MULTIVARIATE BEHAVIORAL RESEARCH
2025, VOL. 60, NO. 3, 598–619

Estimated Factor Scores Are Not True Factor Scores

Mijke Rhemtulla[a]* and Victoria Savalei[b]*

[a]Department of Psychology, University of California, Davis, Davis, CA, USA; [b]Department of Psychology, University of British Columbia, Vancouver, BC, Canada

ABSTRACT

In this tutorial, we clarify the distinction between estimated factor scores, which are weighted composites of observed variables, and true factor scores, which are unobservable values of the underlying latent variable. Using an analogy with linear regression, we show how predicted values in linear regression share the properties of the most common type of factor score estimates, regression factor scores, computed from single-indicator and multiple indicator latent variable models. Using simulated data from 1- and 2-factor models, we also show how the amount of measurement error affects the reliability of regression factor scores, and compare the performance of regression factor scores with that of unweighted sum scores.

KEYWORDS

Factor scores; latent variables; sum scores

A central feature of structural equation models (SEMs) is their ability to represent abstract constructs (e.g., neuroticism, life satisfaction, executive function) as error-free latent variables, or factors, each measured by a set of unreliable observed variables, or indicators (e.g., scale items). When they are correctly specified, SEMs produce asymptotically unbiased estimates of the associations (e.g., correlations, regression coefficients) among latent variables in the model. In contrast, simply adding up indicators of each construct and then modeling the relations among the resulting scale scores is suboptimal because scale scores will still contain measurement error, and thus lack perfect reliability (Bollen, 1989; Cole & Preacher, 2014).[1] To obtain highly reliable scale scores, it is necessary to have either (a) highly reliable indicators, or (b) a very large number of indicators with lower reliability. For example, one can achieve scale score reliability of 0.95 by having 2 items that each have reliability of 0.9, or by having 50 items with reliability of 0.27 (Spearman, 1910). Latent variables, in contrast, contain no measurement error and are perfectly reliable. But while this difference between scale scores and latent variables is well-known, some confusion remains about the extent to which the properties of latent variables carry over to *estimated* scores on latent variables, obtained from observed data.

The goal of this article is to explain the difference between estimated factor scores, which are weighted composites of observed variables, and actual or true factor scores, which are unobservable values of the underlying latent variable. While many technical sources discuss this distinction (e.g., Waller, 2023), non-technical and intuitive presentations are lacking. Anecdotally, these concepts are often confused. For example, in a recent tweet by a researcher asking how to get the values of the latent variables that generated the observed data when using the `simulateData` function in the R package **lavaan** (Rosseel, 2012), the great majority of replies provided instead answers for how to get *estimated* factor scores from the generated data.[2] Contributing to the confusion, the term "factor score" is frequently used in the published literature without a clear definition: It may refer to the estimated factor scores (e.g., DiStefano et al., 2009; Grice, 2001; Skrondal & Laake, 2001), or to the individual's true standing on the latent factor (e.g., Velicer, 1976; Waller, 2023), or to both, blurring the distinction between them. In this article, we will use the terms "estimated factor

scores" and "true factor scores" to distinguish these two concepts.[3] Following estimation of parameters of a latent variable model, scores on the latent variable can be estimated using a variety of methods (e.g., McDonald & Burr, 1967), but the most common is the so-called regression approach. We will refer to factor score estimates obtained in this way as "regression factor scores."[4]

At the heart of the distinction between true and estimated factor scores is the very concept of a latent variable. In classical test theory (CTT; Crocker & Algina, 1986), each observed score is hypothesized to be a sum of the true score plus measurement error. The measurement error is assumed to be completely random and uncorrelated with everything (including itself across repeated assessments). For example, participants may sometimes circle an unintended answer due to momentary distraction. The true score is the hypothetical score that would be obtained if measurement could be repeated infinitely often, canceling out this random error in the long run.[5] Because it can never be observed, CTT's true score represents the earliest invocation of a latent variable, but in a narrow sense because the "latent variable" could be the true length of a table or a participant's true standing on a single questionnaire item.

The common factor model is a generalization of CTT where the latent variable, sometimes referred to as the common score, now symbolizes the underlying true level of a broader construct that gives rise to scores on multiple observed indicators, or items, which measure specific consequences of being high or low on this construct. Each observed variable is still a sum of the latent variable (scaled by the factor loading, capturing that different items have different sensitivities to changes in the latent construct) plus error. The error term, while still often referred to as "measurement error", now captures the unique variance in each item, which includes completely random measurement noise plus any systematic component of a response to a given item that is not due to the latent variable and is not shared with other items. For example, a response to an item about appetite on a depression scale may partly due in differences in metabolism, reflecting the systematic part of a response to that item that is unrelated to depression. Other latent variable models are generalizations of the common factor model, containing multiple latent variables.

For the *i*th individual, their true factor score, f_i, is simply their score on the latent variable f. Because the variable is latent, its values for any individual are unknowable and cannot be obtained from the observed indicators; if they could be, it would no longer be a latent variable. On the other hand, the *i*th individual's *estimated* factor score, $\hat{f}_i$, is a score on a composite observed variable $\hat{f}$ that is constructed as a weighted linear combination of the observed indicators of f, where the weights are functions of the estimated model parameters (e.g., factor loadings and residual variances). As the sample size increases, the weights become more precisely estimated, and $\hat{f}$ becomes a weighted linear combination of observed variables with known weights, rather than estimated weights. But a linear combination of observed variables cannot reproduce the scores on a latent variable, even if the weights are precisely known. Estimated factor scores will generally not equal true factor scores: $\hat{f}_i \neq f_i$, so long as none of the observed indicators are free of measurement error. Unlike sampling error, measurement error does not go away with increasing sample size.

Methods for estimating factor score estimates differ in how their weights are computed from the model parameters. The most common type, regression factor scores, have the property of *maximal reliability* in the population. That is, if the model parameters are exactly known, regression factor scores get "closest" to the true factor scores in the sense that the squared correlation between f and $\hat{f}$ is the maximum possible for any linear composite of observed variables (Bentler, 1968; Raykov, 2004). In this note, we clarify the distinction between estimated and true factor scores by drawing an analogy between the derivation of the regression factor scores and how predicted values are obtained in regression. We start with simple regression and compare it to a single-indicator latent variable model. We then move to multivariate regression and draw the analogy to obtaining regression factor scores from a multiple indicator model. We then describe some implications of the regression analogy. In the next sections, we use simulated data from a 1-factor model and 2-factor model to show how the amount of measurement error affects the reliability of regression factor scores; we also compare the performance of regression factor scores with that of unweighted sum scores across these scenarios. We conclude with a discussion.

[3]Many authors (e.g., Maraun, 1996) prefer the term "predicted" rather than "estimated" factor scores. Typically, we estimate parameters of a *distribution* of random variables (e.g., model parameters such as factor loadings) but predict *values* of random variables (though some psychometric models have person parameters). Another term in use is "constructed" factor scores (Beauducel & Rabe, 2009). However, we use "estimated" factor scores as it is more common in the psychometric literature. In addition, the word "predicted" may inadvertently imply a particular type of factor score estimate, namely regression factor score. While we primarily focus on regression factor scores, many of our statements are more general.

[4]"Regression factor score estimates" would be more correct, but the term "regression factor scores" is shorter and is widely used; we just ask the reader to remember that regression factor scores are not true factor scores.

[5]Other assumptions include that participants' memory is wiped between measurements, so that there are no repeated testing effects, and that the true score itself, representing the quantity being measured, is stable over time.

An analogy with regression

Simple regression and single-indicator factor model

Consider two models for an observed variable y, where all variables are scalars:

$$y = \beta x + \epsilon \quad (1)$$

$$y = \lambda f + e \quad (2)$$

Equation (1) describes a simple regression model, where the predictor x is another observed variable.[6] Equation (2) describes a factor model, where the predictor f is a latent variable. In both models, the error term assumed to be uncorrelated with the predictor. In the latent variable model, f is imagined to be a cause of y, and the error term represents measurement error in the sense discussed above. While in regression causality cannot often be assumed, to draw a parallel with factor analysis, in this instance we will assume that x causes y, and we will call the error term ϵ model error. Then, mathematically, the factor model can be viewed as a peculiar kind of a causal regression model where all scores are missing on the predictor.

The fact that x is in the dataset, whereas f is not, has two implications. First, the model parameters β and λ in Equations (1) and (2) are estimated differently. In a regression model, an estimate of β can be obtained straightforwardly from the data on x and y. In a factor analysis model, an estimate of λ can only be obtained if we are able to find other observed variables that are also predicted by f. That is, we will need to extend the model from a single variable y to multiple observed indicators $\mathbf{y}$, which we will do in the next section. In this section, we will assume that model parameters are known, so differences in estimation between the regression and the factor model are not relevant, and y is a single observed indicator.

The second implication is more esoteric and is known as the problem of factor score indeterminacy (Maraun, 1996). Briefly, because the values on f are not observed for *any* person, there are multiple latent variables that can fit the description given by Equation (2), even if λ is exactly known. The most important practical consequence of this indeterminacy is that the correlations of these different possible fs with other variables external to the model (say z) will be different, unless further assumptions are made (e.g., that z and e are uncorrelated). Stated differently, the precise location of the latent factor f, when viewed as a vector in the variable space that includes other variables (Wickens, 2014), is not fully known, which implies that its correlations with other variables are to some degree indeterminate. The indeterminacy issue does not affect our analogy, so we set it aside for now but give a fuller explanation in Appendix A.

We now use the analogy between the models in Equations (1) and (2) to explain the concept of regression factor scores, which are the most popular type of factor score estimates. Because the latent factor is the predictor in Equation (2), attempting to estimate scores on f by predicting them from the indicator y is analogous to predicting individuals' values on the predictor x from their values on the outcome variable y in regression. To make this prediction, we need to invert the regression equation in 1, as follows:

$$x = \beta^* y + \epsilon^*, \quad (3)$$

where the error term is different from that in Equation 1 because it now must be orthogonal to y and not to x. Because this inverse model is no longer causal, we will call this error term prediction error. Importantly for our analogy, if the parameters of the original regression model are known, the parameters of the inverse regression model are also known. In fact, when x and y are standardized, $\beta = \beta^*$. More generally, $\beta^* = \beta\sqrt{\frac{\text{var}(x)}{\text{var}(y)}}$, where $\text{var}(y) = \beta^2\text{var}(x) + \text{var}(\epsilon)$; that is, we are able to write the inverse regression coefficient in terms of the parameters of the original regression model (which include predictor variance and error variance). Then, for an individual with a known score on y, we can predict their score on x as $\hat{x}_i = \beta^* y_i$. This predicted value will not equal the actual (unknown) value x_i for that individual, even if the population value of β^* is known. Figure 1 (left panel) shows the discrepancy between actual x_i and $\hat{x}_i$. Because the values of x and y are observed for at least some individuals in the sample, the value of the residual or the error of prediction can also be obtained for those individuals simply by subtracting $\hat{x}_i$ from x_i.

To estimate factor scores using the regression method, we write the corresponding inverse regression model as follows:

$$f = \lambda^* y + e^*, \quad (4)$$

where again the error term is different from the one in Equation (2) because it now must be orthogonal to y. Since this equation is no longer causal but is just a prediction equation, e^* is prediction error, not measurement error (McDonald, 2011). This is the model from which regression factor scores are obtained. While *all* values on f are missing for everyone, the

[6]Item intercepts are omitted for simplicity. We assume a random regression model.

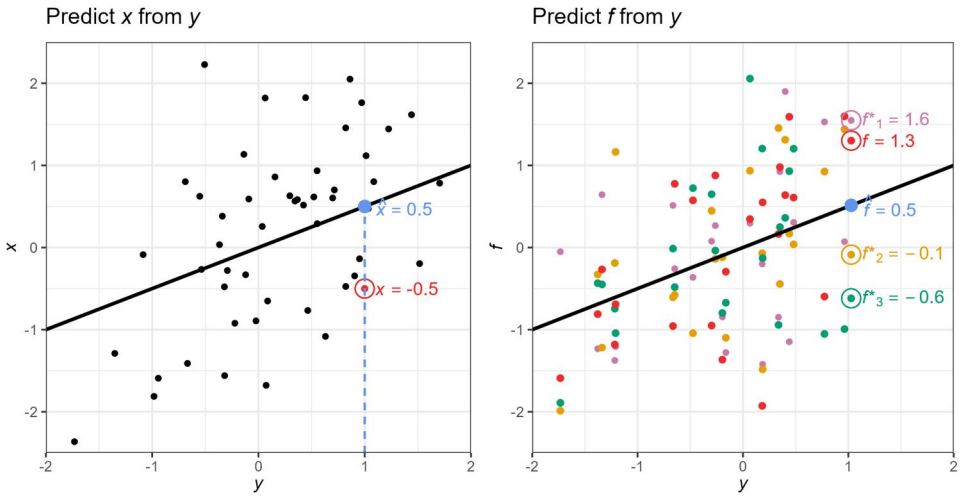

Figure 1. Discrepancy between observed and predicted values in simple linear regression, and between latent and predicted values in factor analysis. *Note.* In the left plot, the circled point is the observed value of x_i. In the right plot, the four circled points are four possible values (out of an infinite number of possible values) of f_i. Neither the observed value of x_i nor the possible values of f_i are equal to the predicted values $\hat{x}_i$ and $\hat{f}_i$, which fall on the regression lines.

prediction is done in the same way as in regression. For an individual with a score y_i, we can predict their true factor score with $\hat{f}_i = \lambda^* y_i$, which is their regression factor score. Then, completely parallel to regression, $\lambda^* = \lambda\sqrt{\frac{\text{var}(f)}{\text{var}(y)}}$, where $\text{var}(y) = \lambda^2\text{var}(f) + \text{var}(e)$. Because the latent variable does not have assigned units, we set $\text{var}(f) = 1$ for identification; therefore, $\lambda^* = \frac{\lambda}{\sqrt{\lambda^2 + \text{var}(e)}}$. We have solved for the weight λ^* needed to obtain the regression factor score in terms of the parameters of the original factor model.

Figure 1 (right panel) illustrates that, just as in regression, the predicted values, i.e., the regression factor scores, will always fall on the regression surface (here, a line), and therefore will not equal the true factor scores. However, unlike in regression, factor score indeterminacy means that it is not possible to obtain or estimate residuals for anyone, because no values of f_i are observed. The four different colors of points in this figure show four sets of possible factor scores that could have given rise to the observed values of y, but an infinite set of factor scores are possible. Further, this two-dimensional plot does not capture the fact that we also do not have enough information to "ground" the location of these residuals in a higher-dimensional space involving other variables. If we were to expand this figure to three dimensions by adding a third axis for some other variable z, we would not have the information on how to position the plane defined by y and f relative to the axis defined by z. Thus, the factor model allows us to estimate the correlations between the factor and its indicators, but we lack precise information on how the factor is related to any variables that are not in the model.

Multivariate regression and the factor model

In the previous section, we have worked with a single-indicator factor model with known parameters for simplicity. In reality, latent variables require multiple indicators if model parameters are unknown and are to be estimated. As well, it is common to have multiple latent factors. In this section we will assume that there are k latent variables (although the most common case for factor score estimation remains $k = 1$), and each latent variable f_t $(t = 1, \ldots, k)$ has at least three indicators, for a total of p observed variables.[7] The appropriate regression analogy is then to multivariate regression, where there are multiple predicted variables (to parallel multiple indicators of a factor) and potentially multiple predictors as well (to parallel one or more latent factors). We now develop this analogy; however, this section can be skipped by more applied readers without loss of continuity.

The multivariate regression model is given by:

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \epsilon, \tag{5}$$

where $\mathbf{y}$ is now a $p \times 1$ vector of criterion variables, $\mathbf{x}$ is a $k \times 1$ vector of predictor variables, $\mathbf{B}$ is the $p \times k$ matrix of regression coefficients, and ϵ is a $p \times 1$ vector of model errors. We can use covariance algebra (e.g., Bollen, 1989) to obtain:

$$\text{cov}(\mathbf{y}, \mathbf{x}) = \text{cov}(\mathbf{B}\mathbf{x} + \epsilon, \mathbf{x}) = \mathbf{B}\text{cov}(\mathbf{x}, \mathbf{x}) = \mathbf{B}\text{var}(\mathbf{x}),$$

and it follows that

$$\mathbf{B} = \boldsymbol{\Sigma}_{\mathbf{yx}}\ \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}, \tag{6}$$

[7] We assume this for simplicity. Factor models can be estimated with fewer indicators than three under additional constraints or within a context of larger models ($k > 1$ with factor correlations.

where $\mathbf{\Sigma_{yx}} = \text{cov}(\mathbf{y}, \mathbf{x})$ and $\mathbf{\Sigma_{xx}} = \text{var}(\mathbf{x})$. To see the parallel to the factor model, we also note the model-implied covariance structure for $\mathbf{y}$ implied by the regression model:

$$\mathbf{\Sigma_{yy}} = \text{var}(\mathbf{y}) = \mathbf{B}\text{var}(\mathbf{x})\mathbf{B}' + \text{var}(\epsilon) = \mathbf{B}\mathbf{\Sigma_{xx}}\mathbf{B}' + \mathbf{\Psi}_\epsilon, \tag{7}$$

where $\text{var}(\epsilon) = \mathbf{\Psi}_\epsilon$. We will use this equation shortly.

To predict scores on x from the scores on y, we need the inverse regression model:

$$\mathbf{x} = \mathbf{B}^*\mathbf{y} + \epsilon^*, \tag{8}$$

It follows by the same mathematics (i.e., covariance algebra) that the $k \times p$ matrix of inverse regression coefficients $\mathbf{B}^*$ is given by

$$\mathbf{B}^* = \mathbf{\Sigma_{xy}}\mathbf{\Sigma_{yy}^{-1}}. \tag{9}$$

We can use Equations (6) and (7) to obtain an expression for $\mathbf{B}^*$ in terms of $\mathbf{B}$:

$$\mathbf{B}^* = \mathbf{\Sigma_{xx}}\mathbf{B}'\mathbf{\Sigma_{yy}^{-1}} = \mathbf{\Sigma_{xx}}\mathbf{B}'(\mathbf{B}\mathbf{\Sigma_{xx}}\mathbf{B}' + \mathbf{\Psi}_\epsilon)^{-1}. \tag{10}$$

The matrix of original coefficients $\mathbf{B}$ will be analogous to the matrix of factor loadings, and the matrix of inverse regression coefficients $\mathbf{B}^*$ will be analogous to the matrix of optimal weights used to obtain regression factor scores.

Lastly, it will also be helpful to draw an analogy between the coefficient of determination in regression (i.e., R-squared), and reliability or "construct replicability" (Hancock & Mueller, 2001) in a factor model. In regression, the R-squared gives the proportion of variance in each criterion variable explained by the predictors, or, equivalently, it is the squared correlation between the criterion variable and its predicted value. We will only give the R-squared expressions for the inverse regression in Equation (8). The proportion of variance in each variable x_t $(t = 1, \ldots, k)$ explained by the variables in $\mathbf{y}$ is

$$R_t^2 = \frac{\text{var}(\boldsymbol{\beta}_t^{*'}\mathbf{y})}{\text{var}(\mathbf{x}_t)} = \frac{\boldsymbol{\beta}_t^{*'}\text{var}(\mathbf{y})\boldsymbol{\beta}_t^*}{\text{var}(\mathbf{x}_t)}, \tag{11}$$

where $\boldsymbol{\beta}_t^{*'}$ is the tth row of $\mathbf{B}^*$. Equation (10) can be used to further re-write this expression in terms of the original regression weights $\mathbf{B}$.

We now state parallel expressions to Equations (5)–(11) for the factor model. They only involve a change in notation. The factor model is

$$\mathbf{y} = \mathbf{\Lambda}\mathbf{f} + \mathbf{e}, \tag{12}$$

where $\mathbf{y}$ is a $p \times 1$ vector of indicators, $\mathbf{f}$ is a $k \times 1$ vector of latent factors, and $\mathbf{\Lambda}$ is the $p \times k$ matrix of latent regression coefficients, i.e., factor loadings. The expression parallel to Equation (6) is:

$$\mathbf{\Lambda} = \mathbf{\Sigma_{yf}}\mathbf{\Phi^{-1}}, \tag{13}$$

where $\mathbf{\Sigma_{yf}} = \text{cov}(\mathbf{y}, \mathbf{f})$ and $\mathbf{\Phi} = \text{var}(\mathbf{f})$. Unlike in regression, however, this expression cannot be used directly to obtain $\mathbf{\Lambda}$ because $\mathbf{\Sigma_{yf}}$ involves unknown correlations between observed and latent variables. The parallel expression to Equation (7) is:

$$\mathbf{\Sigma_{yy}} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}, \tag{14}$$

where $\mathbf{\Psi} = \text{var}(\mathbf{e})$, which is the familiar covariance structure under a factor analytic model. Model estimates for the parameters on the right-hand side can be obtained by fitting this model to the sample covariance matrix of $\mathbf{y}$. Once the model has been fit, we have the matrices $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$ (or their estimates).

To obtain regression factor scores, we invert the regression implied by the factor model:

$$f = \mathbf{W}\mathbf{y} + \mathbf{e}^*, \tag{15}$$

where $\mathbf{W}$ is the $k \times p$ matrix of weights.[8] To obtain regression factor scores, we use the expression parallel to Equation (9):

$$\mathbf{W} = \mathbf{\Sigma_{fy}}\mathbf{\Sigma_{yy}^{-1}}. \tag{16}$$

Only the model in Equation (12) can be estimated from the data, and the inverse model in Equation (15) cannot be fit directly. This is the main difference between regression and factor analysis models. Therefore, to compute the regression factor score weights $\mathbf{W}$, we express $\mathbf{W}$ in terms of the parameters of the original fitted factor model, parallel to the expression in Equation (10) for regression:

$$\mathbf{W} = \mathbf{\Phi}\mathbf{\Lambda}'\mathbf{\Sigma_{yy}^{-1}} = \mathbf{\Phi}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi})^{-1} \tag{17}$$

The $k \times 1$ vector of estimated (regression) factor scores for the ith individual can be obtained as $\hat{\mathbf{f}}_i = \mathbf{W}\mathbf{y}_i$, where $\mathbf{y}_i$ is the $p \times 1$ vector of their scores on $\mathbf{y}$. Estimated (regression) factor scores $\hat{\mathbf{f}} = \mathbf{W}\mathbf{y}$ are random variables that are just weighted linear composites of the observed variables.[9] In the case of a 1-factor model $(k = 1)$, $\hat{\mathbf{f}}$ is a scalar (a single score for each person), whereas $\mathbf{y}$ is a vector of p observed variables, so $\mathbf{W}$ reduces to a $1 \times p$ row vector of optimal weights for each variable. For multi-factor models, each row of $\mathbf{W}$ gives

[8] For consistency, we could have called it $\mathbf{\Lambda}^*$, but $\mathbf{W}$ is a much more common notation.

[9] Two other types of factor score estimates use the original, rather than the inverted, regression line to reverse-engineer estimates of latent variable scores from the observed values of $\mathbf{y}$. Under the "idealized variables" approach, $\hat{\mathbf{f}}_I = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{y}$, and under the Bartlett approach, which assumes the regression errors are heteroscedastic, $\hat{\mathbf{f}}_B = (\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{y}$. An expression connecting regression and Bartlett factor scores is $\hat{\mathbf{f}}_B = (\mathbf{I}_k + \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Phi}^{-1}\hat{\mathbf{f}}$. In the case of the 1-factor model $(k = 1)$, they are proportional.

optimal weights for obtaining a different weighted composite of the variables in $\mathbf{y}$ to best capture each factor.[10]

The R-squared values for the inverse regression in Equation (15) will give the proportion of variance in each latent variable f_t $(t = 1, ..., k)$ explained by all the observed indicators y, or equivalently, the squared correlation between each latent variable f_t and its prediction, i.e., the tth regression factor score $\hat{f}_t$. The squared correlation between the latent variable and the observed composite designed to measure it is also known as reliability. In the population, regression factor scores have the property of maximal reliability: That is, no other weighted linear composite of observed variables y can have a higher squared correlation with the latent variable. For this reason, we will use the abbreviation "MR" to refer to these R-squared values. Parallel to Equation (11), they are given by:

$$\begin{aligned} \mathrm{MR}_t = R_t^2 &= \frac{\mathrm{var}(\mathbf{w}_t'\mathbf{y})}{\mathrm{var}(\mathbf{f}_t)} = \frac{\mathrm{var}(\mathbf{w}_t'\mathbf{y})}{1} = \mathbf{w}_t'\mathrm{var}(\mathbf{y})\mathbf{w}_t \\ &= \phi_t'\mathbf{\Lambda}'\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{\Lambda}\phi_t, \end{aligned} \tag{18}$$

where $\mathbf{w}_t' = \phi_t'\mathbf{\Lambda}'\mathbf{\Sigma}_{\mathbf{yy}}^{-1}$ is the tth row of $\mathbf{W}$ (see Equation (17)) and ϕ_t is the tth row (or column) of $\mathbf{\Phi}$. As a reminder, the variance of each latent variable is set to 1 for identification. The square-roots of the values in Equation (18) have also been called factor determinacy (FD) indices (Grice, 2001; Rodriguez et al., 2016).

In the special case of the 1-factor model, the expression in Equation (18) simplifies to the following formula:

$$\mathrm{MR} = \boldsymbol{\lambda}'\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\boldsymbol{\lambda} = (\boldsymbol{\lambda}'\boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{\Psi})^{-1}\boldsymbol{\lambda}, \tag{19}$$

When $\mathbf{\Psi}$ is diagonal, the resulting further simplified version of Equation (19) has also been labeled as "construct replicability" or "coefficient H" (Hancock & Mueller, 2001), but we will refer to it always as maximal reliability (MR).

Implications of the regression analogy

The analogy with regression has a number of implications for understanding factor score estimation. We will continue to focus on regression factor scores, but our discussion applies more generally, since other factor score estimates will have worse prediction properties. First, as the sample size grows, estimation or prediction of factor scores only improves up to a point. A larger sample size will ensure that the estimated model parameters (such as factor loadings) become increasingly precise estimates of their population values, and that the factor score estimates are computed using an increasingly accurate set of weights (in the case of regression factor scores, these are optimal weights). In other words, sampling error decreases. But increasing sample size will never remove model error in regression (i.e., ϵ in Equation (1)) or measurement error in factor analysis (i.e., e in Equation (2)). To the extent that there is model error in regression, $\mathbf{x}$ does not explain all of the variation in $\mathbf{y}$, and the values of the predicted variables will not fall exactly on the regression surface defined by the predictors, in either the original or the inverse regression equation. Similarly, in factor analysis, the true factor scores will not fall exactly on the inverse regression surface defined by the observed variables to the extent that there is measurement error in all of the observed variables (i.e., $\mathbf{f}$ does not explain all of the variation in $\mathbf{y}$).

The coefficient of determination (R-squared) can be used to quantify the success of prediction in both regression and factor analysis. R-squared is the squared correlation between the predicted value (which is a weighted linear combination of the predictors) and the criterion. For the 1-factor model, the R-squared for predicting regression factor scores from the observed variables is called maximal reliability (MR; see Equation (19)). It is equal to the squared correlation between the best linear combination of the observed variables (i.e., the regression factor scores), and the true scores on the latent variable. We will compute MR for the illustrations in the next section.

A second implication of the analogy with regression is that as prediction error decreases, regression factor scores approach true factor scores. Prediction error decreases as individual indicators become more reliable (measurement error for some or all indicators decreases in the original model, or equivalently standardized loadings increase) and/or as the set of indicators becomes large. In fact, under minimal assumptions, factor score estimates will approach true factor scores as the number of indicators of that factor goes to infinity (Bentler & Kano, 1990; Ellis & Junker, 1997). Recall that in classical test theory, the true score is defined as the long-run average of repeated observations. In the factor analysis model, repeated observations are multiple indicators, so as their number increases, the latent variable can be approximated more and more precisely by a linear combination of them. In the next section, we illustrate how the

[10]Interestingly, in correlated factor models, these weights will generally all be non-zero; that is, all variables in the model contribute to the estimation of scores on each factor, not just indicators of that factor.

accuracy of regression factor scores depends on the measurement properties of the indicators.

Illustrations

In this section, we illustrate the ideas explained above in a few ways: First with a single dataset with a covariance matrix that perfectly represents a population (i.e., no sampling variability), then with several such datasets, varying a range of population parameters, such as factor loadings, and finally with repeated draws of samples from a given population, this time with sampling variability. We do this with both 1- and 2-factor models to show some of the ways in which regression factor scores behave differently from true factor scores, and how these differences depend on whether the weights are known or estimated (i.e., with versus without sampling variability).[11] Alongside these comparisons, we include unweighted sum scores to show the ways in which unit weights, which are theoretically sub-optimal but not affected by sampling variability, can produce scores that differ (sometimes for the worse, sometimes for the better) from those based on regression factor score weights, which are theoretically optimal but subject to sampling variability.

Illustration 1: a single "population" dataset, 1-factor model with equal loadings

Using the code in Appendix B, we generated a single dataset from a 1-factor model with 3 indicators, where each observed variable has a factor loading of 0.60 and an error variance of 0.64, and the factor f has variance 1. To remove the influence of sampling fluctuations, we generated a "population" dataset, i.e., one where the covariance matrix is perfectly described by the population model. In addition to simulating scores on observed variables, we also simulated true factor scores (and true error scores), to enable the comparison between the true and estimated factor scores. We then fit the 1-factor model to the generated data. Because the model fits the simulated data exactly, this fitting procedure produced parameter estimates that exactly match the population values. We obtained the regression factor scores from this analysis using the `lavPredict` function with `method = "regression"` in **lavaan**. The R code in Appendix B also shows how to reproduce the estimated factor scores produced by **lavaan** by applying Equation (16) to the parameter matrices obtained from the fitted model.

[11]Code to reproduce all simulations can be downloaded at osf.io/a68wm/.

Table 1. Scores on the observed variables, true factor scores, and regression factor scores for the first ten rows of generated data in Illustration 1: 1-factor model with known parameters.

y_1	y_2	y_3	f	$\hat{f}$
1.450	−0.883	0.964	1.028	0.534
0.433	−0.384	1.070	0.709	0.390
0.158	1.378	0.302	1.329	0.641
0.833	0.487	0.893	0.944	0.772
0.758	0.754	−0.061	0.307	0.506
−0.768	0.643	−0.814	0.352	−0.328
1.313	2.152	1.213	1.483	1.632
1.272	0.060	−0.572	−0.815	0.265
0.050	−2.008	−0.702	−0.750	−0.928
−1.836	−1.691	−1.356	−2.098	−1.703

Table 1 displays the first 10 rows of data, including the observed variables y_1 to y_3, the true factor scores on f and the regression factor scores $\hat{f}$. Because sampling variability was removed from the simulation, these estimated factor scores are the closest we can get to the true factor scores using only observed data, maximizing the correlation between f and $\hat{f}$. However, it is apparent that these are quite different values. Their aggregate properties are also quite different: the true factor scores have been generated to have variance of 1, whereas the estimated factor scores have variance .63. The same phenomenon can be observed in regression (predicted values vary less than true values).[12]

The difference between true factor scores and their estimated values is also apparent in the correlation between them: In this example, the correlation is 0.79, which is the square-root of maximum reliability (MR).[13] Maximum reliability is the reliability of an optimally weighted sum score of the items, which is precisely what the regression factor scores (obtained when the population parameters are known, as is the case here) are. In this example, however, because all factor loadings are equal, estimated factor scores are equally-weighted composites, and thus are perfectly correlated with the sum scores. In this situation, MR has the same population value as coefficient omega (McDonald, 1978), which describes the reliability of an unweighted sum score.

Illustration 2: multiple "population" datasets, 1-factor models with varied loadings

Next, we simulated 1000 datasets from a 1-factor model with 3 indicators, but each dataset came from a

[12]In practice, factor score estimates are often subsequently standardized, so that their sample variance becomes 1.

[13]This value can be obtained from the model parameters using Equation (19); the R code is given in Appendix B.

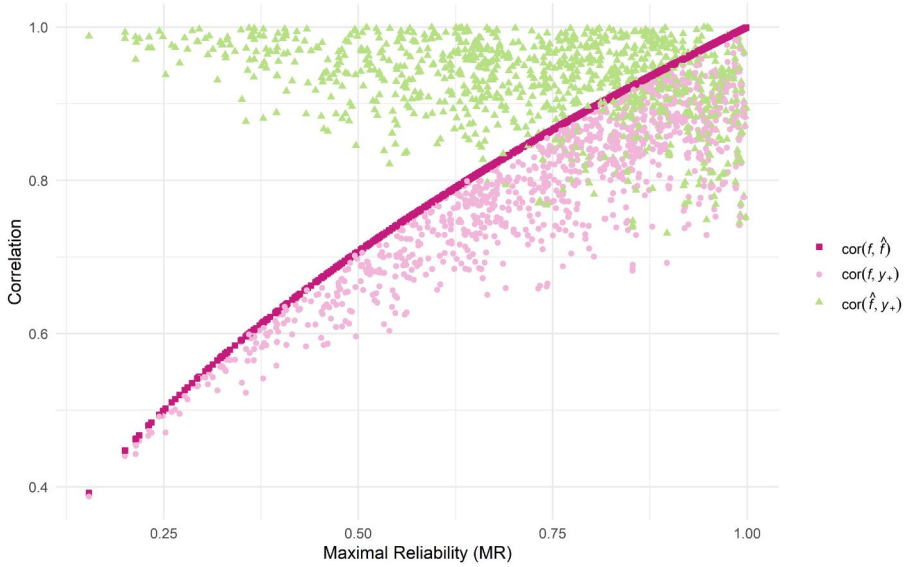

Figure 2. Correlations among true factor scores, regression factor scores, and unweighted sum scores for Illustration 2: 1-factor models with varied loadings, parameters assumed to be known. *Note.* f = true factor scores, $\hat{f}$ = estimated factor scores, and y_+ = unweighted sum scores. Data are generated from a 1-factor model with factor loadings drawn randomly from $U[.2, 1]$. Regression factor scores are obtained using Equation (16).

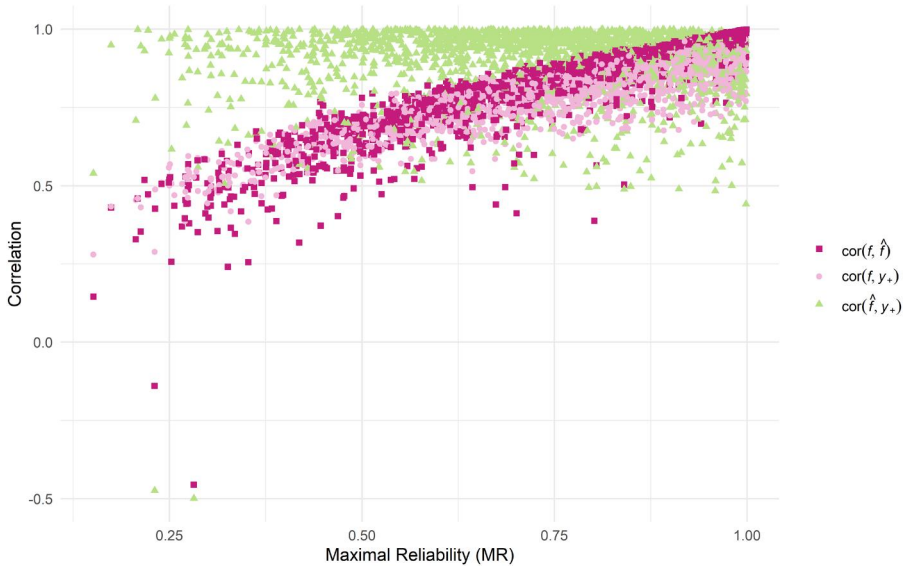

Figure 3. Correlations among true factor scores, regression factor scores, and unweighted sum scores for Illustration 3: 1-factor models with varied loadings, parameters estimated from $N = 200$. *Note.* f = true factor scores, $\hat{f}$ = estimated factor scores, and y_+ = unweighted sum scores. Population covariance matrices are generated from a 1-factor model with standardized factor loadings drawn randomly from $U[.2, 1]$, and sample data of size $N = 200$ drawn from a multivariate normal distribution. Regression factor scores are obtained using Equation (16).

population with different factor loadings. The factor loading values were randomly drawn from a uniform distribution $U[0.2, 1]$. The influence of sampling fluctuations was again removed by forcing each dataset to have the exact covariance matrix specified by the population values of the loadings. Figure 2 displays the correlations among true factor scores, estimated (regression) factor scores, and the unweighted sum scores for these 1000 populations. Along the x-axis is maximum reliability. The correlations between true and estimated factor scores (dark pink squares) fall on a curve for the square-root function because this correlation is simply the square root of MR. The correlations between regression factor scores and unweighted sum scores (light green triangles) reflect the amount of variation in the three randomly drawn factor

loadings; the more disparate these loadings, the lower the correspondence between regression factor scores and unweighted sum scores. Finally, the correlations between true factor scores and unweighted sum scores (light pink circles) are always equal to or lower than the correlations between true and estimated factor scores (dark pink squares), reflecting that regression factor scores outperform unweighted sum scores *when the true factor loadings are known*.[14] In the next simulation, we complicate this story by adding sampling variability, so that regression factor scores are computed using *estimated* rather than true factor loadings, which means they are no longer optimally weighted composites.

Illustration 3: multiple "sample" datasets drawn from 1-factor models with varied loadings

Adding sampling variability to the simulation depicted in Figure 2 introduces estimation error into the factor loadings, resulting in regression factor scores that are computed using error-laden weights rather than the optimal weights. Figure 3 displays the correlations among true factor scores, regression factor scores, and the unweighted sum scores in 1000 datasets of size $N = 200$, drawn from different populations. For each dataset, population factor loadings were again randomly drawn from the uniform distribution $U[0.2, 1]$, but the datasets were not forced to conform to the population covariance matrix (i.e., sampling variability was left in). Along the x-axis is the *population* value of MR.

The correlations between true and regression factor scores (dark pink squares) no longer fall on the curve for the square-root function because the regression factor scores are computed from estimated factor loadings, rather than population factor loadings. It is interesting to compare these correlations with the correlations between the true factor scores and unweighted sum scores (light pink circles). It is no longer the case that the regression factor scores are always more highly correlated with the true factor scores than are the unweighted sum scores. Regression factor scores perform worse than sum scores when the estimated factor loadings happen to be very different from their population values. For example, if the true population loadings for y_1 and y_2 are 0.80 and 0.10, but their sample estimates are reversed (i.e., y_1 is estimated to have a loading near 0 and y_2 is estimated to have a high loading), the regression factor scores will be strongly correlated with y_2, and thus weakly correlated with f. In contrast, a simple sum score does not differentially weigh items on the basis of their estimated loadings, so the quality of estimated loadings does not affect its correlation with the true factor scores.

Figure 4 displays the dark and light pink dots of Figure 3 in a violin plot to enable a clearer comparison across distributions. The distribution of correlations between true factor scores and regression factor scores is right-skewed. The average correlation between the true and regression factor scores is higher than the average correlation between true factor scores and unweighted sum scores (0.82 vs. 0.79), but the variance of the correlations between true factor scores and regression factor scores is larger ($SD = .14$ vs. 0.11).[15]

Illustration 4: a single "population" dataset, 2-factor model with equal loadings

When there is more than one factor in the model, the true and estimated factor scores also differ in their correlations with each other.[16] We now simulate data from a 2-factor model to illustrate the difference in the correlation matrices of true factor scores versus regression factor scores. Here, we encounter a complicating factor: We must choose whether to predict each factor score with all of the indicators of the 2-factor model or with only its own indicators (i.e., from the parameters of a one-factor model fit to only those indicators) (Logan et al., 2022).[17] We illustrate both approaches. In the former approach, Equation (17) is applied to the parameter matrices obtained from the two-factor model, producing a weight matrix that contains non-zero weights for all indicators of both factors (unless the factors are uncorrelated). That is, regression factor scores for each factor are a weighted combination of not only the indicators of that factor, but also of the

[14]This figure shows that variability of the correlations increases with increasing MR. This is because it is a function of the variance in the (randomly drawn) loadings. When MR is low, that corresponds to a draw of 3 low loadings. When MR is high, it could be that there is one high loading and the other two are near zero, or that there are 3 medium-high loadings. When there is more variability in the loadings, there is more discrepancy between the regression factor scores and the sum scores.

[15]This result appears to be in conflict with the results of McNeish (2023). We explore this discrepancy in Discussion.

[16]Here, we focus on regression factor scores. Other types of estimated factor scores exist that preserve correlations among the factors, but they have other suboptimal properties (McDonald & Burr, 1967).

[17]In sample data, there are actually three options, because regression factor score weights from only the indicators of that factor can be obtained either from the 2-factor model (i.e., by subsetting the model matrices) or from a 1-factor model. These two approaches will result in different estimates, but when the larger model is correct, these methods are asymptotically the same.

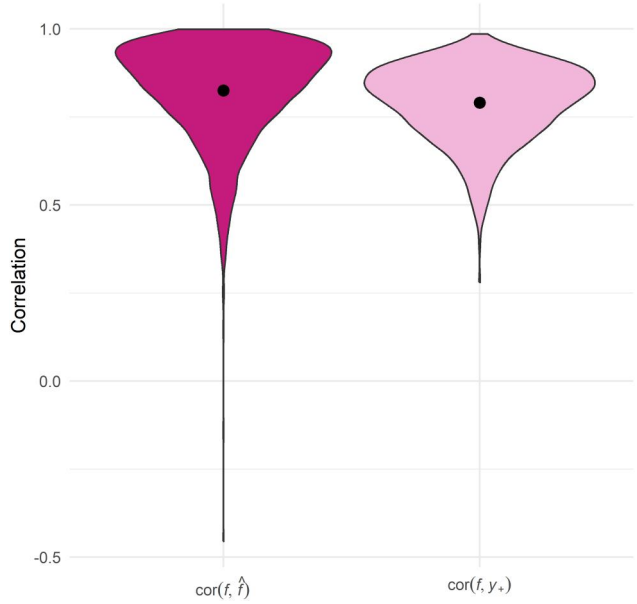

Figure 4. Correlations between true and estimated factor scores, and between true and unweighted sum scores in Illustration 3: 1-factor models with varied loadings, parameters estimated from $N = 200$. *Note.* f = true factor scores, $\hat{f}$ = estimated factor scores, and y_+ = unweighted sum scores. Population covariance matrices are generated from a 1-factor model with standardized factor loadings drawn randomly from $U[.2, 1]$, and sample data of size $N = 200$ drawn from a multivariate normal distribution. Regression factor scores are obtained using Equation (16).

indicators of the other factor. These regression factor scores have the property of maximal reliability. To obtain estimated factor scores using *only* the indicators of each factor, we use the model matrices corresponding to one-factor models for each factor. Prediction based on a subset of predictors will naturally result in lower R-squared (reliability); However, this is still maximal reliability for the smaller set of indicators.

We simulated data from a 2-factor model with 3 indicators per factor, where each observed variable has a factor loading of 0.60 and an error variance of 0.64, the factors f_1 and f_2 have variance 1, and the correlation between the factors is .30. We again generated data that were perfectly described by the population model, and we fit (a) a 2-factor model to the data, and (b) two 1-factor models to each factor separately. Table 2 displays the correlations among the true factor scores (f_1, f_2), regression factor scores obtained from the two-factor model ($f_1^\dagger$, $f_2^\dagger$), and regression factor scores obtained from two one-factor models ($\hat{f}_1, \hat{f}_2$).

Even though the data were simulated to reproduce the exact population parameters, so that the regression factor score weights are at their population values, the correlations among estimated factor scores do not match the correlations among true factor scores.[18] While the true factor correlation is 0.30, the correlation between estimated factor scores is either 0.39 (when obtained from the full 2-factor model) or 0.20 (when obtained from individual 1-factor models). The actual estimated factor scores for each person are of course different as well. While the regression factor scores from the 2-factor model will have higher squared correlations with the actual factors (as given by Equation (18)) than the regression factor scores from 1-factor models (as given by 19), in practice factor scores are typically estimated from unidimensional models. For this reason, in the remaining two illustrations, we focus on the regression factor scores obtained from two separate one-factor models.

Table 2. Correlations among true factor scores and regression factor scores in illustration 4: 2-factor model with known parameters.

	f_1	f_2	$f_1^\dagger$	$f_2^\dagger$	$\hat{f}_1$	$\hat{f}_2$
f_1	1.00					
f_2	*0.30*	1.00				
$f_1^\dagger$	0.80	0.34	1.00			
$f_2^\dagger$	0.31	0.86	*0.39*	1.00		
$\hat{f}_1$	0.79	0.24	0.99	0.28	1.00	
$\hat{f}_2$	0.26	0.86	0.32	1.00	*0.20*	1.00

Note. f_1 and f_2 are true factor scores, $f_1^\dagger$ and $f_2^\dagger$ are regression factor scores obtained from the 2-factor model, and $\hat{f}_1$ and $\hat{f}_2$ are regression factor scores obtained from 2 separate 1-factor models of f_1 and f_2.

Illustration 5: multiple "population" datasets, 2-factor models with varied loadings

We now extend the previous illustration from equal loadings to randomly drawn loadings. We simulated 1000 datasets from a 2-factor model with 6 indicators, where

[18]This is a known problem in factor score estimation (Grice, 2001)

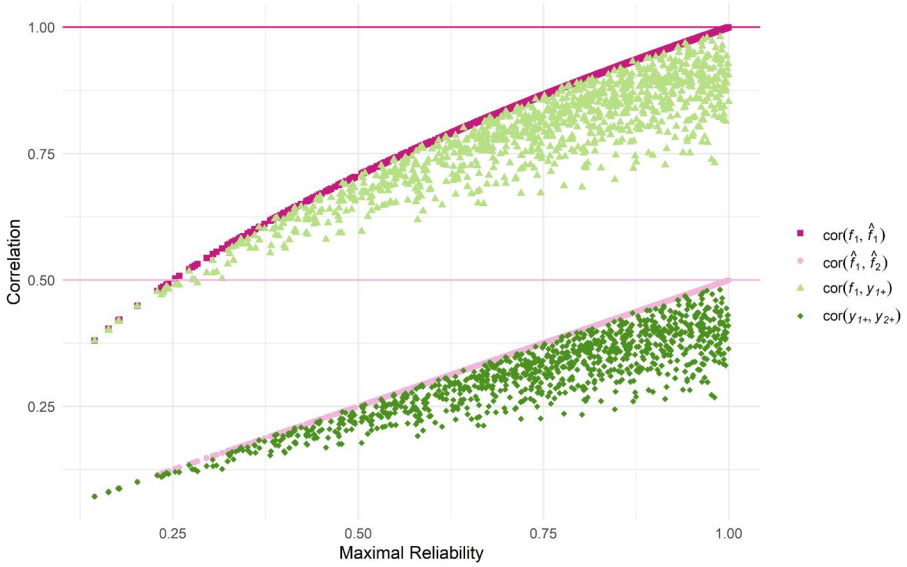

Figure 5. Correlations among true factor scores, regression factor scores, and unweighted sum scores in Illustration 5: 2-factor model with known parameters. *Note*. f_1 and f_2 are true factor scores, $\hat{f}_1$ and $\hat{f}_2$ are regression factor scores from one factor models, and y_{1+} and y_{2+} are unweighted subscale sum-scores. Population covariance matrices are generated from a 2-factor model with 3 indicators per factor, one set of 3 standardized factor loadings drawn randomly from *U*[.2, 1] repeated across the 2 factors, and $\text{cor}(f_1, f_2) = .5$. Regression factor scores are obtained from separate 1-factor models, fit to each of the 2 factors, using Equation (16).

each dataset is perfectly described by a population model with different factor loadings. To determine the parameters of each population, three factor loadings are randomly drawn from the uniform distribution *U*[0.2, 1], and these loadings are repeated across factors (i.e., the 3 loadings for Factor 2 are identical to those of Factor 1). The correlation between true f_1 and f_2 is always .5. The covariance matrix of each dataset is perfectly described by the population model.

Figure 5 shows the correlations among true factor scores, regression factor scores from 1-factor models, and unweighted sum scores, as a function of MR (MR is the same for both factors because they have the same loadings). First we examine the correlation between the estimated factor scores for f_1 and f_2 (light pink circles, which overlap each other so as to appear as a solid line). While the true scores for the two factors are always correlated at 0.50, the correlation between estimated scores is downwardly biased to the degree that that these weighted composites are unreliable (i.e., as a function of MR)—this is the famous attenuation due to unreliability, and it again illustrates the fact that estimated factor scores are observed composites, not actual latent variables. The correlations between unweighted sum scores (shown as dark green diamonds) computed from the indicators of f_1 and f_2 are also attenuated (downwardly biased) due to unreliability, but because their reliability is given by coefficient ω and not MR, the relationship with MR is not deterministic. In these population simulations, the correlations between the sum scores are always at least as attenuated or more attenuated than the correlation among the regression factor scores, reflecting their lower reliability. Lastly, the correlations of the true factor scores with regression factor scores and with sum scores (the dark pink squares and the light green triangles, respectively) show the same patterns as for the one factor model in Figure 2.

Illustration 6: multiple "sample" datasets drawn from 2-factor models with varied loadings

In our final simulation, we modify the previous simulation by adding sampling variability; that is, we draw random samples of size $N = 200$, without forcing the sample covariance matrix to equal the population covariance matrix. The population characteristics are otherwise the same as in the previous simulation. Here, to obtain the estimated regression weights based on a 1-factor submodel, estimates from the fitted 2-factor model can be used, or two separate 1-factor models can be fit to each subset of indicators, resulting in slightly different estimates. We chose to refit the individual 1-factor models, but when the larger model is correct, these methods are asymptotically the same. Figure 6 displays the correlations among true factor scores, regression factor scores, and unweighted sum scores. The

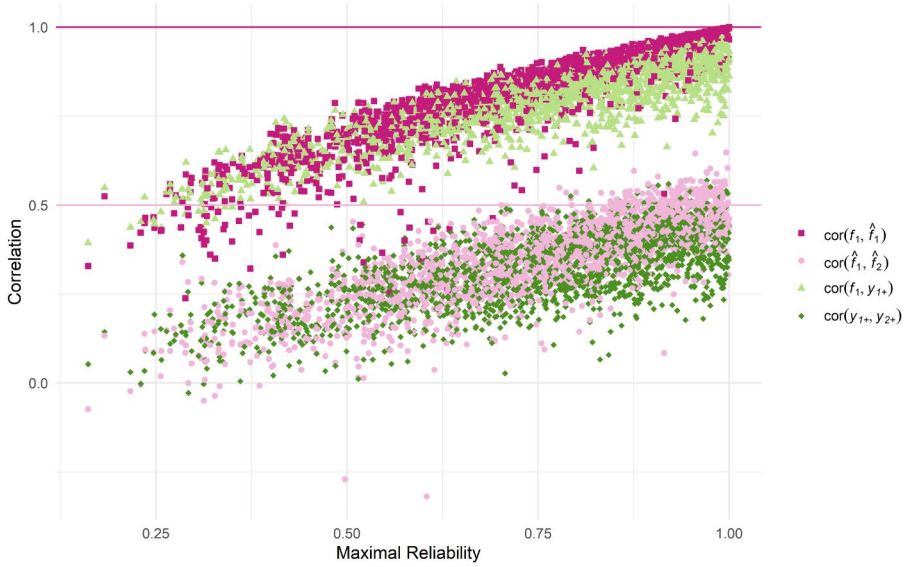

Figure 6. Correlations among true factor scores, regression factor scores, and unweighted sum scores in Illustration 6: 2-factor models with varied loadings, parameters estimated from $N = 200$. *Note.* f_1 and f_2 are true factor scores, $\hat{f}_1$ and $\hat{f}_2$ are regression factor scores from one factor models, and y_{1+} and y_{2+} are unweighted subscale sum-scores. Population covariance matrices are generated from a 2-factor model with 3 indicators per factor, one set of 3 standardized factor loadings drawn randomly from *U[*.2, 1*]* repeated across the 2 factors, and $\text{cor}(f_1, f_2) = .5$. Sample data of size $N = 200$ are drawn from a multivariate normal distribution. Regression factor scores are obtained from separate 1-factor models, fit to each of the 2 factors, using Equation (16).

correlations between estimated factor scores for the two factors (light pink circles) again tend to underestimate the true factor correlation, but there is now a great deal of variability in the estimated values, and for higher values of MR, there are a few cases of *over*estimation. Sum score correlations (dark green diamonds) are also quite variable, with visibly more variability relative to estimated factor scores for higher values of *H*.

The patterns of correlations between true factor scores and regression factor scores (dark pink squares), and between true factor scores and sum scores (light green triangles) replicate the results observed for the 1-factor model in a similar scenario (see Figure 3). Regression factor scores are, on average, more strongly correlated with true factor scores than sum scores are (mean correlations = .83 and 0.79), but are also more variable (standard deviations = .13 and 0.10).

Discussion

The goal of this article was to make more intuitive the distinction between true and estimated factor scores. Using an analogy with estimated values in regression, we clarified that estimated factor scores are never equal to the true factor scores, and that their proximity to true scores is a function of the amount of measurement error in the set of indicators (Rigdon et al., 2019), as captured by MR. Focusing on regression factor scores, we showed that estimated factor scores differ from true factor scores in terms of individual values, overall variance, and correlations with each other. These differences exist even when the model parameters are known.

Estimated factor scores do not approach true factor scores with increasing sample size

It may seem plausible that estimated factor scores would approach factor scores as the sample size goes to infinity, but they do not. When researchers work with large samples or when methodologists generate simulated data and use the regression method to get factor score estimates, they may mistakenly think they are dealing with estimated values that are as good as actual factor scores. However, our analogy with regression has aimed to highlight that model error (in regression) and measurement error (in factor analysis) are distinct from sampling error. In the simple case of a single predictor (of either an observed or a latent variable), if the standardized population regression coefficient is not 1, that means there is some model error (variation around the regression line), and estimated values are never the same as the actual values, no matter how much data one has. In the case of a factor model, the regression plane defined by the observed variables will be a sub-space of the larger latent variable space, defined by the errors and the

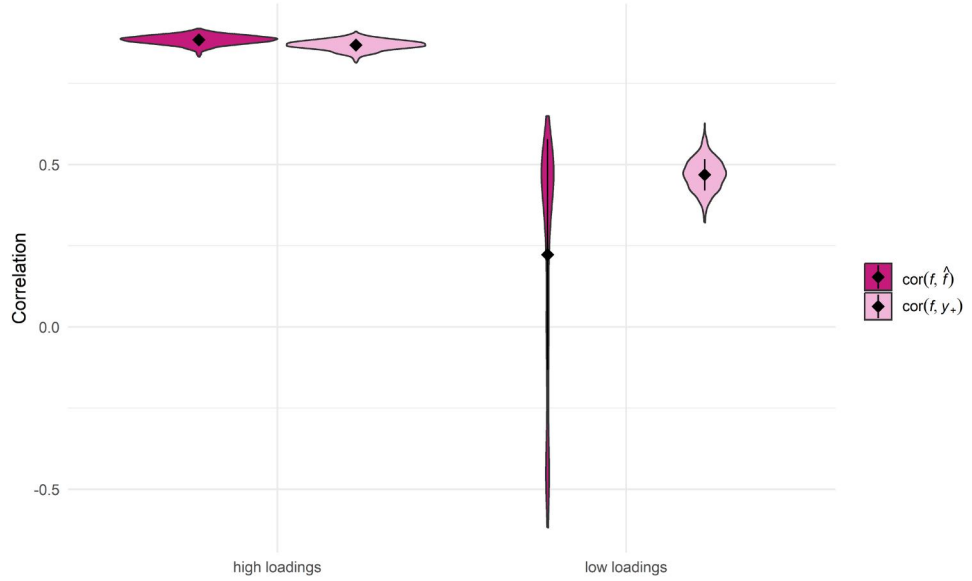

Figure 7. Correlations between true and estimated factor scores, and between true factor scores and unweighted sum scores, for high vs. low loadings. *Note.* f = true factor scores, $\hat{f}$ = regression factor scores, y_+ = unweighted sum scores. Population factor loadings in the "high loadings" condition are $\{.40, .50, .65, .80, .85\}$ and those in the "low loadings" condition are $\{.05, .15, .25, .35, .40\}$. Each distribution describes results from 1000 sample draws ($N = 250$) from the same population. A full description of the simulation, including cutpoints used to discretize each observed variable, can be found in McNeish (2023), and code for this figure can be found at https://osf.io/a68wm/.

factors. This is the difference between sampling error (the kind that disappears with sample size) and measurement error or model error (the kind that doesn't).

Different methods of estimating factor scores result in slightly different properties. We have focused on regression factor scores because they are the default method of factor score estimation in popular software, but some of the issues highlighted in our simulation results will differ as a function of factor score estimation method. For example, Logan et al. (2022) recommended using ten Berge estimates (Krijnen et al., 1996; Ten Berge et al., 1999), which are estimated from a multi-factor model and have the property that correlations among the estimated factor scores are equivalent to those among the latent variables. For these estimates, therefore, the distortions presented in Table 2 should be eliminated. While this is a nice property, only regression factor scores enjoy the property of maximum reliability, which is lost if other methods are employed. However, while factor score prediction methods differ in which properties are optimized in the estimated scores (McDonald & Burr, 1967), all methods share the same basic limitation that they cannot account for the missing dimension(s) on which the true factor scores lie. That is, all estimated factor scores are observed composites, and they cannot fully account for the information contained in the latent variables they try to approximate, no matter the sample size.

However, there does exist a limit in which estimated factor scores will approach factor scores: If the number of indicators goes to infinity, estimated factor scores approach true factor scores (or, in another sense, factors become principal components; Bentler & Kano, 1990). Another limit also exists, although it is admittedly less useful in practice: As we showed *via* simulation, holding the number of indicators constant, factor score estimates will approach true factor scores as their reliability (i.e., MR) approaches 1. For this coefficient to approach 1, it is not required that all indicators become more precise, but in fact, only one increasingly precise indicator is required. The need for just one good indicator is a useful reminder that in the reflective measurement model, all indicators carry the same information such that a single perfect indicator is better than 100 unreliable ones. That is, the reflective model does not allow different indicators to capture unique facets of a latent variable (Bollen & Bauldry, 2011).

We have focused on regression factor scores in this article.

Do regression factor scores outperform unweighted sum scores?

While the comparison of different observed composites was not our direct focus here, given recent resurgent interest in properties of sum scores (e.g.,

McNeish, 2023; Widaman & Revelle, 2023), we have also investigated the accuracy and stability of regression factor scores relative to unweighted sum scores. Our illustrations confirmed that when the population measurement model parameters are known, regression factor scores outperform unweighted sum scores, as measured by higher resulting correlations with the true factor scores. These differences will be most pronounced when the population loadings are highly heterogeneous, because when loadings are equal, regression factor scores (estimated from one-factor models) and unweighted sum scores become colinear. However, the property of maximal reliability holds only in the population, and in sample data where factor loadings are estimated rather than known, regression factor scores can perform better or worse than unweighted sum scores, depending on how close the estimated factor loadings are to their population values, which in turn depends on the mean and variance of the true factor loadings, as well as on the sample size.

Some of these results seemingly contradict those of McNeish (2023), who found that regression factor scores always outperform unweighted sum scores. The primary difference between our simulation designs is that we sampled from a wider range of factor loadings that included much lower values. For example, McNeish (2023) drew samples of size $N = 250$ from a 1-factor model with 5 indicators and loadings of $\{.40, .50, .65, .80, .85\}$ and found that not only was the average correlation between true and estimated factor scores higher than that between true factor scores and sum scores (0.89 vs. 0.87), but that the variance of those correlations was also slightly lower ($SD = .011$ vs. 0.015). We replicated these findings and extended them by adding a condition with substantially lower factor loadings of $\{.05, .15, .25, .35, .40\}$. Figure 7 displays the distribution of correlations from the original and the modified set of factor loadings. In the low loadings condition, the correlation between true and estimated factor scores was lower on average than the correlation between true factor scores and unweighted sum scores (0.20 vs. 0.47) and had a much larger variance (0.36 vs. 0.05). While factor loadings in the range of 0.05 to 0.40 are not common, this extreme condition demonstrates that the relative performance of sum scores and estimated factor scores depends dramatically on the reliability of the observed indicators. Uanhoro (2019) examined a broader range of conditions and suggested that when there is a great deal of uncertainty in the model (e.g., when factor loadings are lower and when sample size is smaller), unweighted sum scores outperform estimated factor scores. In addition, sample estimates of MR (reliability of regression factor scores) have been found to be more positively biased than sample estimates of coefficient omega (reliability of sum scores) in such conditions (Aguirre-Urreta et al., 2019). Lastly, while the details of these arguments are not our focus here, we note that there may be other reasons researchers may prefer sum scores over predicted factor scores as the observed composite of choice, such as consistent weights across samples (e.g., Widaman & Revelle, 2023).

Recommendations for practice

The most obvious alternative to computing estimated factor scores and carrying those forward to a subsequent analysis is to keep variables latent by doing all analysis in the context of SEM. But SEM is not always the best approach. For example, the full model may be too large or complex to estimate on the available sample, or individuals' estimated factor scores are of direct interest. In these cases, estimated factor scores can be a viable approach, but it is important to keep in mind that they differ from true factor scores. So long as their reliability is less than perfect (which will virtually always be the case in practice), estimated factor scores will not equal true factor scores. Depending on estimation method, their correlations with each other may be biased estimates of the corresponding correlations among true factor scores, even in the population. For all factor score estimation methods, the correlations of factor score estimates with other variables external to the measurement model will always be biased (i.e., attenuation due to unreliability) (McDonald & Burr, 1967). In theory, regression factor scores are superior to simple unweighted sum scores, because they are closer to true factor scores than any other composite could be. In practice, this optimal property is threatened by sampling variability—in small samples and with low factor loadings, it can be safer to use unweighted sum scores than to rely on imprecisely estimated factor loadings to derive regression factor scores (Uanhoro, 2019; Widaman & Revelle, 2023).

Conclusion

There is a great deal of confusion among practitioners of factor analysis about the difference between estimated factor scores and actual scores on a latent variable. While many technical presentations on this

distinction exist, here we have attempted to present a relatively less technical and intuitive explanation of the difference between estimated and true values on the latent variable. We urge both researchers and methodologists to take care to clarify what they mean by "factor scores" when using this shorthand in papers; in the vast majority of cases, they mean "estimated factor scores" when discussing values obtained from observed data, and not the actual scores on the latent variable.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported by any grant.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Aguirre-Urreta, M. I., Rönkkö, M., & McIntosh, C. N. (2019). A cautionary note on the finite sample behavior of maximal reliability. *Psychological Methods*, *24*(2), 236–252. https://doi.org/10.1037/met0000176

Beauducel, A., & Rabe, S. (2009). Model-related factor score predictors for confirmatory factor analysis. *The British Journal of Mathematical and Statistical Psychology*, *62*(Pt 3), 489–506. https://doi.org/10.1348/000711008X336146

Bentler, P. M. (1968). Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33*(3), 335–345. https://doi.org/10.1007/BF02289328

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*(1), 67–74. https://doi.org/10.1207/s15327906mbr2501_8

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

Bollen, K. A., & Bauldry, S. (2011). Three cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Carpenter, T. (2023, March). *Speaking of lavaan, does anyone know when simulating data how to export the values of the latent variables used in the simulation? If I simulate a latent variable with 3 indicators @ 500 obs, I need the latent variable values IN ADDITION to the observed values? How to get?* Retrieved from https://x.com/tcarpenter216/status/1633615211893751810

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. https://doi.org/10.1037/a0033805

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. ERIC.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (p. 195–223) Lawrence Erlbaum.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis. *Methodology*, *13*(Supplement 1), 31–38. https://doi.org/10.1027/1614-2241/a000130

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*(1), 20.

Ellis, J. L., & Junker, B. W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika*, *62*(4), 495–523. https://doi.org/10.1007/BF02294640

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

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and Future*, *195*, 216.

Krijnen, W. P., Wansbeek, T., & ten Berge. J. M. (1996). Best linear predictors for factor scores. *Communications in Statistics - Theory and Methods*, *25*(12), 3013–3025. https://doi.org/10.1080/03610929608831883

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. https://doi.org/10.1037/a0024776

Logan, J. A., Jiang, H., Helsabeck, N., & Yeomans-Maldonado, G. (2022). Should I allow my confirmatory factors to correlate during factor score extraction? Implications for the applied researcher. *Quality & Quantity*, *56*(4), 2107–2131. https://doi.org/10.1007/s11135-021-01202-x

Maraun, M. D. (1996). Metaphor taken as math: Indeterminancy in the factor analysis model. *Multivariate Behavioral Research*, *31*(4), 517–538. https://doi.org/10.1207/s15327906mbr3104_6

McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, *38*(1), 75–79. https://doi.org/10.1177/001316447803800111

McDonald, R. P. (2011). Measuring latent quantities. *Psychometrika*, *76*(4), 511–536. https://doi.org/10.1007/s11336-011-9223-7

McDonald, R. P., & Burr, E. (1967). A comparison of four methods of constructing factor scores. *Psychometrika*, *32*(4), 381–401. https://doi.org/10.1007/BF02289653

McNeish, D. (2023). Psychometric properties of sum scores and factor scores differ even when their correlation is 0.98: A response to widaman and revelle. *Behavior Research Methods*, *55*(8), 4269–4290. https://doi.org/10.3758/s13428-022-02016-x

Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modelling approach. *The British Journal of Mathematical and Statistical Psychology*, *57*(Pt 1), 21–27. https://doi.org/10.1348/000711004849295

Rigdon, E. E., Becker, J.-M., & Sarstedt, M. (2019). Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. *Multivariate Behavioral Research*, *54*(3), 429–443. https://doi.org/10.1080/00273171.2018.1535420

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

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

Savalei, V. (2019). A comparison of several approaches for controlling measurement error in small samples. *Psychological Methods*, *24*(3), 352–370. https://doi.org/10.1037/met0000181

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575. https://doi.org/10.1007/BF02296196

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271–295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x

Ten Berge, J. M., Krijnen, W. P., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and Its Applications*, *289*(1–3), 311–318. https://doi.org/10.1016/S0024-3795(97)10007-6

Uanhoro, J. (2019). *When data are not so informative, it pays to choose the sum score over the factor score*. Retrieved August 18, 2023, from https://www.jamesuanhoro.com/post/2019/08/02/when-data-are-not-so-informative-it-pays-to-choose-the-sum-score-over-the-factor-score/

Velicer, W. F. (1976). The relation between factor score estimates, image scores, and principal component scores. *Educational and Psychological Measurement*, *36*(1), 149–159. https://doi.org/10.1177/001316447603600114

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Waller, N. G. (2023). Breaking our silence on factor score indeterminacy. *Journal of Educational and Behavioral Statistics*, *48*(2), 244–261. https://doi.org/10.3102/10769986221128810

Wickens, T. D. (2014). *The geometry of multivariate statistics*. Psychology Press.

Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788–806. https://doi.org/10.3758/s13428-022-01849-w

Appendix A. Factor score indeterminacy

In this appendix, we provide a geometric illustration of factor score indeterminacy. Figure A1 shows the difference between predicting observed variables and estimating factor scores. In this figure, variables are represented by vectors, and correlations between variables are cosines of angles between the vectors (e.g., Wickens, 2014). In the left panel, the variable x is predicted by the variable y. Both x and y have well-defined correlations (i.e., well-defined angles) with any third variable (here, one such variable, z, is shown), because all are observed variables. In the right panel, the latent variable f is predicted by the observed variable y (y may be a linear combination of a set of observed variables, e.g., y is proportional to $\hat{f}$). However, because f is entirely latent, we only know the exact location of y and the *length* of the residual (i.e., amount of prediction error, in the case of regression factor scores), but not its location. Stated equivalently, we know the angle between f and y (its cosine is maximum reliability, or reliability of the regression factor scores), but there are infinitely many locations where f could be that have the same angle with y. Four possible locations for f are shown, but any vector running along the wall of this "hyper-cone" (or, any vector that ends on the circle) would be a legitimate alternative representation. We include the variable z in this representation to make clear that different possible locations of F will have somewhat different correlations (angles) with z, if no other assumptions are made about z. Two such angles are shown for f_1 and f_2.

This consequence of factor score indeterminacy—that correlations between the latent variable and external variables are undefined—also has implications for identification of latent variable models that include associations between a given latent variable and other variables (beyond its observed indicators), whether latent or observed. For

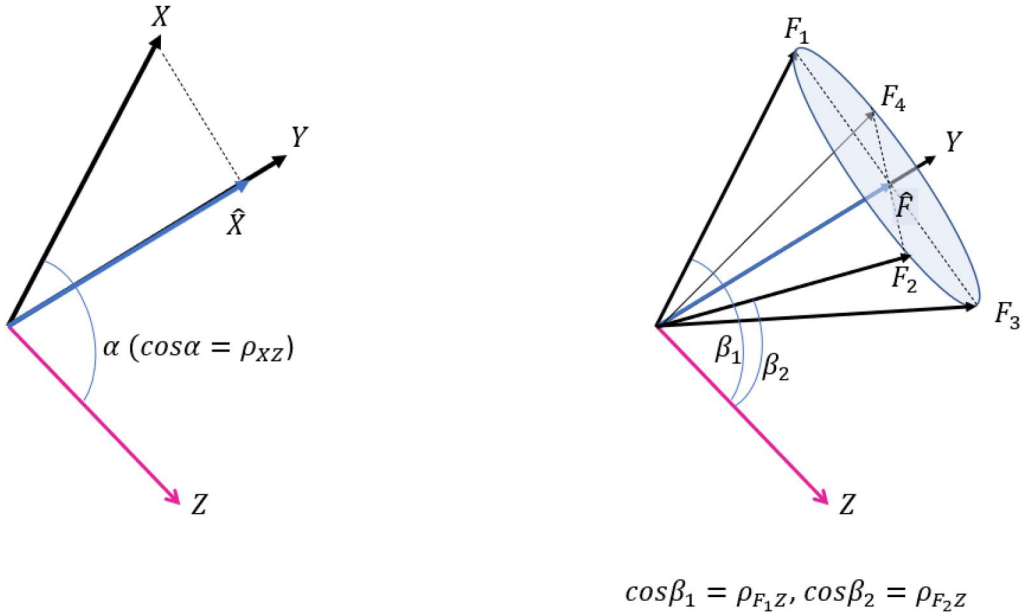

Figure A1. A geometric illustration of the difference between predicting an observed variable and predicting a latent variable (factor score estimation), illustrating factor score indeterminacy. *Note.* Left: An observed variable *x* is predicted from another observed variable, *y* (the prediction $\hat{x}$ is colinear with *y* and the dotted line represents the residual). Because *x* is observed, its location in space is known, not only relative to *y* but also to any third variable *z* (one such *z* is shown; its correlation with *x* is the cosine of its angle with *x*). Right: A latent variable *f* is predicted from an observed variable *y* (or a best linear combination of observed variables), resulting in the prediction $\hat{f}$ (the blue vector, colinear with *y*). The exact location of *f* is not known, and four possibilities are shown (f_1 to f_4); they all result in the same orthogonal projection onto the line colinear with $\hat{f}$, and all dotted lines represent the residuals from these projections (all of the same length; this squared length gives the amount of residual variance). Infinitely many *f*s are possible, one for each point on the circle around $\hat{f}$, with the radius given by the length of the residual. These different *f*s have different correlations with an arbitrary third variable (*z*), because they have different angles with *z*. Two such angles are shown. The less factor indeterminacy there is, the smaller the radius of the circle (the shorter the residual vectors) and the more similar are the angles between different *f*s and *z*.

example, if an observed variable *z* is added to the 1-factor model, its correlation with the latent variable would not be uniquely defined if *z* is also permitted to correlate with the measurement errors of all the indicators of *f*, leading to lack of identification for the latent variable model. However, if *z* is assumed to be orthogonal to all the measurement errors, its correlation with *f* becomes estimable, as this assumption reduces the dimensionality of the "hyper-cone" of possibilities for the location of *f* (see Figure A1). Admittedly showing orthogonality between *z* and measurement errors is difficult and is not attempted in this figure. It is difficult because in the three-dimensional simplification, which is what is attempted in Figure A1, a vector orthogonal to the residuals from factor score prediction will have to be shown as colinear with *y* (and $\hat{f}$), but in a high-dimensional space this does not have to be the case. The assumption of orthogonality with the measurement errors is traditionally made by the approaches of "factor score regression" (e.g., Croon, 2002; Devlieger & Rosseel, 2017): In this family of methods, one regresses the criterion (say *z*) on estimated factor scores obtained from a latent variable model, and then corrects the regression coefficient for predicting *z* from $\hat{f}$ for bias; this correction assumes that *z* and $\hat{f}$ are orthogonal controlling for *f*.

Appendix B: Simulation code

This appendix describes how to use the R package **lavaan** (Rosseel, 2012) to simulate the data used in Illustrations 1 and 4, as well as how to extract and examine regression factor scores and examine their properties. Illustrations 2, 3, 5, and 6 are straightforward extensions of 1 and 4. Full R code to reproduce all illustrations and figures can be found at osf.io/a68wm/.

We first generate a simulated dataset of size $N = 100$ from a 1-factor model with three variables in the R package **lavaan** (Rosseel, 2012). The syntax to generate data only on observed variables would be as follows, where we specify each model parameter explicitly for maximum clarity:

```
set.seed(123)
genmod0 < - '
            F =~ 0.6*Y1 + 0.6*Y2 + 0.6*Y3
            Y1 ~~ 0.64*Y1
            Y2 ~~ 0.64*Y2
            Y3 ~~ 0.64*Y3
            F ~~ 1*F
            '
data0 < - simulateData(genmod0, sample.
nobs = 100)
```

This method produces values on the *observed variables only*, which we see by examining the first few rows of the simulated data:

```
> round(head(data0),3)
      Y1         Y2        Y3
1  0.916     -1.057     1.414
2  0.823     -0.657     0.356
3 -1.413     -1.028    -1.100
4 -0.022     -0.433     0.295
5 -0.757      0.096     0.368
6 -1.509     -0.993    -1.394
```

This is because when simulating observed data from a factor model, the approach implemented in SEM software such as *lavaan* is to use the user-supplied model parameters to produce a model-implied covariance matrix of observed variables, and then use a multivariate normal data generation method (e.g., Venables & Ripley, 2013) to draw a sample of observations from a multivariate normal distribution with that matrix as the population covariance matrix.

Because our goal is to compare estimated factor scores to true factor scores, this method does not work for us. We want the generated datasets to include participants' scores on both observed and latent variables. To generate the values on *f* as well, we set up the model as a system of regression equations of indicators *y* on the latent factor *f*, so the latent variable is treated as observed. That is, instead of using the operator =~ , which indicates the creation of a latent variable on the left by observed variables listed on the right, we use the usual regression operator ~ , as follows:

```
set.seed(123)
genmod1 < - '
            Y1 ~ 0.6*F + 1*E1
            Y2 ~ 0.6*F + 1*E2
            Y3 ~ 0.6*F + 1*E3
            Y1 ~~ 0.64*Y1
            Y2 ~~ 0.64*Y2
            Y3 ~~ 0.64*Y3
            F ~~ 1*F
            '
data1 < - simulateData(genmod1, sample.nobs = 100, empirical = TRUE)
```

We can now view the first 6 rows of data to see that the dataset simulated in this way contains not only the observed variables but also the common factor *f* as well as the residuals *e*1-*e*3:

```
> round(head(data1), 3)
      Y1        Y2        Y3        F        E1        E2        E3
1  0.720     1.762    -1.638    0.537     0.025    -0.004    -0.433
2 -0.490     0.215    -2.078   -0.990    -0.385     0.404    -0.153
3 -1.687     1.166     0.094   -0.831    -0.171     0.956    -0.296
4 -0.223     1.095     1.337   -0.543    -0.034     0.922     1.430
5 -0.725     0.345    -0.486    1.300    -1.387    -0.063    -0.721
6 -2.661    -1.196    -1.157   -1.750    -0.642    -0.291    -0.630
```

In this syntax, we have also specified `empirical = TRUE`, which means that the sample covariance matrix of the resulting dataset will be exactly equal to the population model implied covariance matrix. This means that all estimated model parameters from this sample will be equal to true parameters, and there is no sampling error. To see this, we can compute the sample covariance matrix for the dataset:

```
N = dim(data1)[1]
S = cov(data1)*(N-1)/N
```

and then view it:

```
> round(S, 3)
      Y1   Y2   Y3   F   E1   E2   E3
Y1  2.00 0.36 0.36 0.6    1    0    0
Y2  0.36 2.00 0.36 0.6    0    1    0
Y3  0.36 0.36 2.00 0.6    0    0    1
F   0.60 0.60 0.60 1.0    0    0    0
E1  1.00 0.00 0.00 0.0    1    0    0
```

```
E2        0.00        1.00        0.00        0.0        0        1        0
E3        0.00        0.00        1.00        0.0        0        0        1
```

We can see that the correlations among observed variables are equal to the product of the corresponding loadings (each is 0.6), and the variances of the observed variables are 1 ($= .6^2 + .64$), because that is the sum of the squared loading and the error variance. The correlation between the factor and each indicator is equal to the corresponding factor loading. This is the method we used to generate "population datasets" in Illustrations 1, 2, 4, and 5. In Illustrations 3 and 6 where sampling variability is introduced, we simply omitted `empirical = TRUE`.

To obtain regression factor scores, we first fit a 1-factor model to the generated observed data and then use the `lavPredict()` function:

```
data1.obs <- data1[, c("Y1","Y2","Y3")]
mod1 <- 'F =~ Y1 + Y2 + Y3
'
fit1 <- cfa(mod1, data1.obs, std.lv=TRUE)
#append estimated factor scores to data generated from latent variable model:
data1[, "Fpred"] <- as.numeric(lavPredict(fit1, method = "regression"))
```

We can then examine the first 6 rows of data to see these estimated factor scores:

```
> round(head(data1), 3)
      Y1       Y2       Y3       F       E1       E2       E3      Fpred
1  0.720    1.762   -1.638   0.537   0.025   -0.004   -0.433    0.186
2 -0.490    0.215   -2.078  -0.990  -0.385    0.404   -0.153   -0.519
3 -1.687    1.166    0.094  -0.831  -0.171    0.956   -0.296   -0.094
4 -0.223    1.095    1.337  -0.543  -0.034    0.922    1.430    0.487
5 -0.725    0.345   -0.486   1.300  -1.387   -0.063   -0.721   -0.191
6 -2.661   -1.196   -1.157  -1.750  -0.642   -0.291   -0.630   -1.106
```

The following code shows how to replicate these estimated values printed by **lavaan** using Equation (16).

```
#Note: Phi = 1 because there is only one factor
Lambda = lavInspect(fit1, what = "coef")$lambda #matrix of loadings
Theta = lavInspect(fit1, what = "coef")$theta #matrix of residual variances
W = t(Lambda) %*% solve(Lambda %*% t(Lambda) + Theta) #equation 15
Fpred.manual <- as.numeric(W %*% t(data1.obs))
```

We can view the first six values to confirm that they match the values shown in the previous chunk of output:

```
> round(Fpred.manual[1:6], 3)
[1]  0.186  -0.519  -0.094   0.487  -0.191  -1.106
```

Comparing the true factor scores in the column labeled `'F'` and the estimated (regression) factor scores in the column labeled `'Fpred'`, it is clear that they are quite different values. Their aggregate properties are also quite different. Below is their variance-covariance matrix, as well as their correlation:

```
> round(var(data1[, c("F", "Fpred")]), 3)
           F  Fpred
F      1.010  0.401
Fpred 0.401 0.401
> round(cor(data1[, c("F", "Fpred")]), 3)
          F  Fpred
F      1.00  0.63
Fpred 0.63  1.00
```

In the second example (Illustration 4), we simulate data from a model with two latent factors.

```
set.seed(123)
genmod2 < - '
              Y1 ~ 0.6*F1
              Y2 ~ 0.6*F1
              Y3 ~ 0.6*F1
              Y4 ~ 0.7*F2
              Y5 ~ 0.7*F2
              Y6 ~ 0.7*F2

              Y1 ~~ 0.64*Y1
              Y2 ~~ 0.64*Y2
              Y3 ~~ 0.64*Y3
              Y4 ~~ 0.51*Y4
              Y5 ~~ 0.51*Y5
              Y6 ~~ 0.51*Y6

              F1 ~~ 1*F1 + 0.3*F2
              F2 ~~ 1*F2
              '
data2 < - simulateData(genmod2, sample.nobs = 100, empirical = TRUE)
data2.obs < - data2[, 1:6]
```

As we explain in the main text, if Equation (16) is applied to a multi-dimensional factor model (2 or more factors) then the regression factor scores for each factor will be a weighted sum of not only the indicators of that factor, but also those of other factors. To get estimated values that are a function of only the indicators of that factor, the elements of Equation (16) must be derived from the results of a 1-factor model fit to just those indicators. In the code below we show both methods.

First, in Method 1 (regression factor scores are weighted sums of indicators of all factors), regression factor scores for f_1 and f_2 are estimated from the 2-factor model.

```
mod2 < - 'F1 =~ Y1 + Y2 + Y3
          F2 =~ Y4 + Y5 + Y6
         '
fit2 < - cfa(mod2, data2.obs, std.lv = TRUE)
data2[, c("Fpred1.2F", "Fpred2.2F")] < - lavPredict(fit2, method = "regression")
```

which produces data with the following estimated factor scores:

```
> round(head(data2),3)
       Y1     Y2     Y3     Y4     Y5     Y6     F1     F2   Fpred1.2F   Fpred2.2F
1  -0.009  0.614 -0.331 -1.061 -0.751 -2.541 -0.108 -1.604   -0.087      -1.508
2   0.600  1.269 -0.453 -0.841 -1.722 -1.374  0.456 -1.281    0.319      -1.331
3   1.235 -0.352  0.608 -0.311 -0.974 -0.162 -0.279  0.169    0.447      -0.462
4   0.011 -1.059 -0.793  0.725 -0.649 -0.301 -0.522 -0.021   -0.634      -0.130
5   0.505  0.338 -0.745 -1.971 -0.250 -0.429 -0.259 -1.135   -0.076      -0.920
6   1.850  1.118  1.254  0.897  0.394  0.984  1.288  1.420    1.528       0.911
```

For interested readers, below is the computation of these estimated factor scores using Equation (16):

```
Lambda = lavInspect(fit2, what = "coef")$lambda #matrix of loadings
Psi = lavInspect(fit2, what = "coef")$psi #matrix of factor correlations
Theta = lavInspect(fit2, what = "coef")$theta #matrix of error variances
W = Psi %*% t(Lambda) %*% solve(Lambda %*% Psi %*% t(Lambda) + Theta) #equation 15
Fpred.2F.manual < - W %*% t(data2.obs)
```

which produces the same values as the previous code chunk:

```
> t(Fpred.2F.manual)[1:6, 1:2]
              F1           F2
[1,] -0.08656863 -1.5079484
[2,]  0.31901981 -1.3307737
[3,]  0.44708646 -0.4621398
[4,] -0.63448178 -0.1301551
[5,] -0.07620193 -0.9196783
[6,]  1.52847602  0.9110895
```

Next, in Method 2, f_1 and f_2 scores are estimated from two separate 1-factor models.

```
mod3 <- 'F1 =~ Y1 + Y2 + Y3'
mod4 <- 'F2 =~ Y4 + Y5 + Y6'
fit3 <- cfa(mod3, data2.obs, std.lv=TRUE)
fit4 <- cfa(mod4, data2.obs, std.lv=TRUE)
data2$Fpred1.1f <- predict(fit3)
data2$Fpred2.1f <- predict(fit4)
round(head(data2), 3)
```

which produces different factor score estimates (see the right-most two columns):

```
> round(head(data2), 3)
     Y1     Y2     Y3     Y4     Y5     Y6     F1     F2   Fpred1.2F  Fpred2.2F   F1     F2
1 -0.009  0.614 -0.331 -1.061 -0.751 -2.541 -0.108 -1.604  -0.087     -1.508    0.095 -1.539
2  0.600  1.269 -0.453 -0.841 -1.722 -1.374  0.456 -1.281   0.319     -1.331    0.494 -1.392
3  1.235 -0.352  0.608 -0.311 -0.974 -0.162 -0.279  0.169   0.447     -0.462    0.520 -0.512
4  0.011 -1.059 -0.793  0.725 -0.649 -0.301 -0.522 -0.021  -0.634     -0.130   -0.642 -0.080
5  0.505  0.338 -0.745 -1.971 -0.250 -0.429 -0.259 -1.135  -0.076     -0.920    0.034 -0.937
6  1.850  1.118  1.254  0.897  0.394  0.984  1.288  1.420   1.528      0.911    1.473  0.805
```

Below is the covariance matrix of the true factor scores, which is the same as the model-estimated matrix Ψ (off by a factor of $(N-1)/N$), because the data were generated to reproduce the population parameters exactly:

```
> round(var(data2[, c("F1", "F2")]), 3)
      F1    F2
F1 1.010 0.303
F2 0.303 1.010
```

Below are the covariance and correlation matrices of the estimated factor scores, both those estimated from the 2-factor model and then those estimated from separate one-factor models:

```
> round(var(data2[, c("Fpred1.2F", "Fpred2.2F")]), 3)
          Fpred1.2F Fpred2.2F
Fpred1.2F     0.644     0.273
Fpred2.2F     0.273     0.754
> round(cor(data2[, c("Fpred1.2F", "Fpred2.2F")]), 3)
          Fpred1.2F Fpred2.2F
Fpred1.2F     1.000     0.391
Fpred2.2F     0.391     1.000
> round(var(data2[, c("Fpred1.1f", "Fpred2.1f")]), 3)
          Fpred1.1f Fpred2.1f
Fpred1.1f     0.634     0.141
Fpred2.1f     0.141     0.750
> round(cor(data2[, c("Fpred1.1f", "Fpred2.1f")]), 3)
          Fpred1.1f Fpred2.1f
Fpred1.1f     1.000     0.205
Fpred2.1f     0.205     1.000
```

When estimated factor scores are obtained from 2 separate 1-factor models, the true correlation of 0.3 is estimated to be 0.2, and when estimated factor scores are obtained from a 2-factor model, it is estimated to be 0.39.

Lastly, here are the correlations among the pairs of true factor scores and estimated factor scores:

```
> round(cor(data2[, "F1"], data2[, "Fpred1.1f"]), 3)
          F1
[1,] 0.792
> round(cor(data2[, "F2"], data2[, "Fpred2.1f"]), 3)
          F2
[1,] 0.862
> round(cor(data2[, "F1"], data2[, "Fpred1.2F"]), 3)
[1] 0.798
> round(cor(data2[, "F2"], data2[, "Fpred2.2F"]), 3)
[1] 0.864
```

These correlations are higher for the factors that had higher factor loadings in the model specification. Whether they are estimated from the full 2-factor vs. 1-factor models does not appear to make a difference.