Psychometrika*, *47*(4), 517–521. https://doi.org/10.1007/BF02293712

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*(2), 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer Series in Statistics.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65–83. https://doi.org/10.1207/s15327906mbr4101_5

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika*, *52*(1), 99–111. https://doi.org/10.1007/BF02293958

Goretzko, D., Pham, T. T. H., & Bühner, M. (2019). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-019-00300-2

Gorsuch, R. L. (1974). *Factor analysis*. W. B. Saunders.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. https://doi.org/10.1037/1082-989X.6.4.430

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, *8*(2), 65–81. https://doi.org/10.1111/j.2044-8317.1955.tb00321.x

Harman, H. H. (1960). *Modern factor analysis* (2nd ed.). The University of Chicago Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Heywood, H. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *134*(824), 486–501. https://doi.org/10.1098/rspa.1931.0209

Higham, N. J. (2002). Computing the nearest correlation matrix—A problem from finance. *IMA Journal of Numerical Analysis*, *22*(3), 329–343. https://doi.org/10.1093/imanum/22.3.329

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jackson, D. L., Voth, J., & Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural Equation Modeling*, *20*(1), 86–97. https://doi.org/10.1080/10705511.2013.742388

Jung, S. (2013). Exploratory factor analysis with small sample sizes: A comparison of three approaches. *Behavioural Processes*, *97*, 90–95. https://doi.org/10.1016/j.beproc.2012.11.016

Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods*, *43*(3), 701–709. https://doi.org/10.3758/s13428-011-0077-9

Jung, S., & Takane, Y. (2008). Regularized exploratory factor analysis. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 141–149). University Academic Press.

Kano, Y. (1998). Improper solutions in exploratory factor analysis: Causes and treatments. *Advances in data science and classification* (pp. 375–382). Springer.

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*(3), 457–477. https://doi.org/10.1207/s15327906mbr2603

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociological Methods & Research*, *41*(1), 124–167. https://doi.org/10.1177/00491241124 42138

Lai, K. (2019). Creating misspecified models in moment structure analysis. *Psychometrika*, *84*(3), 781–721. https://doi.org/10.1007/s11336-018-09655-0

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Thomson Brooks/Cole.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, *28*(4), 612–625. https://doi.org/10.1111/j.1468-2958.2002.tb00828.x

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods*, *24*(4), 452–467. https://doi.org/10.1037/met0000230

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Psychology Press.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

MacCallum, R. C. (2003). 2001 Presidential Address: Working with imperfect models. *Multivariate Behavioral Research*, *38*(1), 113–139. https://doi.org/10.1207/S15327906MBR3801_5

MacCallum, R. C., & Tucker, L. (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502–511. https://doi.org/10.1037/0033-2909.109.3.502

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1

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

McArdle, J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, *25*(1), 81–87. https://doi.org/10.1207/s15327906mbr2501_10

McDonald, R. (1985). *Factor analysis and related methods*. Erlbaum.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*(1), 3–31. https://doi.org/10.1207/s15327965pli0102_1

Mulaik, S. A. (2010). *Foundations of factor analysis*. CRC Press.

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics*, *32*(2), 153–161. https://doi.org/10.1023/A:1015210025234

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. https://www.R-project.org/

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287–297. https://doi.org/10.1037/1040-3590.12.3.287

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research (R package version 1.8.12) [Computer software manual]. https://CRAN.R-project.org/package=psych

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research*, *13*(1), 109–119. https://doi.org/10.1177/004912418401300 1004

Ryan, J. J., Carruthers, C. A., Miller, L. J., Souheaver, G. T., Gontkovsky, S. T., & Zehr, M. D. (2003). Exploratory factor analysis of the Wechsler Abbreviated Scale of Intelligence (WASI) in adult standardization and clinical samples. *Applied Neuropsychology*, *10*(4), 252–256. https://doi.org/10.1207/s15324826an1004_8

SAS Institute Inc. (2015). SAS/STAT® 14.1 user's guide. https://support.sas.com/

Sato, M. (1987). Pragmatic treatment of improper solutions in factor analysis. *Annals of the Institute of Statistical Mathematics*, *39*(2), 443–455. https://doi.org/10.1007/BF02491481

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*(1), 148–154. https://doi.org/10.1037/0033-2909.106.1.148

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In I. Borg & P. Ph. Mohler (Eds.), *Trends and Perspectives in Empirical Social Research* (pp. 201–224). Walter de Gruyter.

Steiger, J. H., & Lind, J. (1980, May). *Statistically based tests for the number of common factors* [Paper presentation]. Annual meeting of the Psychometric Society, Iowa City, IA.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. University of Chicago Press.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421–459. https://doi.org/10.1007/BF02290601

Van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, *43*(2), 225–243. https://doi.org/10.1007/BF02293865

Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231–251. https://doi.org/10.1037/1082-989X.3.2.231

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*(1), 1–28. https://doi.org/10.1207/s15327906mbr2501_1

Waller, N. G. (2019). fungible: Psychometric functions from the Waller lab (Version 1.90) [Computer software manual]. https://cran.r-project.org/web/packages/fungible/index.html

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Sage.

Ximénez, C. (2006). A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling*, *13*(4), 587–614. https://doi.org/10.1207/s15328007sem1304_5

Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis*, *52*(10), 4842–4858. https://doi.org/10.1016/j.csda.2008.03.030

Yuan, K.-H., Wu, R., & Bentler, P. M. (2011). Ridge structural equation modelling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology*, *64*(Pt. 1), 107–133. https://doi.org/10.1348/000711010X497442

Zhang, G. (2014). Estimating standard errors in exploratory factor analysis. *Multivariate Behavioral Research*, *49*(4), 339–353. https://doi.org/10.1080/00273171.2014.908271

Zhang, G., Preacher, K. J., & Luo, S. (2010). Bootstrap confidence intervals for ordinary least squares factor loadings and correlations in exploratory factor analysis. *Multivariate Behavioral Research*, *45*(1), 104–134. https://doi.org/10.1080/00273170903504836

Received November 4, 2019
Revision received October 8, 2020
Accepted October 28, 2020 ■