

Modeling Data with Measurement Errors but without Predefined Metrics: Fact versus Fallacy

Ke-Hai Yuan and Zhiyong Zhang

University of Notre Dame
kyuan@nd.edu

Abstract. Data in social and behavioral sciences typically contain measurement errors and also do not have predefined metrics. Structural equation modeling (SEM) is commonly used to analyze such data. This article discusses issues in latent variable modeling as compared to regression analysis with composite-scores. Via logical reasoning and analytical results as well as the analyses of two real datasets, several misconceptions related to bias and accuracy of parameter estimates, standardization of variables, and result interpretation are clarified. The results are expected to facilitate better understanding of the strength and limitations of SEM and regression analysis with weighted composites, and to advance social and behavioral data science.

Keywords: Measurement error · Attenuation · Standardization · Scales of latent variables

1 Introduction

Two key features of data in social and behavioral sciences are measurement errors and no predefined metrics. Associated with the features are latent variables whose scales need to be subjectively chosen. These features pose challenges to data analysis and result interpretation. A conventional method to address the issue of measurement errors is structural equation modeling (SEM), while standardized solution is used to address the issue of lack of metrics. In particular, textbooks contain formulas showing that the least-squares (LS) method yields attenuated or biased regression coefficients when predictors contain measurement errors, and SEM effectively addresses the issue. Textbooks on regression analysis and SEM also contain formulas for computing the regression coefficients with standardized variables (e.g., Bollen, 1989; Cohen, Cohen, West, & Aiken, 2003; Loehlin & Beaujean, 2017), which are available in the output of commonly used software and routinely reported in papers. Because the notions that *measurement errors cause biased estimates* and *standardized solutions facilitate result interpretation* have been imprinted and routinely taught in the discipline,

a rigorous examination on their validity not only facilitates better understanding of these concepts but also advances behavioral data science.

The purpose of this article is to bring attention of both quantitative and applied researchers to these potential issues, and for proper and better applications of multivariate methods. In particular, we aim to answer the following questions.

- Q1. In social and behavioral sciences, measurements and latent variables typically do not have predefined scales. We need to fix the scales of latent variables subjectively for the parameters of a model to be identified. Under SEM, what is the effect of different scaling options on the accuracy of parameter estimates and the related z -statistics? How can we use the information to serve our purpose?
- Q2. Do measurement errors cause attenuated or biased estimates for the LS method of regression analysis with weighted composites? A weighted composite in this article is a linear combination of values of indicators for a latent variable.
- Q3. Does SEM yield more accurate parameter estimates than LS regression with weighted composites?
- Q4. Between SEM and regression analysis with weighted composites, to what level the estimated regression coefficients can be compared with their SEM counterparts?
- Q5. Does standardization advance result interpretation or only facilitate model identification?

We will answer the above questions by combining recent findings from the literature, logical reasoning and analytical results, and fresh numerical results via the analyses of two real datasets. As we are going to show, results do not necessarily support the widely held notions regarding attenuation, bias, accuracy and efficiency of parameter estimates for SEM and regression analysis with composites that contain measurement errors.

Most existing studies comparing SEM and regression analysis with composites are conducted by comparing the values of parameter estimates and their standard errors (SEs). For logical and proper analyses of data that do not have predefined metrics, we propose a new approach under which methods are compared by the sizes of the signal-to-noise ratio (SNR) of their estimates. For a parameter estimate $\hat{\gamma}$ based on a sample of size N , the SNR is defined as

$$\tau = \frac{\gamma}{\text{SD}},$$

where γ and SD are respectively the expected values of $\hat{\gamma}$ and $[\text{Var}(\sqrt{N}\hat{\gamma})]^{1/2}$ or their probability limits as N increases. The new approach is a natural product of our effort in answering the above 5 questions.

The rest of the article is arranged as the following. First, we review the literature for related work and to clarify our contributions. Second, we describe our view on SEM and regression analysis with weighted composites, including key elements to answer the posed questions. Third, we provide a logical analysis

on the utility of standardization. Fourth, two empirical examples are provided and patterns over the results are summarized. Fifth, our answers to the above five questions are subsequently presented by combing the results from our analysis and the results of the examples. Sixth, summary, discussion and take-home messages are provided to conclude the article.

2 Review of the Literature and Clarification of Contributions

There are studies in the psychometric literature that might be regarded as related to the development of the current article. We will review them below to clarify the differences between the existing studies and the topics we are going to cover. Some of our results as well as the framework under which our study is conducted will also be previewed in this section.

2.1 Parameters and z -statistics are scale dependent

It is well known that, in SEM, the scale of a latent variable can be set by 1) fixing one of the loadings of its indicators at a given value, typically 1.0; or 2) for an independent latent variable, fixing its variance at a given value. The choices among the indicators as well as between 1) and 2) are equivalent in the sense that the resulting model implied covariance matrix remains the same. But they can yield quite different parameter estimates. Gonzalez and Griffin (2001) noted that different ways to scaling latent variables can also result in different z -statistics. This implies that the results of null hypothesis testing by the Wald test (or z -test for a single parameter) depend on how the scales of the latent variables are fixed. Subsequently, Gonzalez and Griffin recommended using the likelihood ratio statistic (T_{ml}) or its difference for parameter inference, because T_{ml} remains the same across different scalings of latent variables. However, one has to run a separate model to conduct the likelihood ratio test for each single parameter; whereas the z -statistics for all the parameters are in the default output of standard software following a single run of the base model. This might be why the z -test is widely used in practice. In addition, the validity of T_{ml} as a χ^2 statistic depends on the normal distribution assumption¹ even asymptotically. In contrast, SEs and the corresponding z -statistics based on the sandwich-type covariance matrix are asymptotically valid without the need for the normality assumption.

The sensitivity of z -statistics to the scales of latent variables reflects the dependency of statistical power of the Wald test on model parameterization. Instead of treating this sensitivity as an undesired feature, we should make use of it to serve the purpose of data analysis. In particular, if a test with a greater

¹ While there exist conditions for T_{ml} to follow a chi-square distribution when data are not normally distributed, there is not an effective way to verify the so-called asymptotic robustness conditions in practice.

power is desired, we can choose scales that correspond to the greatest z -statistics. Let's call an indicator whose loading is fixed at 1.0 an *anchor*. An analytical result in [Yuan and Fang \(2023b\)](#) implies that the SNR for the path coefficient between two latent variables increases as the anchor of the dependent latent variable becomes more reliable, where they assumed that all the indicators for the independent latent variable are parallel. A better understanding of the relationship between the z -statistics and the properties of the anchors is needed for the general case. For such a purpose, we will further study the following two characteristics of the z -statistics: 1) What properties of the anchors affect the value of z -statistics? and 2) Are the z -statistics for all path coefficients of the structural model equally affected by the changes of scales of the latent variables? These characteristics are not examined in [Gonzalez and Griffin \(2001\)](#).

2.2 Model identification versus theoretical assumption

[Steiger \(2002\)](#) discussed scenarios for fixing the scale of a latent variable using equality constraints, and fixing a factor loading at 1.0 is regarded as a particular constraint. He emphasized that additional constraints beyond the minimal need to fix the scales of latent variables will affect the value of T_{ml} . That is, the model implied covariance matrix will vary when different extra constraints are implemented. The same message has also been given by others (e.g., [Bentler, 2006](#)). In this article, we are not interested in the effect of extra constraints beyond the minimal need for scaling latent variables. Instead, for the scaling issue, we examine how the values of the SNRs and z -statistics are affected by the psychometric properties of the indicators used to fix the scales of latent variables. The value of T_{ml} remains the same among these choices. [Steiger \(2002\)](#) also discussed statistical issues due to interactions of extra constraints in standardized solutions. We will also discuss standardization but our interest is on issues related to substantive and statistical interpretations instead of issues caused by interactions of extra constraints.

2.3 Accuracy and precision of parameter estimates

For a mediation model with three latent variables, [Ledgerwood and Shrout \(2011\)](#) compared bias and SEs of parameter estimates between SEM and regression analysis via average scores. They used “accuracy” and “precision” to substitute for the statistical terminology bias and SEs, and showed that SEM yields estimates with greater accuracy but less precision. While [Ledgerwood and Shrout \(2011\)](#) contain several interesting observations, they missed two key points. The first is that values of parameters under SEM depend on the scales of the latent variables and those under regression analysis depend on the scales of the composites ([Yuan & Deng, 2021](#)). Thus, accuracy or bias is not a substantively grounded concept for statistical modeling of variables whose metrics are artificially assigned. The second is that SEs are typically proportional to the values of the parameter estimates, precision is also not a meaningful quantity

to compare between SEM and regression analysis with composite scores. Consequently, the conclusions of [Ledgerwood and Shrout \(2011\)](#) for the comparison between SEM and regression analysis with composites are problematic. In addition, the use of parallel indicators in their Monte Carlo studies also made their results of regression analysis with the average scores too optimistic, since the average score enjoys the maximum reliability ([Bentler, 1968](#); [Yuan & Bentler, 2002](#)). Existing results indicate that, following regression analysis with composites, the estimates of the regression coefficients, their SEs, and the resulting R-square are all related to the reliabilities of the composites ([Cochran, 1970](#); [Fuller, 2009](#)).

Instead of comparing different methods by the accuracy or precision of their parameter estimates, we compare methods via their SNRs. We will argue that the SNR is a natural quantity to compare for modeling variables without predefined metrics.

2.4 Factor score vs average score

[McNeish and Wolf \(2020\)](#) discussed rationales in forming composites and suggested treating sum scores as factor scores based on a factor model with parallel measurements. They also recommended factor-scoring items according to the factor model under which the scales are validated instead of using the sum scores by default. A followup discussion by [Widaman and Revelle \(2022\)](#) gave a different perspective on the merit of sum scores. They compared parameter estimates by different scoring methods and noticed little difference. In the current article, we are interested in comparing SEM and regression analysis with weighted composites, and regard both sum scores and factor scores as special cases of weighted composites. In particular, results on SNRs for the estimated path coefficients indicate that larger differences exist between SEM and regression analyses with weighted composites than among regression analyses with differently formulated composites ([Yuan & Fang, 2023b](#)).

2.5 Standardized score versus raw score

Variable standardization and treating the standardized coefficients as effect-size measures are common practices in social and behavioral sciences. Their pros and cons have been discussed under different contexts. Aiming to set out guidelines for what to report and how best to report effect sizes, [Baguley \(2009\)](#) listed advantages of simple (unstandardized) effect sizes but the measures need to have metrics that are well understood or substantiated if not predefined. [Pek and Flora \(2018\)](#) provided an informed discussion on why unstandardized effect sizes tend to be more informative than standardized ones in primary research studies. While their focus is on effect sizes with manifest variables, they stated (p. 214) “We agree that standardization of effects associated with latent variables (e.g., factors in a factor analysis) is useful, but assert that observed variables, and consequently effect sizes based on them, should not always be standardized.”

Kim and Ferree (1981) distinguished the operation of standardization of scales from the use of standardized coefficients. If the groups under consideration do not have comparable distributions, their discussion discourages standardizing variables on the basis of group-specific means and variances. Olejnik and Algina (2000) showed that measures of effect size are affected by the research design used, and warned that effect sizes may not be comparable across different designs when different random components (e.g., individual difference factors) are included in computing the pooled variances for standardizing the effect sizes. They also reviewed various factors that may contribute to the misinterpretation/understanding of effect size.

McGrath and Meyer (2006) discussed the differences of Cohen's d and the point-biserial correlation coefficient (r_{pb}). Both of which can be used when one variable is dichotomous and the other is quantitative. Termed the proportions of 0 and 1 for the dichotomous variable as the base rates, they showed that r_{pb} is a base-rate-sensitive effect-size measure, whereas d is base-rate-insensitive. Standardization is also widely used in epidemiology when estimating and comparing group means, where it is a different operation than z -scoring the variables. Still the group distributions matter in standardizing the means, as was showed by (Schoenbach & Rosamond, 2000, Chapter 6).

While the pros and cons of standardization have been extensively addressed, none of the articles discuss the issue of standardization in SEM. Part of our interest in this article is to examine the aspects of the usefulness of standardizing latent variables. In particular, our focus is on variables that do not have predefined metrics.

2.6 Properties and results of factor scores

Since factor scores will be repeatedly mentioned in our discussion, we briefly review their properties here. First, because latent variables are not observable, there exists an issue of indeterminacy with their scales and orientation. However, parameters of a factor or SEM model can be uniquely estimated once the scales of all latent variables are fixed and the model is identified. Then both the Bartlett-factor scores (BFSs) and the regression-factor scores (RFSs) based on the parameter estimates can be uniquely computed (Lawley & Maxwell, 1971). Unless explicitly mentioned, factor scores in this article always refer to either the BFSs or the RFSs.

It is well-known that the BFS possesses the maximal reliability among all weighted composites (see e.g., Yuan & Bentler, 2002). Yuan and Deng (2021) showed that the RFSs are proportional to the BFSs. That is, one can get the values of the BFSs from those of the RFSs via a linear transformation, conditional on the estimated factor loadings, factor covariances, and error variances. Thus, the RFSs also possess the maximum reliability. In addition, Yuan and Deng (2021) noted that the two types of factor scores are also equivalent in conducting regression analysis in the sense that they yield the same R-square value. Note that the RFSs can be computed jointly for all the factors or separately for each single factor. Yuan and Deng (2021) also noted that when the RFSs

are computed separately, the regression coefficients following RFS-regression are proportional to those following BFS-regression. When the RFSs are computed jointly, the two sets of regression coefficients can still be computed from each others but using a linear transformation.

Skrondal and Laake (2001) noted an important property of factor-score (FS) regression. That is, regression analysis with BFSs as the outcome variables and the jointly computed RFSs as the predictors yields path coefficients that are mathematically identical to those under SEM. This property was also discussed in Croon (2014) and Devlieger, Mayer, and Rosseel (2016), and will be noted in our following discussion.

2.7 Monte Carlo studies comparing parameter estimates

There are Monte Carlo studies comparing the empirical bias, SEs or mean-squared errors (MSE) of parameter estimates across methods (e.g., Forero & Maydeu-Olivares, 2009). For a model whose population values of parameters are held constant, the method that yields the smallest bias or SE or MSE is preferred under the conditions being considered. The finding is still statistically meaningful even when the variables do not have predefined metrics, as in typical simulation studies (e.g., Shi & Tong, 2017; Zhang & Yang, 2020). However, because bias, SE and MSE are scale dependent, additional thoughts are needed to compare the estimates under regression analysis against those under SEM, especially when variables do not have predefined metrics.

Note that bias is defined as the expected value of an estimator minus its population value. The regression coefficients under regression analysis with composites naturally don't have the same population values as their SEM counterparts. But for data that do not have predefined metrics, we can make the two sets of population values identical by properly choosing the scales for the composites or the latent variables. We will formally discuss how to compare estimates across the two classes of methods in the following sections.

3 Structural Equation Modeling versus Regression Analysis with Composites

In this section we present the elements for logically comparing SEM against regression analysis with weighted composites.

3.1 Theoretical constructs by latent variables versus by composites

While latent variables and composites are conceptually different, we regard both of them as representatives of the theoretical constructs. But their degrees of alignment are subject to judgment. It is commonly believed that theoretical constructs are virtually modeled under SEM while weighted composites always

contain measurement errors. However, the goodness of fit² of an SEM model is unlikely to be perfect, implying that the Greek letters in a path diagram only approximately represent the theoretical constructs. Such a discrepancy systematically changes the values of the parameters from those of an ideal model-population match (Yuan, Marshall, & Bentler, 2003), causing biased parameter estimates and biased interpretation. Similarly, the reliabilities of the composites also vary³ as the number of indicators vary, implying that the degree of alignment between the theoretical constructs and the weighted composites also varies. However, the bias due to model misspecification under SEM has been often ignored whereas “bias” caused by measurement errors has been repeatedly warned in textbooks (e.g., Allen & Yen, 1979).

In the development of this article, we do not explicitly consider the approximating nature of latent variable models. But we admit that there always exist differences between theoretical constructs and the latent variables in practice. Although the setup implicitly favors latent variable models, as was typically done in the field, we will discuss the rationale and provide the evidence that regression analysis with weighted composites yields different but more efficient parameter estimates than SEM instead of biased estimates.

3.2 Consistency

Measurements in social and behavioral sciences typically do not have pre-defined metrics in the first place. For such data, Yuan and Deng (2021) and Yuan and Fang (2023b) pointed out that the sizes of parameters under SEM do not enjoy a substantive interpretation since they are determined by the scales of latent variables that are subjectively assigned. Two researchers modeling the same dataset via the same model estimated by the same method (e.g., normal-distributed-based ML) can have very different parameter estimates for the same path coefficient. For this path coefficient, it is impossible for a third researcher who conducts regression analysis with composites to obtain an estimate that is consistent with those under SEM before the two SEM modelers have their difference dissolved. Thus, it does not make sense to claim that LS regression analysis with weighted composites yields biased estimates without a unique set of target population values even under SEM.

Let us consider a simple case to understand the details. With a measurement model

$$y = \lambda_y \eta + e_y, \quad x = \lambda_x \xi + e_x,$$

² Even if a test statistic for the overall model structure is statistically not significant, we are still unable to confirm that the model is correctly specified, since a non-significant statistic only implies that there is not enough power to reject the current model.

³ The content of the true score of a composite might also vary as more indicators are added even when a single factor model fits the corresponding variables adequately (Bentler, 2017).

one can estimate the regression relationship between the two latent variables via the structural model

$$\eta = \gamma_0 + \gamma_1\xi + \zeta,$$

where ξ and η are the latent variables, and e_x , e_y and ζ are the error terms. Alternatively, one can also directly work with the regression model for the observed variables

$$y = a + bx + e.$$

Under the commonly used assumptions about the independence of the error terms for SEM and using σ to denote the population covariance of the variables in its subscript, standard covariance algebra yields

$$b = \frac{\sigma_{xy}}{\sigma_{xx}} = \frac{\lambda_x \lambda_y \sigma_{\xi\eta}}{\lambda_x^2 \sigma_{\xi\xi} + \sigma_{e_x e_x}} = \frac{\rho_x \lambda_x \lambda_y \sigma_{\xi\eta}}{\lambda_x^2 \sigma_{\xi\xi}} = \frac{\rho_x \lambda_y \gamma_1}{\lambda_x}, \quad (1)$$

where ρ_x is the reliability of x . Equation (1) implies that, regardless of the value of ρ_x , $b = \gamma_1$ whenever $\lambda_x = \rho_x \lambda_y$ holds. We can also adjust the values of λ_x and λ_y by rescaling ξ and η to make the value of b greater than that of γ_1 . Alternatively, given the population values of ρ_x , σ_{xy} and σ_{xx} , the value of γ_1 can be made equal to any pre-specified value (except 0) by adjusting the value of λ_x and λ_y . We will further illustrate this via a real data example in a following section.

3.3 Parameter comparison

A common theme in [Ledgerwood and Shrout \(2011\)](#), [McNeish and Wolf \(2020\)](#), [Widaman and Revelle \(2022\)](#), and others is the comparison of the raw values of parameter estimates by different methods. In this article we emphasize that the population values of parameters as well as their estimates are not of substantive interest for models involving latent variables measured by indicators that do not have predefined metrics. In particular, population values of the model parameters depend on scales that are artificially assigned.

Although there is no point to directly compare the values of parameters under regression analysis with weighted composites against those under SEM, they have the following relationship:

- 1) $\gamma = \mathbf{0}$ if and only if $\gamma_w = \mathbf{0}$, where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)'$ and $\gamma_w = (\gamma_{w1}, \gamma_{w2}, \dots, \gamma_{wp})'$ are the vectors of regression coefficients of a given dependent latent variable under SEM and regression analysis with weighted composites, respectively ([Buonaccorsi, 2010](#), Eq. 5.7 on page 109).
- 2) When the (joint) regression-factor scores are used as the predictors and the Bartlett-factor score is used as the dependent variable in regression analysis with weighted composites, there also exist $\gamma_w = \gamma$ and $\hat{\gamma}_w = \hat{\gamma}$ ([Skrondal & Laake, 2001](#)). Then the two types of estimates can be substituted for each other, although they are obtained by different methods.

- 3) Regression analysis with equally weighted composites can also yield parameter estimates that are mathematically identical to those under SEM according to

$$\begin{aligned} y &= a + b_1x_1 + \dots + b_px_p + e \\ &= a + (h_1b_1)(x_1/h_1) + \dots + (h_pb_p)(x_p/h_p) + e, \end{aligned} \quad (2)$$

where the y and x_j s are equally weighted composites and the h_j s are subject to choice (e.g., $h_j = \gamma_j/b_j$).

However, for arbitrarily chosen scales, b_j and γ_j may not be equal. They even can have different signs.

Instead of judging the goodness of an estimate by its accuracy or precision, we propose to compare the efficiency of parameter estimates by the size of their SNR⁴, which plays a key role in statistical inference. Let's term this proposal the new framework in contrast to the old framework that compares methods by precision and accuracy of parameter estimates, the following remarks are in order.

- 4) The new framework allows us to face and address the issue of dependency for the values of parameters or their estimates on the scales of latent variables, especially for data and variables that have no predefined metrics.
- 5) The new framework naturally facilitates the comparison of efficiencies of estimates for parameters that do not have the same population value. For example, the population value of a path coefficient under SEM depends on how the scales of the involved latent variables are fixed. However, the SNRs for the estimates of this path coefficient can be compared across different scaling options, and the one with the largest SNR is the most efficient estimate.
- 6) When the population value of a parameter is uniquely defined across methods, as in a Monte Carlo study where the scales of latent variables are identically scaled across methods, an estimate with a smaller SE is more efficient and corresponds to a greater SNR than that with a greater SE. We will further discuss the issue of empirical bias in Monte Carlo studies in the concluding section.
- 7) Parallel to Cohen's d , the SNR also serves as a summary statistic.

In addition, strengths of SEM and regression analysis with composites can be fairly compared and utilized under the new framework. We will have more results on this point via examples in a following section.

For a model with one dependent latent variable and one independent latent variable, [Yuan and Fang \(2023b\)](#) rigorously compared the SNRs of the estimated path coefficients under regression analyses with weighted composites against those under SEM. They found that, conditional on the population weights, the SNR under factor score (FS) regression is mathematically greater than that under SEM. They also defined a multivariate version of SNR and conjectured

⁴ For a parameter estimate, the SNR is estimated by $z/N^{1/2}$, where z is the z -statistic and N is the sample size.

that its values for the path coefficients under FS regression would be greater than that under SEM. Note that the SNR plays the role of Cohen’s d in null hypothesis testing for parameters. Meta analytical results in [Deng and Yuan \(2023\)](#) showed that, across nine different real datasets and eleven models, SEM yields the least powerful test, even weaker than path analysis with equally weighted composites.

3.4 Different utilities

SEM and regression analysis with weighted composites are different not only in their approaches to modeling the theoretical constructs but also in aims and utilities. For SEM, the relationship among the latent variables is modeled. The corresponding parameters and their estimates are to govern the relationship among the latent variables at the *population* level. In practice, an *individual* with greater pretest scores is expected to perform better on the post-tests. This expected relationship is of interest in many disciplines. We may want to plug the path coefficients estimated under SEM in the regression equation to predict the values of individuals corresponding to the latent variable. For such a purpose, we will have to substitute the independent latent variables by composites of the individuals. However, except in rare situations, values of composites are not error-free, including the factor scores that are psychometrically most reliable.

Alternatively, we can start with regression analysis via weighted composites, and use the estimates of the path coefficients to construct an equation for prediction. The new outcome variable is then predicted according to this equation using the newly observed scores of the independent variables via weighted composites. [Fuller \(2009\)](#) noted that, even when the independent variables contain measurement errors, LS estimates of the regression model still yield the best linear unbiased predictor in the sense that the corresponding MSE is the smallest. [Yuan and Fang \(2023a\)](#) contain the details showing that the predicted value based on the SEM estimates becomes less accurate as the reliabilities of the weighted composites decrease.

Thus, one should start by regression analysis with weighted composites if the purpose is for prediction. However, SEM is preferred if the aim is to describe the relationship among the latent variables at the population level. These different characterizations might be more fundamental than which method generates more accurate parameter estimates.

The analyses and discussions in this section contain our answers to questions Q2, Q3 and Q4.

4 Standardization and Bias-correction

The notion that standardized solutions facilitate result interpretation has been rooted in psychometrics, especially with latent variable modeling and regression analysis. In this section, we conduct a logical analysis on the utility of standardization. We will also discuss the utility of bias correction. For such a purpose,

we distinguish measurements that have predefined metrics from those that have no predefined metrics.

To facilitate understanding of the issue, let's consider the relationship between height ξ (inches) and weight η (pounds), and they are assumed to follow the relationship

$$\eta = \gamma_0 + \gamma_1\xi + \zeta, \quad (3)$$

where ζ is the error term in predicting weight by height. In practice, we only observe x (inches) and y (pounds) due to the deficiencies in technology. A reasonable and also logical measurement model⁵ in this case is

$$x = \xi + \delta \text{ and } y = \eta + \varepsilon, \quad (4)$$

where δ and ε are measurement errors, and they are statistically independent with the latent variables ξ and η . With a sample (x_i, y_i) of size N , if we estimate the model

$$y_i = \gamma_{*0} + \gamma_{*1}x_i + e_i \quad (5)$$

by the LS method, then the LS estimate $\hat{\gamma}_{*1}$ is expected to be smaller than the γ_1 of Equation (3). In the measurement error literature (e.g., Fuller, 2009), emphasis was placed on getting a consistent estimate of γ_1 by correcting $\hat{\gamma}_{*1}$. Let the corrected estimate be denoted by $\tilde{\gamma}_1$. Then the value of $\tilde{\gamma}_1$ provides us the information that, with one inch increase in height, a person is expected to increase by $\tilde{\gamma}_1$ pounds.

Let's standardize the variables in Equations (3) and (4), resulting in

$$\eta_s = \gamma_{s1}\xi_s + \zeta_s, \quad x_s = \lambda_x\xi_s + \delta_s, \quad y_s = \lambda_y\eta_s + \varepsilon_s. \quad (6)$$

By the first equation in Equation (6), we would conclude that an individual with an increase of one SD (inches) in height is expected to increase by γ_{s1} SD (pounds) in weight. Such standardized scales might not prevent us from understanding the relationship between height and weight if we are familiar with the two SD units. However, if the SDs are as inexplicable as the values of ξ or η , then standardization does not facilitate interpretation but only facilitates identifying a set of unique values of the γ and λ .

In the case of height and weight, the use of standardized scales certainly hinders our understanding of their relationship. When there are no established metrics for x and y , the values of their SDs are at least as inexplicable as the values of x and y themselves. Standardization may even block possible attempts to think about the issue because it is hard to get a sense out of the SDs even in the case of height and weight, and each SD depends on the distribution shape as well as the range of the variable.

For height and weight, the bias-corrected estimate $\tilde{\gamma}_1$ does provide a more accurate quantification for the relationship between the two variables. When the

⁵ When repeated measurements on height and weight for each individual are available, a typical practice is to substitute the x and y in Equation (4) by their respective averages.

scales of x or y are arbitrary, as for typical variables in social and behavioral sciences where data are obtained via Likert items or the averages/sums of such items, parameters under standardized scales do not advance the understanding of the relationship of the involved variables. Bias-corrected estimates may not help with better substantive interpretation either.

Standardized regression coefficients across groups might be comparable if the ranges and distribution shapes of the groups are similar. Otherwise, equal standardized coefficients may still imply different relationships between the outcome variable and the predictors in separate groups.

5 Real-data Examples

In this section we use two real data examples to illustrate some of the points noted in the previous sections. Because the value of a z -statistic is simply the value of the SNR multiplied by the square root of the sample size⁶, comparison of the z -statistics between different methods will be directly followed from our comparison of the SNRs.

5.1 Example 1

Data Mardia, Kent, and Bibby (1979; Table 1.2.1) contain test scores on 5 topics from $N = 88$ students. The five topics are: C_1 =Mechanics, C_2 =Vectors, O_1 =Algebra, O_2 =Analysis, and O_3 =Statistics. The scores for the first two topics were obtained with closed-book exams and for the last three were with open-book exams. Tanaka, Watadani, and Ho Moon (1991) fitted the dataset by a two-factor model, one factor represents the trait for taking closed-book tests, and the other for taking open-book tests. This dataset has been used to illustrate new developments in SEM and other multivariate methods (e.g., Cadigan, 1995). We will use it to show how different methods perform in estimating the regression parameter between the two constructs. Because this dataset is open to public, we expect readers to easily replicate our results.

The path diagram for the two-factor model is given in Figure 1, where ξ_o and ξ_c represent the latent traits for taking the open- and closed-book tests, respectively. Let $\phi_o = \text{Var}(\xi_o)$ and $\phi_c = \text{Var}(\xi_c)$. Fitting the model implied covariance matrix with $\phi_o = 1.0$ and $\phi_c = 1.0$ to the sample covariance matrix by normal-distribution-based maximum likelihood (NML) yields $T_{ml} = 2.073$, indicating that the model fits the data very well when referred to χ_4^2 . The parameter estimates, their SEs, and the corresponding z -statistics for the confirmatory factor model are reported in Table 1. The reliabilities of the individual indicators estimated via the factor model are also included in Table 1 and so are those of the two Bartlett-factor scores (BFS_c , BFS_o).

⁶ The values of the SNR in Tables 2, 3 and 5 are obtained by dividing the z -statistics by $(N - 1)^{1/2}$, where N is the original sample size.

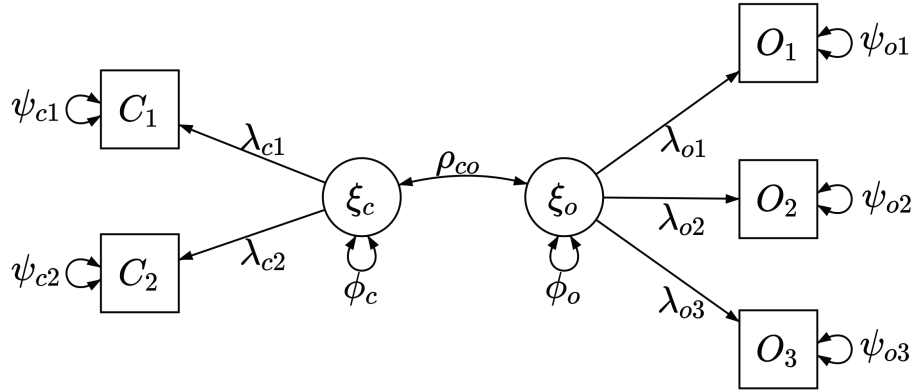


Figure 1. A two-factor model for the open- and closed-book test dataset.

Model For illustration purpose, let's consider the following two structural models under SEM

$$\xi_c = \gamma_{co}\xi_o + \zeta_c, \text{ and } \xi_o = \gamma_{oc}\xi_c + \zeta_o. \quad (7)$$

That is, we predict the latent trait for the closed-book test by that for the open-book test, and the latent trait for the open-book test by that for the closed-book test, respectively. Note that our purpose here is to illustrate the properties of different methods rather than to testify the causal directions of the two traits. Actually, the two models in Equation (7) are mathematically equivalent to the confirmatory factor model in Figure 1 with respect to the overall model structure.

Parallel to the two structural models in Equation (7), we also estimate the following regression models

$$\hat{\xi}_c = \gamma_{*co}\hat{\xi}_o + e_c, \text{ and } \hat{\xi}_o = \gamma_{*oc}\hat{\xi}_c + e_o \quad (8)$$

by the LS method, where $\hat{\xi}_c$ and $\hat{\xi}_o$ are composite-scores. There are many ways to formulate composite-scores, we will only consider equally-weighted composites (EWC), the BFSs and the RFSs in the study. Note that the EWCs are least selective among all composites since they don't use any of the psychometric properties of the individual indicators, whereas the two types of factor scores are most selective since they optimally use these properties. Also note that both the sum scores and the simple averages are special cases of EWCs.

For the structural models in Equation (7), we need to fix the scales of ξ_o and ξ_c in order for the models to be identified. There are 6 different options to identify each model by fixing two factor loadings at 1.0; 2 options to identify the model $\xi_o \rightarrow \xi_c$ via fixing $\phi_o = 1.0$; and 3 options to identify the model $\xi_c \rightarrow \xi_o$ via fixing $\phi_c = 1.0$. Thus, there are 8 different sets of scalings to identify the model $\xi_o \rightarrow \xi_c$; and 9 different sets of scalings to identify the model $\xi_c \rightarrow \xi_o$. The overall model remains equivalent among these different scalings, with T_{ml} being the same as that for the confirmatory factor model in Figure 1.

Table 1. Estimates (Est) of factor loadings (λ), error variances (ψ), factor correlation (ρ), and reliability (Rel) of the indicators (Ind) and Bartlett-factor scores (BFS) under the common factor model.

| Param. | Est | SE | z | Param. | Est | SE | z | Ind/BF | Rel Est |
|----------------|--------|-------|--------|-------------|---------|--------|-------|------------------|---------|
| λ_{c1} | 12.253 | 1.843 | 6.649 | ψ_{c1} | 155.632 | 31.679 | 4.913 | C1 | .491 |
| λ_{c2} | 10.383 | 1.379 | 7.530 | ψ_{c2} | 65.036 | 18.099 | 3.593 | C2 | .624 |
| λ_{o1} | 9.834 | 0.929 | 10.588 | ψ_{o1} | 16.186 | 7.261 | 2.229 | O1 | .857 |
| λ_{o2} | 11.490 | 1.403 | 8.192 | ψ_{o2} | 88.352 | 16.773 | 5.268 | O2 | .599 |
| λ_{o3} | 12.517 | 1.667 | 7.508 | ψ_{o3} | 141.074 | 24.881 | 5.670 | O3 | .526 |
| ρ_{co} | 0.818 | 0.073 | 11.258 | | | | | BFS _c | .724 |
| | | | | | | | | BFS _o | .896 |

Table 2. Estimates of the path coefficient, its standard deviation (SD) and the corresponding signal-to-noise ratio (SNR) by SEM, factor-score (FS) regression and equally-weighted-composite (EWC) regression for the relationship $\xi_o \rightarrow \xi_c$.

| Method | Identification | Est. | SD | SNR |
|---------|--|--------|--------|--------------|
| SEM | $\lambda_{c1} = 1.0, \lambda_{o1} = 1.0$ | 1.019 | 1.676 | 0.608 |
| | $\lambda_{c1} = 1.0, \lambda_{o2} = 1.0$ | 0.872 | 1.528 | 0.571 |
| | $\lambda_{c1} = 1.0, \lambda_{o3} = 1.0$ | 0.800 | 1.459 | <u>0.548</u> |
| | $\lambda_{c2} = 1.0, \lambda_{o1} = 1.0$ | 0.863 | 1.216 | 0.710 |
| | $\lambda_{c2} = 1.0, \lambda_{o2} = 1.0$ | 0.739 | 1.132 | 0.653 |
| | $\lambda_{c2} = 1.0, \lambda_{o3} = 1.0$ | 0.678 | 1.094 | 0.620 |
| | $\phi_o = 1.0, \lambda_{c1} = 1.0$ | 10.019 | 16.818 | 0.596 |
| | $\phi_o = 1.0, \lambda_{c2} = 1.0$ | 8.490 | 12.286 | 0.691 |
| FS-reg | $\lambda_{o1} = 1.0, \lambda_{c1} = 1.0$ | | | |
| | BFS(ξ_o) & BFS(ξ_c) | 0.912 | 1.055 | 0.865 |
| | RFS(ξ_o) & RFS(ξ_c) | 0.738 | 0.853 | 0.865 |
| | BFS(ξ_o) & RFS(ξ_c) | 0.661 | 0.764 | 0.865 |
| | RFS(ξ_o) & BFS(ξ_c) | 1.019 | 1.178 | 0.865 |
| | $\phi_o = 1.0, \lambda_{c1} = 1.0$ | | | |
| | BFS(ξ_o) & BFS(ξ_c) | 8.973 | 10.377 | 0.865 |
| | RFS(ξ_o) & RFS(ξ_c) | 7.253 | 8.388 | 0.865 |
| EWC-reg | sum(ξ_o) & sum(ξ_c) | 0.428 | 0.587 | 0.730 |
| | ave(ξ_o) & ave(ξ_c) | 0.643 | 0.880 | 0.730 |
| | sum(ξ_o) & ave(ξ_c) | 0.214 | 0.293 | 0.730 |
| | ave(ξ_o) & sum(ξ_c) | 1.285 | 1.760 | 0.730 |

Note. BFS=Bartlett-factor score, RFS=regression-factor score; sum=sum score, ave=average score; the estimate with the largest SNR under SEM is in bold while the one with the smallest SNR is underlined.

Results The NML estimates of γ for the two models in Equation (7) are given in the upper panel of Tables 2 and 3, respectively. The standard deviation (SD) of each $\hat{\gamma}$ and the corresponding SNR are also included in the tables. Clearly, the value of $\hat{\gamma}$ under SEM changes as the scalings vary. But we do not regard their differences as problematic because their population counterparts are different, and each $\hat{\gamma}$ is consistent and efficient for a different γ (assuming data are normally distributed). As a matter of fact, we can choose the scale of ξ_o or that of ξ_c to make γ_{co} or γ_{oc} to equal any pre-specified (nonzero) value while the test statistic for the overall model structure remains at $T_{ml} = 2.073$.

Table 3. Estimates (Est) of the path coefficient, its standard deviation (SD) and the corresponding signal-to-noise ratio (SNR) by SEM, factor-score (FS) regression and equally-weighted-composite (EWC) regression for the relationship $\xi_c \rightarrow \xi_o$.

| Method | Identification | Est | SD | SNR |
|---------------------------------|--|--------|--------|--------------|
| SEM | $\lambda_{c1} = 1.0, \lambda_{o1} = 1.0$ | 0.656 | 1.101 | 0.596 |
| | $\lambda_{c1} = 1.0, \lambda_{o2} = 1.0$ | 0.767 | 1.433 | 0.535 |
| | $\lambda_{c1} = 1.0, \lambda_{o3} = 1.0$ | 0.835 | 1.614 | <u>0.517</u> |
| | $\lambda_{c2} = 1.0, \lambda_{o1} = 1.0$ | 0.774 | 1.266 | 0.612 |
| | $\lambda_{c2} = 1.0, \lambda_{o2} = 1.0$ | 0.905 | 1.656 | 0.546 |
| | $\lambda_{c2} = 1.0, \lambda_{o3} = 1.0$ | 0.986 | 1.868 | 0.528 |
| | $\phi_c = 1.0, \lambda_{o1} = 1.0$ | 8.040 | 10.424 | 0.771 |
| | $\phi_c = 1.0, \lambda_{o2} = 1.0$ | 9.395 | 14.421 | 0.651 |
| | $\phi_c = 1.0, \lambda_{o3} = 1.0$ | 10.235 | 16.500 | 0.620 |
| FS-reg | $\lambda_{c1} = 1.0, \lambda_{o1} = 1.0$ | | | |
| | BFS(ξ_c) & BFS(ξ_o) | 0.475 | 0.549 | 0.865 |
| | RFS(ξ_c) & RFS(ξ_o) | 0.588 | 0.680 | 0.865 |
| | BFS(ξ_c) & RFS(ξ_o) | 0.425 | 0.492 | 0.865 |
| | RFS(ξ_c) & BFS(ξ_o) | 0.656 | 0.759 | 0.865 |
| | $\phi_c = 1.0, \lambda_{o1} = 1.0$ | | | |
| | BFS(ξ_c) & BFS(ξ_o) | 5.821 | 6.732 | 0.865 |
| | RFS(ξ_c) & RFS(ξ_o) | 7.201 | 8.328 | 0.865 |
| | BFS(ξ_c) & RFS(ξ_o) | 5.213 | 6.029 | 0.865 |
| RFS(ξ_c) & BFS(ξ_o) | 8.040 | 9.299 | 0.865 | |
| EWC-reg | sum(ξ_c) & sum(ξ_o) | 0.824 | 1.128 | 0.730 |
| | ave(ξ_c) & ave(ξ_o) | 0.549 | 0.752 | 0.730 |
| | sum(ξ_c) & ave(ξ_o) | 0.275 | 0.376 | 0.730 |
| | ave(ξ_c) & sum(ξ_o) | 1.648 | 2.257 | 0.730 |

Note. BFS=Bartlett-factor score, RFS=regression-factor score; sum=sum score, ave=average score; the estimate with the largest SNR under SEM is in bold while the one with the smallest SNR is underlined.

In Tables 2 and 3, the largest SNR under SEM was put in bold and the smallest was underlined. The largest SNR in Table 2 corresponds to the condition when both ξ_o and ξ_c are anchored by the most reliable indicators, whereas the

largest SNR in Table 3 corresponds to the condition when ξ_c is scaled by $\phi_c = 1.0$. In both Tables 2 and 3, the smallest SNR under SEM corresponds to the conditions when both ξ_o and ξ_c are anchored by the least reliable indicators.

The middle panel of Table 2 contains the estimates of γ_{*co} for the first regression model in Equation (8), where $\hat{\xi}_o$ and $\hat{\xi}_c$ are factor scores computed following the NML estimates of the parameters under the identification rules $\lambda_{o1} = \lambda_{c1} = 1.0$ and $\phi_o = \lambda_{c1} = 1.0$, respectively. Parallel results for the second regression model in Equation (8) are displayed in the middle panel of Table 3. In particular, the SNRs by the LS method for the two regression models in Equation (8) have the same value. To save space, we did not include the results of FS regression corresponding to all sets of scalings of the two latent variables, while their values of the SNR remain to be 0.865. For each identification condition, there is also a $\hat{\gamma}_*$ that has the same value as its SEM counterpart (i.e., 1.019 and 10.019 in Table 2, and 0.656 and 8.040 in Table 3), verifying the noted result by Skronrdal and Laake (2001).

The lower panels of Tables 2 and 3 contain the LS estimates for the regression models in Equation (8) when $\hat{\xi}_o$ and $\hat{\xi}_c$ are the sum and/or average scores. They are denoted by EWC regression (EWC-reg) in the tables. Clearly, the values of the estimates of γ_{*co} and γ_{*oc} for the regression models in Equation (8) depend on the scales of the composites and so do their corresponding SDs. However, unlike SEM, the values of the SNR (as well as the corresponding z -statistic) under FS regression or EWC regression remain the same across different scalings. Note that the SNRs under EWC regression are smaller than those under FS regression, because the sum scores are not as reliable as the factor scores.

In Table 2, all the eight SNRs for $\hat{\gamma}$ by SEM are smaller than those for $\hat{\gamma}^*$ by EWC regression and by FS regression. In Table 3, all the nine SNRs for $\hat{\gamma}$ by SEM are smaller than that for $\hat{\gamma}_*$ by FS regression; and only one SNR for $\hat{\gamma}$ by SEM is larger than that for $\hat{\gamma}^*$ by EWC regression. Thus, using SNR as a measure for the efficiency of parameter estimates, among the 17 options for identifying the two SEM models, only in one option SEM outperforms EWC regression. None of the 17 SEM identification options yields a greater SNR than FS regression.

Results in Tables 2 and 3 also illustrate the fact that an LS estimate under regression analysis with weighted composites does not have to be smaller than its counterpart under SEM. Another interesting fact is that, unlike in regression analysis under which the SNRs for the coefficients of $x \rightarrow y$ and $y \rightarrow x$ are the same, the SNR under SEM differentiates the path coefficients between $\xi_o \rightarrow \xi_c$ and $\xi_c \rightarrow \xi_o$ even if the two latent variables are scaled by the same set of anchors (e.g., $\lambda_{c1} = \lambda_{o1} = 1.0$).

The results for FS regression in Tables 2 and 3 were obtained by treating the factor scores as the observed variables, which is widely used in practice (DiStefano, Zhu, & Mindrila, 2009; Widaman & Revelle, 2022). But parameter estimates used to compute the factor scores contain sampling errors, which affect the SEs of the resulting $\hat{\gamma}_*$. Results in Yuan and Fang (2023b) indicate that the SEs of the FS regression coefficients by considering the sampling errors in

the estimated weights tend to be smaller than those by treating weights as being given, and FS regression may become even more powerful in detecting the existence of a relationship if the sampling errors in weights are accounted.

5.2 Example 2

Data Table 2 of [Weston and Gore \(2006\)](#) contains a sample covariance matrix for a dataset with $N = 403$ cases and $p = 12$ variables. The dataset was part of a survey of college students who participated in a vocational psychology research project. With three indicators for each construct, the 12 variables are respectively measures of 1) self-efficacy beliefs, 2) outcome expectations, 3) career-related interests, and 4) occupational considerations. Weston and Gore Jr considered two structural models. [Deng and Yuan \(2023\)](#) compared the values of z -statistics of parameter estimates for each model by different methods, where each latent variable was scaled by only one option. We consider only one of their models, and our purpose here is to see how the SNRs react to different options for scaling the latent variables.

Model The path diagram in Figure 2 corresponds to the first model of [Weston and Gore \(2006\)](#), which posits that the effect of self-efficacy beliefs on career-related interests is partially mediated by outcome expectations, while the effect of self-efficacy beliefs on occupational considerations is completely through the two mediator variables (outcome expectations & career-related interests). The structural model has four path coefficients: γ_{11} , γ_{21} , β_{21} , β_{32} .

For each latent variable in Figure 2, we can select one of the three factor loadings and fix it at 1.0 to anchor its scale. So by factor loadings alone there are $3^4 = 81$ different sets of scalings to identify the model. For the independent latent variable self-efficacy beliefs (ξ_1), we can also fix its variance at 1.0, which provides additional $3^3 = 27$ different sets of scalings to identify the model. With a total of $81 + 27 = 108$ ways of model identification, we will only study a subset of them to illustrate our point, and the selected subsets allow us to see how and what parameters are affected by the properties of the anchors.

Results Letting $\phi_{11} = \text{Var}(\xi_1) = 1.0$ and $\lambda_{y_11} = \lambda_{y_42} = \lambda_{y_73} = 1.0$, fitting the model in Figure 2 to the vocational-psychology dataset by NML results in $T_{ml} = 416.061$, which corresponds to a p-value that is essentially 0 when referred to χ_{50}^2 . With CFI=.913, and RMSEA=.135, the model might not be regarded as fitting the data adequately although it is substantively derived (see [Weston and Gore, 2006](#) and references therein). Such a discrepancy between theory and goodness of model-fit is not unusual in empirical modeling, reflecting our earlier observation that the theoretical constructs may not be perfectly⁷ represented by

⁷ When letting all the four latent variables be freely correlated in Figure 2, we have $T_{ml} = 361.848$, which corresponds to a p-value that is essentially 0 when referred to χ_{48}^2 , indicating that discrepancy between the theoretical constructs and the latent variables also exists in the measurement model.

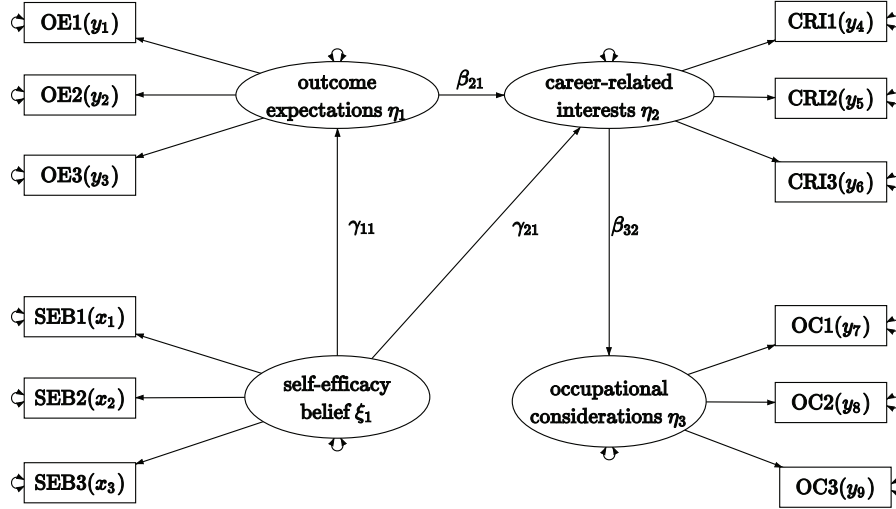


Figure 2. A mediated model for self-efficacy belief on occupational considerations (Weston and Gore, 2006, $N = 403$).

Table 4. Parameter estimates (Est), their SEs (SE) and z -statistics for the model in Figure 2 ($p = 12$, $N = 403$, $T_{ml} = 416.061$, $df = 50$, p -value=.000; RMSEA=.135, and CFI=.913). The right column is the reliability (ρ) of the 12 indicators.

| Param. | Est | SE | z | Param. | Est | SE | z | Rel | Est |
|------------------|--------|-------|--------|----------------------|-------|-------|--------|-------------|-------|
| $\lambda_{x1,1}$ | 2.381 | 0.111 | 21.467 | ψ_{x1} | 1.840 | 0.179 | 10.264 | ρ_{x1} | 0.755 |
| $\lambda_{x2,1}$ | 2.365 | 0.108 | 21.903 | ψ_{x2} | 1.628 | 0.167 | 9.768 | ρ_{x2} | 0.775 |
| $\lambda_{x3,1}$ | 2.402 | 0.104 | 23.145 | ψ_{x3} | 1.186 | 0.148 | 8.013 | ρ_{x3} | 0.830 |
| $\lambda_{y1,1}$ | 1.000 | | | ψ_{y1} | 0.882 | 0.078 | 11.318 | ρ_{y1} | 0.808 |
| $\lambda_{y2,1}$ | 0.976 | 0.030 | 32.930 | ψ_{y2} | 0.325 | 0.048 | 6.827 | ρ_{y2} | 0.916 |
| $\lambda_{y3,1}$ | 0.993 | 0.032 | 30.628 | ψ_{y3} | 0.580 | 0.060 | 9.621 | ρ_{y3} | 0.863 |
| $\lambda_{y4,2}$ | 1.000 | | | ψ_{y4} | 0.044 | 0.003 | 12.853 | ρ_{y4} | 0.394 |
| $\lambda_{y5,2}$ | 1.144 | 0.094 | 12.229 | ψ_{y5} | 0.027 | 0.002 | 11.300 | ρ_{y5} | 0.580 |
| $\lambda_{y6,2}$ | 1.011 | 0.098 | 10.326 | ψ_{y6} | 0.049 | 0.004 | 12.974 | ρ_{y6} | 0.372 |
| $\lambda_{y7,3}$ | 1.000 | | | ψ_{y7} | 0.795 | 0.100 | 7.918 | ρ_{y7} | 0.835 |
| $\lambda_{y8,3}$ | 0.963 | 0.040 | 24.212 | ψ_{y8} | 1.350 | 0.126 | 10.705 | ρ_{y8} | 0.734 |
| $\lambda_{y9,3}$ | 0.795 | 0.033 | 23.934 | ψ_{y9} | 0.962 | 0.089 | 10.867 | ρ_{y9} | 0.725 |
| γ_{11} | 1.186 | 0.096 | 12.364 | | | | | | |
| γ_{21} | 0.046 | 0.009 | 5.161 | $\sigma_{\zeta_1}^2$ | 2.302 | 0.211 | 10.913 | | |
| β_{21} | 0.057 | 0.006 | 10.003 | $\sigma_{\zeta_2}^2$ | 0.008 | 0.001 | 5.542 | | |
| β_{32} | 10.368 | 0.822 | 12.615 | $\sigma_{\zeta_3}^2$ | 0.964 | 0.151 | 6.396 | | |

Table 5. Values of the signal-to-noise ratio (SNR) of $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ for the model in Figure 2 when ξ_1 , η_1 , η_2 , and η_3 are anchored by fixing one of their loadings (λ) at 1.0, or by letting $\phi_{11} = \text{Var}(\xi_1) = 1.0$.

| anchors of | SNR | | | | anchors of | SNR | | | |
|--|---------------------|---------------------|--------------------|--------------------|-----------------------------------|---------------------|---------------------|--------------------|--------------------|
| $\xi_1, \eta_1, \eta_2, \eta_3$ | $\hat{\gamma}_{11}$ | $\hat{\gamma}_{21}$ | $\hat{\beta}_{21}$ | $\hat{\beta}_{32}$ | η_1, η_2, η_3 | $\hat{\gamma}_{11}$ | $\hat{\gamma}_{21}$ | $\hat{\beta}_{21}$ | $\hat{\beta}_{32}$ |
| <u>x_1, y_1, y_4, y_7</u> | 0.631 | 0.258 | 0.499 | 0.629 | ($\phi_{11} = 1.0$) | | | | |
| x_2, y_1, y_4, y_7 | 0.636 | 0.259 | 0.499 | 0.629 | | | | | |
| x_3, y_1, y_4, y_7 | 0.646 | 0.259 | 0.499 | 0.629 | <u>y_1, y_4, y_7</u> | 0.617 | 0.257 | 0.499 | 0.629 |
| x_1, y_2, y_4, y_7 | 0.652 | 0.258 | 0.507 | 0.629 | y_2, y_4, y_7 | 0.636 | 0.257 | 0.507 | 0.629 |
| x_1, y_3, y_4, y_7 | 0.642 | 0.258 | 0.504 | 0.629 | y_3, y_4, y_7 | 0.626 | 0.257 | 0.504 | 0.629 |
| x_1, y_1, y_5, y_7 | 0.631 | 0.267 | 0.569 | 0.785 | y_1, y_5, y_7 | 0.617 | 0.266 | 0.569 | 0.785 |
| x_1, y_1, y_6, y_7 | 0.631 | 0.257 | 0.489 | 0.609 | y_1, y_6, y_7 | 0.617 | 0.256 | 0.489 | 0.609 |
| x_1, y_1, y_4, y_8 | 0.631 | 0.258 | 0.499 | 0.607 | y_1, y_4, y_8 | 0.617 | 0.257 | 0.499 | 0.607 |
| x_1, y_1, y_4, y_9 | 0.631 | 0.258 | 0.499 | 0.605 | y_1, y_4, y_9 | 0.617 | 0.257 | 0.499 | 0.605 |
| <u>x_2, y_2, y_5, y_8</u> | 0.656 | 0.267 | 0.582 | 0.743 | | | | | |
| x_1, y_2, y_5, y_8 | 0.652 | 0.267 | 0.582 | 0.743 | | | | | |
| x_3, y_2, y_5, y_8 | 0.668 | 0.268 | 0.582 | 0.743 | <u>y_2, y_5, y_8</u> | 0.636 | 0.266 | 0.582 | 0.743 |
| x_2, y_1, y_5, y_8 | 0.636 | 0.267 | 0.569 | 0.743 | y_1, y_5, y_8 | 0.617 | 0.266 | 0.569 | 0.743 |
| x_2, y_3, y_5, y_8 | 0.646 | 0.267 | 0.577 | 0.743 | y_3, y_5, y_8 | 0.626 | 0.266 | 0.577 | 0.743 |
| x_2, y_2, y_4, y_8 | 0.656 | 0.259 | 0.507 | 0.607 | y_2, y_4, y_8 | 0.636 | 0.257 | 0.507 | 0.607 |
| x_2, y_2, y_6, y_8 | 0.656 | 0.257 | 0.496 | 0.589 | y_2, y_6, y_8 | 0.636 | 0.256 | 0.496 | 0.589 |
| x_2, y_2, y_5, y_7 | 0.656 | 0.267 | 0.582 | 0.785 | y_2, y_5, y_7 | 0.636 | 0.266 | 0.582 | 0.785 |
| x_2, y_2, y_5, y_9 | 0.656 | 0.267 | 0.582 | 0.739 | y_2, y_5, y_9 | 0.636 | 0.266 | 0.582 | 0.739 |
| <u>x_3, y_3, y_6, y_9</u> | 0.657 | 0.258 | 0.493 | 0.587 | | | | | |
| x_1, y_3, y_6, y_9 | 0.642 | 0.257 | 0.493 | 0.587 | | | | | |
| x_2, y_3, y_6, y_9 | 0.646 | 0.257 | 0.493 | 0.587 | <u>y_3, y_6, y_9</u> | 0.626 | 0.256 | 0.493 | 0.587 |
| x_3, y_1, y_6, y_9 | 0.646 | 0.258 | 0.489 | 0.587 | y_1, y_6, y_9 | 0.617 | 0.256 | 0.489 | 0.587 |
| x_3, y_2, y_6, y_9 | 0.668 | 0.258 | 0.496 | 0.587 | y_2, y_6, y_9 | 0.636 | 0.256 | 0.496 | 0.587 |
| x_3, y_3, y_4, y_9 | 0.657 | 0.259 | 0.504 | 0.605 | y_3, y_4, y_9 | 0.626 | 0.257 | 0.504 | 0.605 |
| x_3, y_3, y_5, y_9 | 0.657 | 0.268 | 0.577 | 0.739 | y_3, y_5, y_9 | 0.626 | 0.266 | 0.577 | 0.739 |
| x_3, y_3, y_6, y_7 | 0.657 | 0.258 | 0.493 | 0.609 | y_3, y_6, y_7 | 0.626 | 0.256 | 0.493 | 0.609 |
| x_3, y_3, y_6, y_8 | 0.657 | 0.258 | 0.493 | 0.589 | y_3, y_6, y_8 | 0.626 | 0.256 | 0.493 | 0.589 |

Note. The underlined row in each block is for reference against which the other lines of the block are compared.

the Greek letters in Figure 2. But the discrepancy between the model and data has little to do with our illustration, since the results are essentially the same even if the model fits the data perfectly (i.e., letting the sample covariance matrix equal the model implied covariance matrix). For reference, Table 4 contains the estimates of the factor loadings (λ), error variances (ψ), the path coefficients of the structural model, and the variances of the three prediction errors (σ_ζ^2). The last column of Table 4 indicate that x_3 is the most reliable indicator for ξ_1 while y_2 , y_5 and y_7 are the most reliable indicators for η_1 , η_2 and η_3 , respectively. All the parameter estimates in Table 4 are statistically significant at the level of .05.

Table 5 contains the values of the SNR for the four path coefficients under 48 different identification conditions (out of 108 options). Results on the left side of the table are obtained when one of the loadings of ξ_1 is fixed at 1.0, while those on the right side are obtained by letting $\phi_{11} = \text{Var}(\xi_1) = 1.0$. Note that the values of the four parameter estimates vary across the 48 sets of scalings, while $T_{ml} = 416.061$. Our main interest with this example is the pattern of the SNRs while they vary with the parameter estimates when the latent variables are scaled differently.

There are 6 blocks of results in Table 5, and each block has one set of scalings underlined, which serves the condition for reference. In particular, for each set of scalings within a given block, only one of the four latent variables is rescaled compared to the reference condition. The results in Table 5 exhibit the following patterns.

- 1) When ξ_1 is anchored by x_1 , x_2 , x_3 or by $\phi_{11} = 1.0$, the SNRs for $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ are not affected by the scale change of ξ_1 . In addition, results not included in the table also indicate that the values of $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ as well as their SDs are not affected by the scale change of ξ_1 either. This is because the paths represented by β_{21} and β_{32} are not directly connected with ξ_1 in Figure 2. However, the SNRs or equivalently the z -statistics for both $\hat{\gamma}_{11}$ and $\hat{\gamma}_{21}$ become greater when ξ_1 is anchored by a more reliable indicator.
- 2) In Figure 2, the paths represented by γ_{21} and β_{32} are not directly connected with η_1 . When η_1 is anchored by different indicators, the SNRs for $\hat{\gamma}_{21}$ and $\hat{\beta}_{32}$ are not affected. However, the SNRs for $\hat{\gamma}_{11}$ and $\hat{\beta}_{21}$ become greater as the anchor of η_1 is more reliable.
- 3) In Figure 2, the path represented by γ_{11} is not directly connected with η_2 . When η_2 is anchored by different indicators, the SNR for $\hat{\gamma}_{11}$ is not affected by the scale change of η_2 . However, the SNRs for $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ become greater as η_2 is anchored by a more reliable indicator.
- 4) For the same reason, the SNRs for $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$ and $\hat{\beta}_{21}$ are not affected by the scale change of η_3 . The SNR for $\hat{\beta}_{32}$ becomes greater as η_3 is anchored by a more reliable indicator.

The results in Table 5 suggested that, the SNR or z -statistic for a parameter estimate is invariant to the scale changes of the latent variables that are not directly connected with the path that the parameter represents. In contrast, the SNR for a parameter estimate becomes greater when the directly connected latent variables are anchored by more reliable indicators. In particular, the greatest

SNRs for $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\gamma}_{32}$ are respectively $\text{SNR}_{\gamma_{11}} = .668$, $\text{SNR}_{\gamma_{21}} = .268$, $\text{SNR}_{\beta_{21}} = .582$, and $\text{SNR}_{\beta_{32}} = .785$. They are simultaneously obtained when all the latent variables are anchored by indicators with the greatest reliability (i.e., x_3 for ξ_1 , y_2 for η_1 , y_5 for η_2 , and y_7 for η_3).

Note that, while the values of $\hat{\gamma}$, $\hat{\beta}$ and SNR change when different indicators are used as anchors, the value of the SNR under SEM will remain the same once the anchors are chosen regardless of the particular values of the factor loadings. That is, $\lambda_{o1} = 1.0$ or $\lambda_{o1} = 2.3$ leads to the same SNR in Tables 2 and 3. Similarly, the value of the SNR (or z -statistic) remains the same once the scale of ξ_c is determined by fixing the value of ϕ_c regardless of its particular value, e.g., $\phi_c = 1.0$ or $\phi_c = 3.5$ corresponds to the same SNR. More systematic results in this direction are presented in Yuan, Ling, and Zhang (2024).

The results of the two examples provide the fact for answering questions Q1, Q2 and Q3.

6 Answers to Questions Q1 to Q5

Our analysis and results might have already answered the questions posed in the introduction of the article. As a summary, we will answer them directly in this section. For clarity, we will also include the original questions.

Q1. *Under SEM, what is the effect of different scaling options on the accuracy of parameter estimates and the related z -statistics? How can we use the information to serve our purpose?*

When a latent variable is anchored by an indicator with greater reliability, the SNRs and consequently the z -statistics for path coefficients that are directly related to the latent variables are expected to be greater. However, estimates of the path coefficients that are not directly related to the latent variables are not affected nor their SEs. Also, scaling an independent latent variable by fixing its variance at 1.0 may result in even greater SNRs. Thus, under SEM, we can obtain more efficient parameter estimates of path coefficients by selecting more reliable anchors for latent variables.

Q2. *Do measurement errors cause attenuated or biased estimates for the LS method of regression analysis with weighted composites?*

Measurement errors alone do not cause biased or attenuated regression coefficients. It is the artificially chosen scales that make the path coefficients under regression analysis different from those under SEM.

Q3. *Does SEM yield more accurate parameter estimates than LS regression with weighted composites?*

SEM does not yield more accurate parameter estimates than LS regression with weighted composites. When measured by SNR, it is more likely the other way around. That is, even regression analysis with EWCs may yield more efficient estimates of path coefficients than SEM, especially when the indicators for each latent variable are approximately parallel.

Q4. *Between SEM and regression analysis with weighted composites, to what level the estimated regression coefficients can be compared with their SEM counterparts?*

There are three levels that the path coefficients under regression analysis can be compared with their counterparts under SEM: (1) $\gamma_w = \mathbf{0}$ if and only if $\gamma = \mathbf{0}$, where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)'$ and $\gamma_w = (\gamma_{w1}, \gamma_{w2}, \dots, \gamma_{wp})'$ are respectively the path coefficients of a given dependent variable under SEM and regression analysis with weighted composites. Inference for one set of parameters can be done by a statistical test on the other set of parameters, simultaneously. (2) $\gamma = \gamma_w$ when a BFS is used as the dependent variable and the joint RFSs are used as the independent variables in FS regression. Then, the two sets of parameters as well as their respective estimates can be substituted for each other. (3) Regardless of the scales chosen for the variables under each modeling technique, the size of the multivariate SNR determines the statistical power in testing $\gamma = \mathbf{0}$, and FS regression is expected to correspond to a greater multivariate SNR and consequently to be more powerful than SEM.

Q5. *Does standardization advance result interpretation or only facilitate model identification?*

Whether the measurements have predefined metrics or not, standardization only facilitates model identification. Unless standard deviation is a widely used and well-understood unit in a given context, standardization does not advance our understanding nor facilitates interpretation of the parameter estimates in a SEM or regression model.

7 Discussion and Conclusion

We studied several widely circulated notions in modeling data with measurement errors but without predefined metrics. While modeling such data poses many challenges, SEM still offers important information that regression analysis with composite scores is unable to provide. In particular, SEM gives a platform to assess the goodness of the overall model structure, unidimensionality of different subscales, individual indicator reliability, and measurement invariance for group comparison, etc. However, while inference on parameter estimates under SEM can be statistically sound, substantive interpretation on the size of a path coefficient becomes a challenge if the measurements do not have predefined or well-understood scales to start with.

Because path coefficients under both SEM and regression analysis with composites depend on subjectively assigned scales of the involved variables, there is no point to demand regression analysis to yield estimates that are consistent with those under SEM. For parameters whose population values depend on subjectively assigned scales, a natural criterion for comparing their estimates is SNR, which is parallel to Cohen's d and serves as an index for the efficiency of the estimate. When all the path coefficients of a dependent variable are considered simultaneously, FS regression is expected to correspond to a greater (multivariate) SNR than SEM. If the indicators for each latent variable do not greatly

deviate from parallel, EWC regression is also expected to outperform SEM with respect to efficiency of the estimated regression coefficients.

In addition to numerical differences between estimates by different methods, parameters under SEM are to govern the relationship of variables representing the population, where individuals are treated equally (i.e., a random representation). In practice, individuals with greater pretest scores are expected to perform better on the outcome variable, and parameters under regression analysis with weighted composites are to govern such a relationship. In particular, conditional on the metrics of the observed and latent variables, the predicted values according to the uncorrected LS estimates of the regression model still have the smallest MSE even if the target is the latent-outcome variable and the pretest scores are not error-free. But the SNR and R^2 of the regression model as well as the MSE of the predicted value are related to the size of measurement errors. More reliable composites correspond to more accurate estimates of path coefficients, greater R^2 values, and smaller prediction errors.

We have shown that standardization of latent or manifest variables is useful for model identification, not necessarily advancing our understanding of the relationship among the involved variables nor better interpretation of the parameters of the model either. That said, we do not exclude a context under which the distribution of a variable can be well understood in a standardized scale with mean 0 and variance 1.0. For example, we might transform the distribution of IQ (ξ) by $z_\xi = (\xi - 100)/15$ and the value of z_ξ allows us to judge the standing of the corresponding ξ in percentile according to the standard normal distribution $N(0, 1)$. However, this might belong to the case where the scale of the measurement was known in advance.

Bias can be easily defined and explained for parameter estimates in modeling variables that have predefined metrics. Empirical bias can also be easily evaluated in Monte Carlo studies even when data do not have predefined metrics. However, it is not clear how to interpret bias substantively if the scales of the variables need to be subjectively assigned. In particular, two researchers who conduct SEM analyses can have very different parameter estimates for the same path coefficient while they both are consistent/unbiased. Similarly, one researcher can choose the sum scores while another can choose the average scores in regression analyses, and they have identical t -statistics and R^2 . But their parameter estimates are different. More generally, suppose Researcher A gets an estimate $\hat{\gamma}_a$ while Researcher B get an estimator $\hat{\gamma}_b = c\hat{\gamma}_a$, where $c > 0$ is a constant. We are unable to compare the two estimators with respect to bias since (if needed) Researcher B can always rescale his estimator to $\tilde{\gamma}_b = \hat{\gamma}_b/c$ if the involved variables do not have predefined metrics. Regardless, we recommend the one with a greater SNR, and $\tilde{\gamma}_b$ and $\hat{\gamma}_b$ are equivalent in the sense that they have the same SNR.

The first take-home message from this article is that sizes of parameter estimates and their SEs are not meaningful quantities for models involving (latent or manifest) variables that do not have predefined metrics, and SNR is a logical and also a natural measure of efficiency of parameter estimates. For the same rea-

son, the MSE of parameter estimates is not a logical criterion to compare across methods unless the population values of the parameters are held constant among the methods or when all the involved variables are on the same metrics. When the estimands become equal, the most efficient and accurate estimates have the greatest SNR. The 2nd take-home message is that standardization does not advance interpretation but offers a way to avoid dealing with the issues of lack of metrics. Effort needs to be made to develop substantively rationalized metrics under which parameter estimates are interpreted.

Acknowledgments

This work was supported by a grant from the Department of Education (R305D210023). However, the contents of the study do not necessarily represent the policy of the funding agencies, and you should not assume endorsement by the Federal Government. Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

References

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks-Cole.
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617. doi: <https://doi.org/10.1348/000712608x377117>
- Bentler, P. M. (1968). Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33*(3), 335–345. doi: <https://doi.org/10.1007/bf02289328>
- Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.
- Bentler, P. M. (2017). Specificity-enhanced reliability coefficients. *Psychological Methods*, *22*(3), 527–540. doi: <https://doi.org/10.1037/met0000092>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.
- Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications*. Boca Raton, FL: Chapman & Hall.
- Cadigan, N. G. (1995). Local influence in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(1), 13–30. doi: <https://doi.org/10.1080/10705519509539992>
- Cochran, W. G. (1970). Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association*, *65*(329), 22–34. doi: <https://doi.org/10.1080/01621459.1970.10481059>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale NJ: Lawrence Erlbaum Associates.

- Croon, M. (2014). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Psychology Press.
- Deng, L., & Yuan, K.-H. (2023). Which method is more powerful in testing the relationship of theoretical constructs? a meta comparison of structural equation modeling and path analysis with weighted composites. *Behavior Research Methods*, *55*(3), 1460–1479. doi: <https://doi.org/10.3758/s13428-022-01838-z>
- Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*(5), 741–770. doi: <https://doi.org/10.1177/0013164415607618>
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11. doi: <https://doi.org/10.7275/DA8T-4G52>
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of irt graded response models: Limited versus full information methods. *Psychological Methods*, *14*(3), 275–299. doi: <https://doi.org/10.1037/a0015825>
- Fuller, W. A. (2009). *Measurement error models*. John Wiley & Sons.
- Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every “one” matters. *Psychological Methods*, *6*(3), 258–269. doi: <https://doi.org/10.1037/1082-989x.6.3.258>
- Kim, J.-O., & Ferree, G. D. (1981). Standardization in causal analysis. *Sociological Methods & Research*, *10*(2), 187–210. doi: <https://doi.org/10.1177/004912418101000203>
- Lawley, D. N., & Maxwell, M. A. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworths.
- Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*(6), 1174–1188. doi: <https://doi.org/10.1037/a0024776>
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). New York: Routledge.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.
- McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d . *Psychological Methods*, *11*(4), 386–401. doi: <https://doi.org/10.1037/1082-989x.11.4.386>
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. doi: <https://doi.org/10.3758/s13428-020-01398-0>
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational*

- Psychology*, 25(3), 241–286. doi: <https://doi.org/10.1006/ceps.2000.1040>
- Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, 23(2), 208–225. doi: <https://doi.org/10.1037/met0000126>
- Schoenbach, V. J., & Rosamond, W. D. (2000). *Understanding the fundamentals of epidemiology: an evolving text*. Chapel Hill:: University of North Carolina.
- Shi, D., & Tong, X. (2017). The impact of prior information on bayesian latent basis growth model estimation. *SAGE Open*, 7(3), 215824401772703. doi: <https://doi.org/10.1177/2158244017727039>
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66(4), 563–575. doi: <https://doi.org/10.1007/bf02296196>
- Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, 7(2), 210–227. doi: <https://doi.org/10.1037/1082-989x.7.2.210>
- Tanaka, Y., Watadani, S., & Ho Moon, S. (1991). Influence in covariance structure analysis: with an application to confirmatory factor analysis. *Communications in Statistics - Theory and Methods*, 20(12), 3805–3821. doi: <https://doi.org/10.1080/03610929108830742>
- Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719–751. doi: <https://doi.org/10.1177/0011000006286345>
- Widaman, K. F., & Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, 55(2), 788–806. doi: <https://doi.org/10.3758/s13428-022-01849-w>
- Yuan, K.-H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, 67(2), 251–259. doi: <https://doi.org/10.1007/bf02294845>
- Yuan, K.-H., & Deng, L. (2021). Equivalence of partial-least-squares sem and the methods of factor-score regression. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(4), 557–571. doi: <https://doi.org/10.1080/10705511.2021.1894940>
- Yuan, K.-H., & Fang, Y. (2023a). Replies to comments on “which method delivers greater signal-to-noise ratio: Structural equation modelling or regression analysis with weighted composites?” by yuan and fang (2023). *British Journal of Mathematical and Statistical Psychology*, 76(3), 695–704. doi: <https://doi.org/10.1111/bmsp.12323>
- Yuan, K.-H., & Fang, Y. (2023b). Which method delivers greater signal-to-noise ratio: Structural equation modelling or regression analysis with weighted composites? *British Journal of Mathematical and Statistical Psychology*, 76(3), 646–678. doi: <https://doi.org/10.1111/bmsp.12293>
- Yuan, K.-H., Ling, L., & Zhang, Z. (2024). Scale-invariance, equivariance and dependency of structural equation models. *Struc-*

tural Equation Modeling: A Multidisciplinary Journal, 1–16. doi: <https://doi.org/10.1080/10705511.2024.2353168>

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). 8. assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology*, 33(1), 241–265. doi: <https://doi.org/10.1111/j.0081-1750.2003.00132.x>

Zhang, Q., & Yang, Y. (2020). Autoregressive mediation models using composite scores and latent variables: Comparisons and recommendations. *Psychological Methods*, 25(4), 472–495. doi: <https://doi.org/10.1037/met0000251>