Assessing dimensionality in dichotomous items when many subjects have all zero responses: An example from psychiatry and a solution using mixture models

William F. Christensen, Department of Statistics, Brigham Young University
Melanie M Wall, Department of Psychiatry and Department of Biostatistics, Columbia University
Irini Moustaki, Department of Statistics, London School of Economics

ABSTRACT
Common methods for determining the number of latent dimensions underlying an item set include eigenvalue analysis and examination of fit statistics for factor analysis models with varying number of factors. Given a set of dichotomous items, we demonstrate that these empirical assessments of dimensionality often incorrectly estimate the number of dimensions when there is a preponderance of individuals in the sample with all-zeros as their responses, e.g. not endorsing any symptoms on a health battery. Simulated data experiments are conducted to demonstrate when each of several common diagnostics of dimensionality can be expected to under- or over-estimate the true dimensionality of the underlying latent variable. An example is shown from psychiatry assessing the dimensionality of a social anxiety disorder battery where 1, 2, 3, or more factors are identified, depending on the method of dimensionality assessment. An all-zero inflated exploratory factor analysis model (AZ-EFA) is introduced for assessing the dimensionality of the underlying subgroup corresponding to those possessing the measurable trait. The AZ-EFA approach is demonstrated using simulation experiments and an example measuring social anxiety disorder from a large nationally representative survey. Implications of the findings are discussed, in particular regarding the potential for different findings in community versus patient populations.

Assessing dimensionality in dichotomous items when many subjects have all zero responses: An example from psychiatry and a solution using mixture models

1. Introduction

A substantial psychiatric literature exists that explores the latent dimensional structure underlying different psychiatric symptom batteries (e.g., PTSD – Hukkelberg and Jensen, 2011; Prolonged Grief Disorder – Prigerson et al., 2009, Simon et al., 2011; Substance Use Disorders – Saha et al., 2006). It is instructive to identify shared dimensions underlying different symptoms as it provides a framework for the development of theory related to the characterization of psychiatric disorders. Although we will focus on psychiatric applications herein, analogously in educational testing assessing the underlying dimensionality of test batteries can be instructive for understanding the nature of multifaceted educational outcomes. Assessment of dimensionality, in particular unidimensionality, is commonly done to justify creating a singular measure or sum score representing the severity of the latent disorder based on endorsement of a set of dichotomous symptoms (or criteria as referenced in the Diagnostic and Statistical Manual of Psychiatric Disorders DSM-5). For example, MacCoun (2013) notes that unidimensionality is the basis for characterizing substance use disorders (SUD) in the DSM-5 according to the count of the number of 11 criteria (with Mild, Moderate, and Severe disorder defined at 2, 4, and 6 criteria met).

Given the importance of the unidimensionality conclusion for measurement, the present paper will examine dimensionality assessment for batteries of dichotomous items when the sample has features commonly observed for psychiatric batteries in the general population, i.e. where many subjects do not endorse any of the symptoms. Large surveys of the general population are commonly used to understand the prevalence and nature of mental or physical health conditions, with prominent examples including the National Comorbidity Survey and the National Epidemiologic Survey on Alcohol and Related Conditions. Because the collection of survey participants is comprised of a mix of both healthy subjects and subjects with a condition of interest, data from such surveys contain many responses that are all zeros. We refer to this as "all-zero" inflated data in contrast to simply zero inflated data to emphasize that there is a non-trivial number of subjects with all items equal to zero, not just one item with a lot of zeros. Two mechanisms could lead to all-zero inflated responses: (1) the items are too extreme or severe for some subjects being measured who possess the trait, or (2) the sample contains a subset of subjects who do not possess the trait being measured. The presence of all-zeros depends on the real-world setting from which the sample is drawn. In a treatment-seeking population it is likely that many patients will endorse a substantial number of symptoms, and those that do not would be considered to be at the low end of severity for the disorder trait (an area of the trait not well captured by the battery). Thus, the first type of all-zero response arises from the pathological class of subjects whose status is not severe enough at the time of measurement to lead the subject to endorse any of the symptoms. In contrast, in a community sample where most people do not have the disorder, many individuals are expected to endorse none of the symptoms because the trait is not relevant to them and only a small number are expected to be pathological, i.e. have the disorder for which severity is being measured. We refer to this class as the non-pathological or healthy subjects. Our discussion here focuses on the problems associated with dimensionality assessments in large community-based samples where the presence/absence of symptoms have been characterized. Specifically, we show that a naïve handling of the all-zero responses (e.g., removing them completely or failing to adapt the analysis methods to account for them) can

result in a distortion of the dimensionality assessment. We note that alternative non-normal latent traits including unipolar traits may also lead to such type of all-zero inflated data (Smits, et al., 2020).

Using an extensive simulation study generating data from different settings, we will demonstrate that using traditional exploratory factor analysis (EFA) methods, unidimensionality can incorrectly be concluded when the sample has a high proportion of all-zero inflated responses. We also propose an approach to assess dimensionality that utilizes a latent variable mixture model with two components, one that is similar to the usual EFA model for dichotomous items and one that accounts for the all-zero inflation. The model extends one previously applied to unidimensional IRT models in the presence of zero-inflation (Wall, et al., 2015). Using a latent mixture model allows for the latent trait(s) to be measured one way in the pathological group for whom the trait(s) and severity are relevant and to be all-zero with probability one in healthy/asymptomatic subjects. The behavior of the pathological group is modeled with a q-dimensional factor model, while the healthy group is constrained to be asymptomatic. That is, our model implicitly assumes that healthy subjects will yield all-zero responses. Because each of the groups potentially contains subjects exhibiting no symptoms, we hypothesize that fitting the mixture model to the complete data will yield a more accurate characterization of the dimensionality of the latent trait than standard methods applied to either the complete sample or the sample after filtering out the subjects exhibiting no symptoms.

This paper is organized as follows: first we discuss the traditional reflective latent measurement model and the assessment of factor dimensionality, then we introduce the all-zero inflated EFA (AZ-EFA) for dichotomous items and how it can be used to assess dimensionality in a sample with all-zero inflation. The subsequent section presents a simulation study to illustrate the challenge of assessing factor dimensionality using traditional methods and the advantage of the proposed AZ-EFA for dimensionality assessment. We then apply the methods to the DSM IV situational fears (13 items) for measuring social anxiety disorder (SAD) as found in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), a large national community survey of over 40,000 adults.

2. The traditional EFA model for dichotomous items and related methods for assessing dimensionality

The basic idea of EFA is the following: for a given set of p observable response variables $Y_1, \ldots, Y_p$ one wants to find a set of $q < p$ continuous latent variables also known as factors $f_1, \ldots, f_q$, that contain essentially the same information. The latent factors are supposed to account for the dependencies among the observable response variables in the sense that if the factors are held fixed, the observed variables would be independent. If the response variables $Y_j, j = 1, \ldots, p$ are binary, one needs to specify the probability of each response pattern as a function of $f_1, \ldots, f_q$:

$$Pr(Y_1 = a_1, Y_2 = a_2, \ldots, Y_p = a_p \mid f_1, \ldots, f_q)$$

where $a_j = 0, 1$ represents the two response categories of Y_j respectively.

There are two main approaches for analyzing binary observed variables using latent variable models and within these there are many variants. The first is the underlying variable approach (UVA), which assumes that each observed binary variable Y_j is generated by an underlying unobserved continuous variable Y_j^* assumed to be normally distributed. This approach employs the classical linear factor analysis model:

$$Y_j^* = \lambda_{1j}f_1 + \cdots + \lambda_{qj}f_q + e_j, (1)$$

where $j = 1, \ldots, p$, and e_j is an error term representing a specific factor and measurement error. In addition, e_j is normally distributed with mean 0 and variance 1 and the vector of latent variables, $\boldsymbol{f} = (f_1, \ldots, f_q)'$ follows a multivariate normal distribution, $\boldsymbol{f} \sim N_q(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix. The connection between the binary variable Y_j and the underlying variable Y_j^* is $Y_j = 1$ when $Y_j^* > \tau_j$ and 0 otherwise and τ_j is a threshold parameter estimated by the data.

The second approach for analyzing binary observed variables using latent variable models is the response function approach (RFA), which specifies the conditional distribution of the complete p-dimensional response pattern as a function of the latent variables and makes the assumption that responses to different variables are independent for given latent variables (Y_j independent of $Y_{j'}$, conditional on $\boldsymbol{f}$). Let $P(Y_j = 1|\boldsymbol{f}) = P_j(\boldsymbol{f})$ stands for the probability that a randomly selected individual with latent vector $\boldsymbol{f} = (f_1, \ldots, f_q)'$, responds to the jth item in the affirmative, where $j = 1, \ldots, p$. We consider the multidimensional two-parameter logistic (2PL) model (Birnbaum 1968),

$$logit\ P_j(\boldsymbol{f}) = \beta_{0j} + \sum_{k=1}^{q}\beta_{kj}f_k \quad (2)$$

where β_{0j} and β_{kj} are the item severity parameters and factor loadings respectively. The latent variables are again assumed to be normally distributed, $\boldsymbol{f} \sim N_q(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix.

Estimation proceeds using weighted least squares for the UVA and (usually) maximum likelihood for the RFA. Each approach is easily implemented in standard software (e.g. Mplus 8, Muthén and Muthén, 1998-2017).

In order to use these factor analysis models, it is necessary to specify a-priori q, the dimensionality of the underlying latent factor. A common first step in determining q is to perform an eigenvalue analysis of either the correlation matrix formed between the items observed on a sample of individuals or the adjusted correlation matrix containing estimates of the communalities for the factor model as diagonal elements; we refer to the latter (preferred) option as "the eigenvalues of the factor model." In the case of dichotomous items, the tetrachoric correlation matrix is used. While there is no single test for the dimensionality of an item set using eigenvalues, common rules of thumb consider the dimensionality to be equal to: (i) the number of eigenvalues greater than 1 (or "Rule of 1"), (ii) the number of eigenvalues that occur before the elbow in a scree plot, or (iii) the number of eigenvalues that are significantly greater than eigenvalues generated by chance with the same sample characteristics – so-called parallel analysis (Horn, 1965). Green et al. (2016) demonstrated satisfactory performance of traditional and revised versions of parallel analysis for dimensionality assessment with binary data. Although eigenvalue-based criteria for assessing the goodness of fit of a latent variable model is considered to be too simplistic in many settings, they are considered here because they remain a commonly used rule of thumb for evaluating dimensionality. Other techniques for assessing the number of latent dimensions underlying an item set or for comparing models with varying numbers of factors include the likelihood ratio test for measuring goodness-of-fit. Furthermore, there are many goodness-of-fit measures (e.g., CFI or RMSEA) that aim to determine the minimal number of factors needed for the model to satisfy cut-offs determined to indicate "good fit." For example, commonly used cut-offs include those recommended by Hu and Bentler (1999) (i.e., choose the smallest number of factors needed to obtain a CFI value that is greater than 0.95 or an RMSEA value that is less than 0.06). Additionally, if full maximum likelihood

estimation is used, model comparison can be done using BIC and choosing the number of factors based on the model with smallest value.

3. The AZ-EFA for assessing dimensionality

A shortcoming of the latent factor models described above is their reliance on the homogeneity of the latent factor distribution, typically assumed to be multivariate normal. The more flexible model considered here is to replace the multivariate normality assumption for the underlying factors in a traditional latent factor model with a mixture of a multivariate normal and a degenerate distribution. To account for the excess of zeros in the data, it is assumed that the population is made up of two subgroups, one that has a multivariate normal distribution on the latent variables with observed response profiles coming from the factor analysis model described in (1) and (2) and another that is degenerate without the latent variables and all-zero responses on the observed variables. Because the two subgroups are unobserved, a mixture model is specified in terms of a latent categorical variable (latent class), which here has only two values but in principle there could be more than two subgroups. The model with more than two groups but unidimensional latent variable is discussed in Wall, et al. (2015).

Let us assume that the non-pathological group is represented by the first mixture component that has a fixed mean at an extreme negative value and a zero variance, i.e.

$$\mu_1 = -100 \text{ and } \sigma_1^2 = 0.$$

In other words, this component has a degenerate distribution with mass 1 at the value of the trait equal to -100. By fixing μ_1 at an arbitrarily large negative value and the variance to zero this implies $f^{(1)} = -100$ for all individuals in this class, which then ensures practically zero probability of endorsement to each item. The pathological group is represented by the second mixture component for which the latent variables remain as before, $f^{(2)} \sim N_q(\mathbf{0}, \mathbf{I})$.

Finally, the model with the two components is written as:

$$p(Y_1, \dots, Y_p) = \sum_{m=1}^{2} \eta_m \int p(Y \mid f^{(m)}) p(f^{(m)}) df^{(m)} =$$
$$\eta_1 + \eta_2 \int \prod_{j=1}^{p} p(Y_j \mid f^{(2)}) p(f^{(2)}) df^{(2)}$$

where η_m denotes the mixture proportions, and $p(Y_j \mid f^{(2)})$ is the Bernoulli distribution with probability of a positive response given by $P(Y_j = 1 \mid f^{(2)}) = P_j(f^{(2)})$ and modelled by (2). For simplifying the notation from now on the latent variable in the second mixture component will be denoted just with f instead of $f^{(2)}$. This model falls within the class of hybrid latent class latent factor models and can be estimated via maximum likelihood in existing software such as Mplus (Muthén and Muthén, 1998-2017) and LatentGold (Vermunt and Magidson, 2016). The dimensionality of the set of observed items can then be obtained by identifying the best fitting number of dimensions q for the non-degenerate mixture component choosing the one with the smallest BIC value. Example code for fitting the AZ-EFA with a degenerate component and a two-dimensional factor model underlying a set of 12 dichotomous observed variables is given in the Appendix S5 of the Supplementary Materials.

4. Simulated data experiment

We will consider the following data generating model: twelve dichotomous items $Y_1, Y_2 \ldots, Y_{12}$ are generated from a two-dimensional latent factor model with the following two-parameter logistic (2PL) structure relating the logit of the probability that $Y_i=1$ and the latent factors f_1 and f_2:

$$logit \Pr(Y_i = 1 \mid f_1, f_2) = -\beta_{oj} + \beta_{1j}f_1 + \beta_{2j}f_2 \qquad (3)$$

where $\beta_{oj} = \beta_o$ is the same for all 12 items and will be at a fixed value described below, $\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)'$, $\boldsymbol{\beta}_2 = (0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1)'$, and (f_1, f_2) follow a bivariate uncorrelated standard Normal $N_2(0,I)$. Model (3) exhibits what is often called "simple structure" with the first 6 items measuring the first factor and the second 6 items measuring the second factor. Furthermore, because the non-zero values in β_1 and β_2 are all equal, the strength of the relationship between each item and its respective factor, f_1 or f_2, are all equal. We chose these features of the data generating model for illustration to emphasize a case where identifying the latent dimensionality to be two (rather than one or three) should be straightforward.

Suppose that 4000 subjects are generated from the model (1) with $\beta_o = 0.25$ (note, the sample size corresponds to a large national survey – NESARC – used later in the example). The tetrachoric correlations for pairs of items associated with the same factor are roughly 0.27 while the tetrachoric correlations for pairs of items associated with two different (independent) factors will have an expected value of 0. A two-dimensional latent factor structure is clearly evident from the block diagonal nature of the correlation matrix. But, now consider what happens if we add a large number of asymptomatic subjects (i.e., subjects having all-zeroes) to the dataset. This all-zero inflation makes the relationship between items as measured by the tetrachoric correlation increase dramatically. For example, adding 36,000 asymptomatic subjects (such that 90% of the whole sample is all-zeros) increases the tetrachoric correlation for items associated with the same factor from roughly $r = 0.27$ to $r = 0.64$. Moreover, it also induces correlation for items associated with independent factors, increasing them from $r = 0.00$ before the asymptomatic subjects are added to $r = 0.48$ afterwards. This happens because subjects with a zero on one item are now much more likely to also have a zero on the other items. Thus, with the large number of asymptomatic cases included, differentiating between a one- and two-dimensional latent structures is much more difficult since the tetrachoric correlations among structurally related items are only 33% larger than the correlations between structurally unrelated items. This demonstration indicates that all-zero inflation can lead to a distortion of the apparent latent dimensionality underlying a set of dichotomous items.

We now describe details of the current simulation study to examine the ability of different methods for identifying the correct dimensionality under various data scenarios. Specifically, we study dimensionality assessment when using different sample sizes ($n = 4000$, 12,000, or 20,000) and symptom severity levels ($\beta_o = 0.25$ and 2), and the possible addition of a large number of subjects with all-zeros as responses. The large sample sizes chosen here correspond to the sample size of a nationally representative sample – NESARC – used in the example application. When $\beta_o = 0.25$ and 2, the probability that the response vector under (1) will contain all zero values equals approximately 1% and 20%, respectively. The various scenarios for the simulation are outlined in Table 1. The first three rows of the table illustrate the performance of dimensionality assessment methods under a baseline scenario of $\beta_o = 0.25$— when severity is moderate (about 1% of simulated subjects have all-zero responses) and there are no additional all-zero observations mixed into the data. Rows four through six consider how dimensionality assessments change from the baseline scenarios when the severity of the items

($\beta_o = 2$) increases to the point that roughly 20% of simulated subjects have all-zero responses. The final three rows evaluate dimensionality assessment techniques when the severity of items are identical to the first three rows ($\beta_o = 0.25$), but when the data for analysis contain additional subjects with all-zero responses. That is, scenarios 1a, 2a, and 3a represent data coming from a mixture of pathological subjects like those in Scenarios 1-3 plus healthy subjects without any symptoms. Figure 1 shows histograms of the sum of the total symptoms present for 3 scenarios.

The true dimensionality of the item set $Y_1, Y_2 \ldots, Y_{12}$ is equal to two in all scenarios since the items are always generated with probabilities that are a function of two factors. To empirically assess the dimensionality of the item set we will consider four types of general approaches: (i) eigenvalue based methods, (ii) exploratory factor analysis (EFA) for 1-, 2-, and 3-factor models using weighted least squares based methods to estimate the parameters of model (1) in Section 2, (iii) EFA for 1-, 2-, and 3-factor models using maximum likelihood methods to estimate the parameters of model (2) in Section 2, and (iv) the AZ-EFA approach. The eigenvalue based methods were implemented in R (R Core Team, 2019). We calculate tetrachoric correlations among the 12 items and consider the number of principal component eigenvalues that exceed 1. We also consider the number of eigenvalues of the factor model (defined in Section 2) that exceed the 95th percentile of the eigenvalues obtained from random matrices (as implemented in the `fa.parallel` function from the `psych` package in R; see [Horn, 1965]). All other approaches were implemented in Mplus. For EFA with weighted least squares methods, we consider an approximate X^2 goodness-of-fit test (X^2_{WLS}), the root mean square error of approximation (RMSEA) index of Steiger and Lind (1980), and the comparative fit index (CFI) of Bentler (1989). For EFA with maximum likelihood, we consider the Pearson X^2 goodness-of-fit test (X^2_P) and the Bayesian Information Criterion (which we denote BIC_{ML}). For the AZ-EFA, we used the BIC (which we denote $BIC_{AZ\text{-}EFA}$). When using BIC metrics, the estimated dimensionality was determined by the number of factors that minimizes the criterion. For the other EFA model assessment criteria, following commonly used cutoffs (see references within Hu and Bentler, 1999) we identify the dimensionality as the smallest number of factors that satisfy the following rules: X^2 test p-value > 0.05, RMSEA < 0.06, and CFI > 0.95. Other metrics available in Mplus (such as Akaike's Information Criterion, adjusted BIC, and the Likelihood Ratio X^2 test statistic) were also considered, but in the interest of brevity, were not highlighted in the discussion below because they performed uniformly worse than some other criterion.

4.1 Results of simulated data experiment

In Section 4.1.1, we first consider the fit of the traditional factor analysis model and the AZ-EFA model to the complete data set. Because an intuitive approach is to fit a traditional factor analysis to the data after removing the all-zero responses, Section 4.1.2 evaluates this strategy. As will be confirmed later, this strategy is expected to perform poorly when the all-zero cases arise at least in part from pathological subjects responding negatively to high-severity items.

4.1.1 Comparing methods for assessing dimensionality

Table 1 summarizes the analyses of 100 simulated data sets generated for each scenario and the percent of time that the correct number of factors (i.e., two) is chosen. Scenarios 1, 2, and 3 in Table 1 illustrate the performance of dimensionality-assessment techniques when the data are "well behaved." That is, moderate-severity items are generated from a two-factor model so that few subjects ($\approx$1%) have all zeros for responses, and no additional all-zero responses are appended to the data. Regardless of the sample size (4,000, 12,000, or 20,000), most of the

techniques have adequate performance, identifying the correct number of factors (i.e., two) at least 90% of the time. Among the dimensionality assessments that performed adequately in the scenarios characterized by the first three rows of Table 1, the top performers were the eigenvalue-based "Rule of 1" (the number of eigenvalues exceeding 1 for the tetrachoric correlation matrix) and BIC_{AZ-EFA}; each of these techniques identified the correct dimensionality 100% of the time in the well-behaved data scenarios.

The rows for Scenarios 4, 5, and 6 in Table 1 illustrate the performance of dimensionality-assessment techniques when data are generated from a two-factor model where items have high severity (i.e., roughly 20% of the subjects have all-zero responses), but no additional all-zero responses are appended to the data. The results for these high-severity scenarios are similar to the results for Scenarios 1, 2, and 3, but for these data, two additional methods break down. The criteria associated with X_P^2 and RMSEA (from the ML and WLS fits of the traditional model, respectively) are also unable to consistently diagnose the dimensionality of the item set. As in the moderate-severity scenarios (Scenarios 1 through 3), the eigenvalue-based "Rule of 1" and BIC_{AZ-EFA} always identified the correct dimensionality of the item set for the high-severity scenarios.

Data-generation Scenarios 1a, 2a, and 3a are described in the seventh through ninth rows of Table 1. These scenarios are similar to the well-behaved scenarios (1, 2, and 3), except that in these scenarios, a set of all-zero observations were added to the pathological cases so that the total sample size was 40,000 in each case. These scenarios simulate a large community-based sample wherein a relatively small number of pathological cases are mixed with a large number of healthy subjects who exhibit none of the measured symptoms. In these scenarios, most of the commonly-used dimensionality-assessment tools fail badly. Although the eigenvalue-based "Rule of 1" performed well when the 40,000 subjects are split evenly between pathological and healthy (all-zero) subjects, for the two cases where the healthy subjects outnumber the pathological subjects (Scenarios 1a and 2a in Table 1), the eigenvalue-based "Rule of 1" consistently diagnoses the item set to be unidimensional. This tendency toward unidimensional assessments in the presence of additional all-zero cases also holds true for the RMSEA and CFI measures obtained from the WLS fit of the traditional factor analysis model. The parallel analysis based on the eigenvalues of the factor model and BIC_{AZ-EFA} are each able to properly identify the proper dimensionality in the data for all cases.

For a more detailed discussion of the impact of all-zero cases on tetrachoric correlations, see Appendix S1 in the Supplemental Materials. We also provide additional simulations in Appendix S2 in the Supplemental Materials which examine the dimesionality-assessment approaches when: (i) factors are correlated, (ii) the data-generation model is truly unidimensional, and (iii) sample sizes are small. To summarize the conclusions drawn from these simulations in Appendix S2, we conclude that identifying true unidimensionality is relatively easy for the majority of methods considered, and that generating data using correlated factors (versus uncorrelated ones) has little impact on these dimensionality assessments. However, we do see that dramatically reducing the sample sizes (e.g., to 80 pathological cases plus 320 healthy—and asymptomatic—cases) causes all of the methods to break down and erroneously conclude unidimensionality at unacceptably high rates. This is not entirely surprising, but emphasizes the fact that the AZ-EFA approach was designed for large mixed-status samples (such as the NESARC study discussed in Section 5) where hundreds or thousands of pathological cases are intermingled with many more healthy cases.

Although it is not central to this article's primary focus on dimensionality assessment, also of importance to practitioners is the ability of the various models to estimate the parameter values of the underlying model. Among the model parameters of greatest interest are the factor loadings and—for the AZ-EFA model—the mixture coefficient that characterizes the split of healthy from pathological subjects. Supplemental Appendix S3 in the Supplemental Materials discusses the issue of factor loading estimation in greater detail, but here we summarize that discussion by noting that the addition of all-zero observations to the pathological cases causes the factor loadings to exhibit substantial bias. In contrast, the estimated factor loadings for the non-degenerate (pathological) class in the AZ-EFA fit are unbiased. With respect to the estimation of the mixing proportion of healthy versus pathological subjects in the data, the AZ-EFA is nearly perfect in its ability to identify the proportions of all-zero cases arising from (1) pathological subjects whose current state is not severe enough to warrant endorsement of any of the severe items and (2) healthy subjects who exhibit no symptoms. See Supplemental Appendix S4 in the Supplemental Materials for a thorough discussion of the estimation of the mixing proportion.

To summarize the conclusions drawn from these simulations, the presence of the inflated zeros is problematic only in those scenarios where the additional all-zero subjects are healthy and the trait is not relevant. We conclude that the presence of additional all-zero subjects (as when healthy subjects are mixed with pathological subjects in community-based samples) will often lead researchers to underestimate the dimensionality of the phenomena measured by the items. Finally, when the number of pathological cases among the observations is expected to be small (e.g., ≤ 200), the true dimensionality of a phenomenon is likely to be understated using any of these methods; AZ-EFA is a useful tool only when the data contain a moderate to large number of pathological subjects. These simulations indicate that $BIC_{AZ\text{-}EFA}$ is the best indicator of dimensionality, with the parallel analysis based on the eigenvalues of the factor model and the X^2_{WLS} approaches performing only slightly worse.

4.1.2 A close look at deleting the all-zero responses

When working with data consisting of many all-zero cases, one possible method an industrious researcher may consider is to assess dimensionality using the traditional methods after removing all of the all-zero cases. For our simulation, we also consider this approach, repeating the traditional methods described above but in datasets where the subjects with all-zeros are dropped (Table 2). That is, Scenarios 1-6 in Table 2 have the same data-generation structure as previously described, but the all-zero observations are omitted after the data are generated. Note that Scenarios 1a, 2a, and 3a are not included in Table 2 because after removing the all-zero observations, these scenarios yield data that are identical to Scenarios 1, 2, and 3. Figure S2 in the Supplemental Materials illustrates the average eigenvalue profile for Scenarios 1 and 4 of Table 2. For moderate-severity items, removing the all-zero items has little impact on the eigenvalue profile for the data; the profile for Scenario 1 is nearly identical when comparing the "use all data" approach in Figure S1 and the "remove all-zero observations" approach in Figure S2 (found in the Supplementary Materials). This might seem to imply that removing the all-zero observations before analysis is an easy solution to the problem of many all-zero observations in a data set. However, when the severity of items is high (as in Scenario 4), removing the all-zero observations distorts the eigenvalue profile and leads to underestimation of the second eigenvalue and overestimation of the subsequent eigenvalues (compare the eigenvalue profile for Scenario 4 in Figures S1 and S2). Thus, without understanding the underlying severity of the measured items, it is difficult or impossible to accurately determine

whether the removal of all-zero items will improve or destroy the integrity of standard assessments of dimensionality.

Table 2 validates the concerns raised by the changes in eigenvalue profiles when removing the all-zero observations. When analyzing moderate-severity items as in Scenarios 1, 2, and 3 (and correspondingly, Scenarios 1a, 2a, and 3a), all methods but the X_P^2 perform adequately (i.e., all other methods identify the correct number of factors at least 90% of the time). However, Scenarios 4, 5, and 6 tell a different story. In Scenarios 1a, 2a, and 3a, the all-zero observations in a data set originate from the healthy subjects in the sample's mixture of pathological and healthy subjects (i.e., a mixture of high-severity cases and non-cases). In contrast, Scenarios 4, 5, and 6 draw subjects from a homogeneous population of pathological persons, but all-zero observations arise because the items are sufficiently severe to yield a large number of sampled subjects with a negative response (i.e., "no" or "incorrect") on every item. In Scenario 4, the "Rule of 1" and the parallel analysis of the factor model's eigenvalues tend to overestimate dimensionality. Most other methods conclude unidimensionality in Scenario 4. Scenarios 5 and 6 are problematic for all methods except the "Rule of 1."

As noted earlier, the difficulty in choosing whether to retain or remove the all-zero data observations lies in the fact that the answer depends on the source of the all-zero observations. When the all-zero observations in data arise from the inclusion of healthy subjects in the sample (true zero-inflation since the trait(s) are not relevant to them), removing the all-zero observations results in adequate dimensionality assessments associated with standard approaches like the "Rule of 1," parallel analysis, RMSEA, and CFI. However, if the all-zero observations in data arise not from a mixed sample but from a set of high-severity items, then removing the all-zero observations ruins the performance of all the standard dimensionality-assessment methods. In practice, removing the all-zero observations in this latter scenario will cause some otherwise-trustworthy methods to underestimate the true dimensionality while others overestimate it. A possible aid in apprehending the source of the all-zero observations is a histogram of the total number of exhibited symptoms for all subjects. When the all-zero observations arise from measurements of moderate-severity items on a mixed sample, one might expect to see a large spike at 0, with a distinguishable second mode at some positive value, as in Figure 1(c). When the all-zero observations arise solely from a high-severity set of items, one would expect a mode at or near zero and a monotonically decreasing probability distribution function for all values greater than the mode, as in Figure 1(b). Notwithstanding the utility of such exploratory tools, it is not knowable whether the all-zero observation arises from a mixed sample (e.g., of pathological and healthy subjects), from high-severity items, or some combination of the two. Because it is difficult to determine which analyses might be facilitated by dropping the all-zero observations, such a strategy is discouraged.

5. Dimensionality assessment of social anxiety disorder from the NESARC study

In this section we consider the underlying dimensionality exhibited by responses to the 13 DSM IV situational fears for measuring social anxiety disorder (SAD) as found in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), a large national U.S. community survey. The NESARC study contained 43,093 subjects, including 5044 reporting presence of at least one symptom of SAD and 38,093 subjects with all-zero responses (88% of the sample). Distributions for the total number of symptoms exhibited are shown in Figure 2. Note that the distribution of exhibited symptoms reflects a scenario where the representation of symptomatic subjects within the total sample is low and—from a comparison of the distributions for total symptoms exhibited on the plot on the right of Figure 2 and the plot on the right of

Figure 1—the aggregate severity of the symptoms for these data appears to be in between the two levels of severity in the simulations from the previous section. That is, we believe that most of the all-zero responses are due to a large number of healthy subjects combined with a small number of pathological subjects who have low severity on the trait. Consequently, based on our simulation experiment we suspect that when assessing the latent dimensionality of the SAD symptoms in the NESARC sample, the "Rule of 1" and other traditional methods (RMSEA and CFI) could each lead to erroneous conclusions of reduced dimensionality or unidimensionality.

The 13 fear symptoms are listed in the left column of Table 3. Table 4 reports the dimensionality assessment metrics calculated in Tables 1 and 2 from the simulated data experiments. Using BIC with the recommended AZ-EFA method, it is apparent that the 1-factor model is vastly inferior to relatively comparable 2- and 3-factor models and that the 3-factor model is considered optimal using the recommended BIC statistic. The "Rule of 1" yields strong but likely misleading evidence for the 1-factor model, with the first three eigenvalues equal to 11.13, 0.50, and 0.29. Recall from the simulations that a false indicator of unidimensionality is exactly what we expect from the "Rule of 1" if the number of all-zero (asymptomatic) subjects dominates the number of pathological subjects and if the all-zero observations are not likely to have arisen from high-severity items.

Using the full sample with standard ML and WLS methods leads to a suite of incongruous conclusions when using the standard dimensionality-assessment metrics; the "Rule of 1" and some of the statistics associated with the WLS estimation (RMSEA and CFI) all choose the 1-factor model, the parallel analysis of the factor model's eigenvalues gives slight preference to the 2-factor model over the 3-factor model, and the maximum-likelihood-based χ_P^2 and BIC_{ML} indicate that three or more factors are necessary. These results generally correspond to the dimensionality assessments obtained from Scenario 1a in Table 1, which showed that most methods ("Rule of 1," RMSEA, and CFI) identify a number of factors that is less than that identified by the top-performing BIC_{AZ-EFA} and parallel analysis approaches. In contrast, simulations show that the χ_P^2 and BIC_{ML} methods often identify a number of factors that is greater than that identified by BIC_{AZ-EFA}. When the asymptomatic subjects are removed (an approach we generally discourage) and the same estimation approaches are used, the methods choose 2 or 3 or more factors. Based on the simulations, these changes in dimensionality assessments are exactly as expected when the all-zero observations are removed before analysis.

The AZ-EFA approach yields factor loadings for the pathological portion of the mixed-sample. The estimated loadings and associated p-values for the three-factor AZ-EFA are given in Table 3. Note that the first factor is dominated by symptoms 1-4, the second factor is dominated by symptoms 4-10, and the third factor is dominated by symptoms 4 and 9-13. An examination of the symptoms in these groups leads us to characterize the first factor as a measurement of an anxiety related to the subject being placed in a "spotlight"—the subject is an object of focused attention in a class, meeting, or performance setting. The second factor is characterized as relating to social interactions. The negative loading on the item related to taking an important exam serves to distinguish the meaning of factor 2 ("social") from the meaning of factor 3 ("scrutiny"). Factor 2 has a 0.36 correlation with factor 1 ("spotlight"). We characterize factor 3 as an anxiety associated with social scrutiny. The "scrutiny" factor includes anxieties related to the personal judgment associated with writing while being watched, taking an exam, being interviewed, speaking to an authority figure, or dating. The negative loading on the item associated with speaking in front of people accentuates that the "scrutiny" factor is not capturing the anxiety associated with the mechanics of a social interaction, but rather with the social

phenomenon related to being judged. The correlations between the third ("scrutiny") factor and each of the first two factors (0.60 and 0.82, respectively) are stronger than the correlation between the first ("spotlight") and second ("social") factors. These correlations characterize the importance of the second and third factors in much the same way as the values of BIC for the one-, two-, and three-factor AZ-EFA models. That is, there is a great deal of additional latent structure that is captured when moving from one factor to two (BIC drops 1241 units) whereas the case for the addition of a third factor (numerically warranted by a further 106 unit drop in BIC) is less compelling than the addition of a second factor.

6. Discussion

We have presented evidence to demonstrate the problem of determining the dimensionality associated with of a set of items when a preponderance of subjects have all-zeros as their responses. Examples include educational testing where answers to exam questions for many students are all incorrect, diagnostic assessments where some subjects exhibit none of the measured symptoms, or marketing assessments where none of the measured products are purchased by some respondents. In all of these examples, interest is in assessing the latent dimensionality associated with items measured on an effected subset of the sample (e.g., students who have exceeded a level of mastery, patients in a pathological class, or consumers of a class of products.) We note for clarity that other authors have discussed the latent variable modeling of count responses with zero inflation (Magnus and Thissen, 2017), but our inquiry focuses on sets of dichotomous response items where there is a large number of all-zero responses.

When all-zero inflated responses arise from items that have high severity, most dimensionality assessment techniques are unaffected as long as the researcher includes all observations in the data set used for analysis; removing the all-zero observations is likely to result in incorrect assessment of dimensionality. For scenarios where high severity of items is the mechanism driving the large number of all-zero observations, either the BIC statistic in conjunction with the AZ-EFA model or one of the eigenvalue-based methods (the "Rule of 1" or the parallel analysis of the factor model's eigenvalues) can be recommended for identifying the latent dimensionality of the item set.

When all-zero inflated responses arise from a mixed-status sample where some subjects exhibit the trait and others do not, the BIC statistic in conjunction with the AZ-EFA (BIC_{AZ-EFA}) is again the recommended approach for identifying the latent dimensionality in an item set, with the parallel analysis of the factor model's eigenvalues also proving successful. In this case, removing the all-zero observations could render many standard dimensionality assessment methods viable, but such a strategy is ill-advised without knowing if it is a mixed-status sample or else just items with high severity. An additional reason to fit the AZ-EFA instead of a standard factor analysis model is that parameter estimates for the standard factor analysis model will exhibit extreme bias when drawing data from a community-based sample while the AZ-EFA parameter estimates are unbiased for the parameters associated with the pathological subset of the sample. We see in both the simulations involving mixed-status samples and the example analysis of the Social Anxiety Disorder symptoms from the NESARC data that the "Rule of 1" and common metrics such as RMSEA, CFI, and SRMR tend to erroneously indicate unidimensionality.

Using simulation studies, we have shown that for assessing latent dimensionality of a phenomenon that is exhibited by only a subset of respondents, an implementation of the AZ-EFA in concert with the BIC metric is superior to all commonly used dimensionality-assessment approaches. Parallel analysis of the factor model's eigenvalues is a reasonable (albeit slightly

inferior) alternative. However, when the number of pathological cases among the observations is expected to be small (e.g., less than 200), the true dimensionality of a phenomenon is likely to be understated using any of the methods considered here; AZ-EFA is a useful tool only when the data contain a moderate to large number of pathological subjects. Estimation of latent variable models to mixture distributions in small samples remains an area of open research. Regardless of the method for assessing dimensionality, the only approach for adequate estimation of model parameters (e.g., factor loadings) is the use of AZ-EFA, which yields factor loading estimates that do not exhibit the biases associated with the standard factor analysis model for dichotomous data. (See Supplemental Appendix S3 in the Supplemental Materials.)

A drawback of fitting the AZ-EFA in Mplus is the increased computational requirement, particularly when fitting more than two factors. For example, using a Mac Pro Intel(R) Xeon(R) CPU E5-1680 v2 @ 3.00GHzwith 64 GB of memory, fitting the one-, two-, and three-factor AZ-EFA model to the social anxiety disorder data from NESARC took 1, 25, and 84 minutes, respectively. Some of the scenarios in the simulation had average computation times for the three-factor model that approached 8 hours.

Applied to the NESARC social anxiety disorder (SAD) data, we have identified and interpreted three underlying dimensions of fear scenarios for SAD, rejecting the unidimensional characterization implied by both the commonly used dimensionality assessment approaches. Our assessment of the latent dimensionality of the SAD symptoms uses three factors: one associated with being placed in a "spotlight," another associated with small-group social interactions, and a third associated with heightened scrutiny or judgment in interpersonal settings.

An important implication of this research is that researchers need to be particularly careful in the ways they treat a community sample as compared to a patient sample. Improperly accounting for the mixture nature of community-based samples can lead to improper conclusions of unidimensionality. Further, historical assessments of unidimensionality for symptom batteries and other test item sets may warrant reevaluation when those assessments were based on community-based or other mixed-status sample data sets.

Acknowledgments

References

Bartholomew D.J., Knott M., & Moustaki I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach, Third Edition*. John Wiley & Sons.

Flora, D. B. & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, *9*(4), 466.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, *16*(4), 625-641.

Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, *76*(1), 5-21.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hukkelberg, S.S. & Jensen, T.K. (2011). The dimensionality of posttraumatic stress symptoms and their relationship to depression in children and adolescents, *Journal of Traumatic Stress*, *24*, 326-333.

Lei, P. W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, *43*(3), 495.

MacCoun, R.J. (2013). The puzzling unidimensionality of DSM-5 substance use disorder diagnoses, *Frontiers in Psychiatry*, *4*, 153. https://doi.org/10.3389/fpsyt.2013.00153

Magnus, B. E., & Thissen, D. (2017). Item response modeling of multivariate count data with zero inflation, maximum inflation, and heaping. *Journal of Educational and Behavioral Statistics*, *42*(5), 531-558.

Muthén, L.K. & Muthén, B.O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Muthén & Muthén.

Prigerson H.G., Horowitz M.J., Jacobs S.C., Parkes CM, Aslan M, Goodkin K, et al. (2009). Prolonged Grief Disorder: Psychometric validation of criteria proposed for *DSM-V* and *ICD-11*. *PLoS Med 6*(8). https://doi.org/10.1371/journal.pmed.1000121

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Saha, T.D., Chou, S.P., & Grant B.F. (2006). Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions, *Psychological Medicine*, *36*, 931–941.

Simon N., Wall M.M., Keshaviah A., Dryman M., LeBlanc N., & Shear K. (2011). Informing the symptom profile for Complicated Grief. *Depression and Anxiety*. *28*(2), 118-126.

Smits, N., Öğreden, O., Garnier-Villarreal, M., Terwee, C. B., & Chalmers, R. P. (2020). A study of alternative approaches to non-normal latent trait distributions in item response theory models used for health outcome measurement. *Statistical Methods in Medical Research*, *29*, 1030–1048.

Vermunt, J. K., & Magidson, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, advanced and syntax*. Statistical Innovations.

Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, *39*(8), 583–597. https://doi.org/10.1177/0146621615588184

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*(3), 392-423.

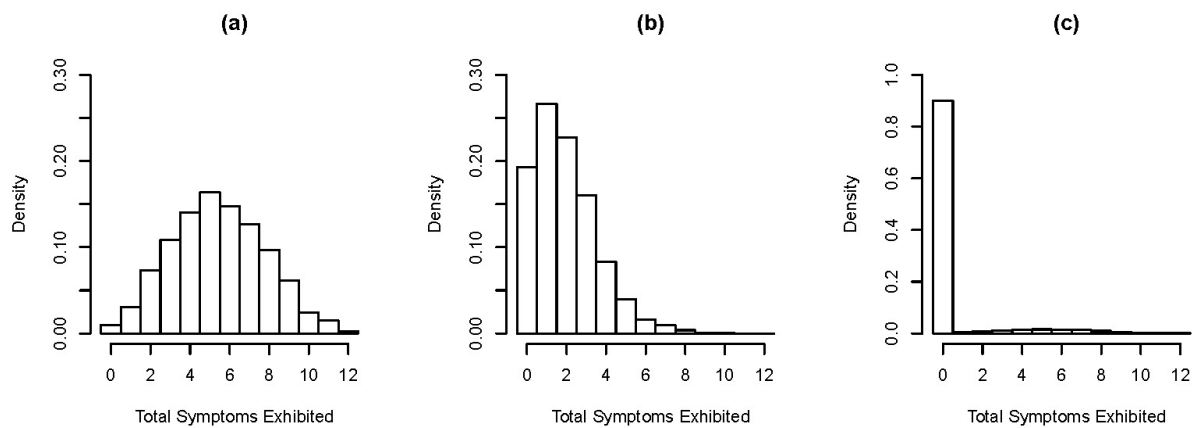Figure 1: Probability distributions for the total number of exhibited symptoms when generating data from: (a) Scenario 1, (b) Scenario 4, and (c) Scenario 1a (as described in Table 1). Note the different vertical scale in plot (c).
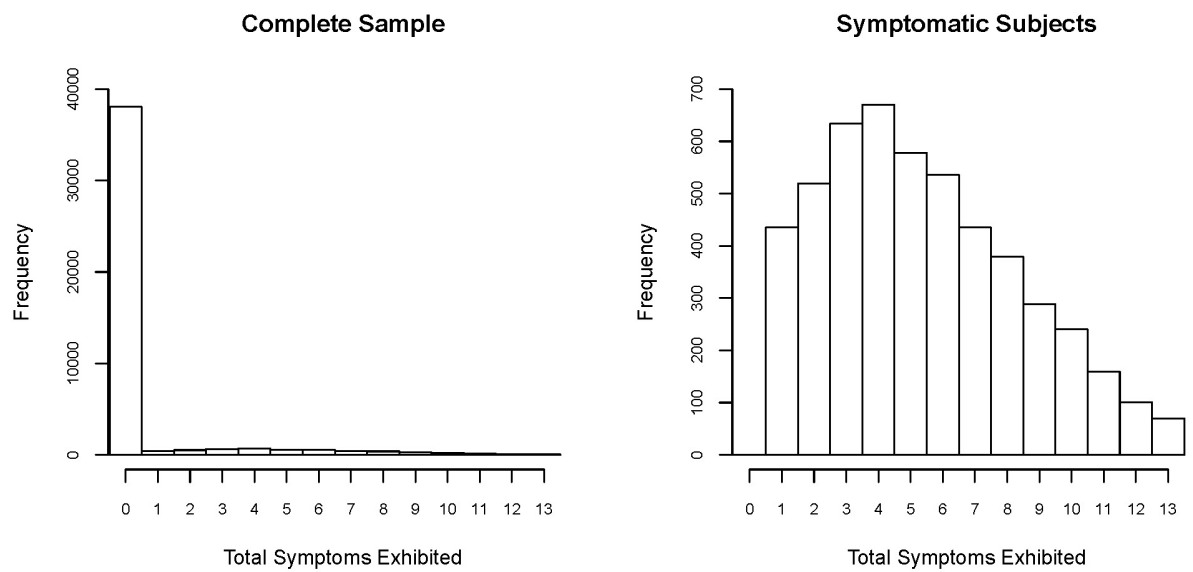

Figure 2: Total Social Anxiety Disorder symptoms exhibited within the complete NESARC data set (left), and among subjects exhibiting at least one symptom (right).

Table 1. Dimensionality assessments. For each scenario, the body of the table gives the percent of the cases where the method selects 1 factor, 2 factors, or 3+ factors. Results for the correct (2-factor) model are bolded. Each of the nine scenarios is characterized by the number of pathological cases in the data, the number of all-zero (i.e., non-pathological or healthy) cases added to the pathological sample, and the percent of the pathological cases that would be expected to be all zeros (as determined by the value of β_0 in equation (1)).

| | | | | Eigenvalue-based methods | | | | | | Traditional FA model (1) (fit with WLS) | | | | | | | | | Traditional FA model (2) (fit with ML) | | | | | | AZ-EFA (fit with ML) | | |
| | | | | "Rule of 1" (# evals > 1) | | | Parallel analysis (factor model) | | | X^2_{WLS} | | | RMSEA | | | CFI | | | X^2_P (Pearson) | | | BIC_{ML} | | | $BIC_{AZ\text{-}EFA}$ | | |
Scenario	# of pathological cases	# of all-zero cases added	% expected all-zero (β_0)	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+
1	4000	0	1% (0.25)	0	**100**	0	0	**96**	4	4	**92**	4	4	**96**	0	4	**96**	0	0	**92**	8	0	**100**	0	0	**100**	0
2	12000	0	1% (0.25)	0	**100**	0	0	**95**	5	2	**95**	3	2	**98**	0	2	**98**	0	0	**93**	7	0	**100**	0	0	**100**	0
3	20000	0	1% (0.25)	0	**100**	0	0	**99**	1	1	**99**	0	1	**99**	0	1	**99**	0	0	**92**	8	0	**100**	0	0	**100**	0
4	4000	0	20% (2.00)	0	**100**	0	0	**95**	5	3	**95**	2	57	**43**	0	3	**97**	0	0	**93**	7	0	**100**	0	0	**100**	0
5	12000	0	20% (2.00)	0	**100**	0	0	**94**	6	3	**94**	3	31	**69**	0	3	**97**	0	0	**85**	15	0	**100**	0	0	**100**	0
6	20000	0	20% (2.00)	0	**100**	0	0	**96**	4	1	**95**	4	22	**78**	0	1	**99**	0	0	**92**	8	0	**100**	0	0	**100**	0
1a	4000	36000	90.1% (0.25)	100	**0**	0	0	**100**	0	0	**93**	7	100	**0**	0	100	**0**	0	0	**0**	100	0	**5**	95	0	**100**	0
2a	12000	28000	70.3% (0.25)	85	**15**	0	0	**100**	0	0	**96**	4	100	**0**	0	100	**0**	0	0	**0**	100	0	**1**	99	0	**100**	0
3a	20000	20000	50.5% (0.25)	0	**100**	0	0	**100**	0	2	**95**	3	2	**98**	0	100	**0**	0	0	**0**	100	0	**56**	44	0	**100**	0

Table 2. Dimensionality assessments when removing the all-zero cases from the sample data. For each scenario, the body of the table gives the percent of the cases where the method selects 1 factor, 2 factors, or 3+ factors. Results for the correct (2-factor) model are bolded.

Scenario	# of path. cases	% expected all-zero among path. cases (β₀)	# of cases after removing all-zero cases	"Rule of 1" (# evals > 1)			Parallel analysis			X^2_{WLS}			RMSEA			CFI			X^2_P (Pearson)			BIC$_{ML}$		
				1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+	1	2	3+
1	4000	1% (0.25)	≈3960	0	**100**	0	0	**97**	3	3	**92**	5	3	**97**	0	3	**97**	0	0	**93**	7	0	**100**	0
2	12000	1% (0.25)	≈11880	0	**100**	0	0	**98**	3	3	**94**	3	3	**97**	0	3	**97**	0	0	**80**	20	0	**100**	0
3	20000	1% (0.25)	≈19800	0	**100**	0	0	**100**	0	1	**99**	0	1	**99**	0	1	**99**	0	0	**43**	57	0	**100**	0
4	4000	20% (2.00)	≈3200	0	**33**	67	5	**42**	53	64	**34**	2	99	**0**	0	94	**5**	0	37	**13**	50	100	**0**	0
5	12000	20% (2.00)	≈9600	1	**99**	0	0	**60**	40	8	**90**	2	100	**0**	0	100	**0**	0	0	**0**	100	100	**0**	0
6	20000	20% (2.00)	≈16000	2	**98**	0	0	**81**	19	5	**91**	4	100	**0**	0	100	**0**	0	0	**0**	100	95	**5**	0

Column group headers: Eigenvalue-based methods | Traditional FA model (1) (fit with WLS) | Traditional FA model (2) (fit with ML)

Table 3. Items in Social Anxiety Disorder inventory and associated AZ-EFA factor loadings when using the 3-factor AZ-EFA approach.

SAD Inventory Item	Factor 1 ("Spotlight")		Factor 2 ("Social")		Factor 3 ("Scrutiny")	
	estimate	(p-value)	estimate	(p-value)	estimate	(p-value)
2. Taking part/speaking in class	1	--	0	--	0	--
3. Taking part/speaking at a meeting	2.38	<0.001	0.37	0.321	0.06	0.875
4. Performing in front of people	1.19	<0.001	-0.01	0.963	0.59	0.006
1. Speaking in front of people	3.03	<0.001	0.90	0.002	-1.51	<0.001
9. Eating/drinking in front of other people	-0.36	0.012	1.54	<0.001	0.44	0.042
10. Having conversations with people you don't know well	0.31	0.091	2.54	<0.001	-0.55	0.007
11. Going to parties or social gatherings	0	--	1	--	0	--
13. Being in a small group situation	-0.37	0.022	1.69	<0.001	0.53	0.137
12. Dating	-0.64	<0.001	0.87	0.001	1.06	<0.001
6. Taking an important exam	-0.69	<0.001	-1.90	<0.001	3.68	<0.001
5. Being interviewed	0.33	<0.001	-0.40	0.011	1.87	<0.001
7. Speaking to an authority figure	0	--	0	--	1	--
8. Writing when someone watches	-0.34	0.004	0.14	0.482	1.25	<0.001

Table 4. Dimensionality assessment for the Social Anxiety Disorder symptom battery using the same metrics as those in the simulation summarized in Table 1. The 1-, 2-, and 3-factor models were fit using each method and the resulting statistic or p-value is given, with the bolded value indicating the number of factors chosen by each method.

	Eigenvalue-based methods							Traditional FA model (1) (fit with WLS)									Traditional FA model (2) (fit with ML)						AZ-EFA (fit with ML)		
	"Rule of 1" (# evals > 1)			Parallel analysis (factor model)				X^2_{WLS}			RMSEA			CFI			X^2_P (Pearson)			BIC_{ML}			$BIC_{AZ\text{-}EFA}$		
	1	2	3	1	2	3	4	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
All cases	x			x				<0.01	<0.01	**<0.01**	**0.02**	0.01	0.01	**0.99**	1.00	1.00	<0.01	<0.01	**<0.01**	103897	101278	**100597**	101551	100310	**100204**
Removing all-zero cases		x					x	<0.01	<0.01	**<0.01**	0.07	**0.05**	0.03	0.88	**0.96**	0.99	<0.01	<0.01	**<0.01**	66695	64959	**64544**	-	-	-