2

Creating a Confirmatory Factor Analysis Model

This chapter will focus on creating and specifying a confirmatory factor analysis (CFA) model, beginning with the role of theory and prior research in CFA. We will then discuss how a CFA model is specified, examining the role of observed and latent variables and model parameters, followed by a discussion of the importance of model identification, scaling latent variables, and estimation methods. We will end this chapter with a detailed example of testing a CFA model.

Specifying the Model

Theory and/or prior research are crucial to specifying a CFA model to be tested. As noted in Chapter 1, the one-factor solution of the Rosenberg Self-Esteem Scale was tested based on the conceptualization of self-esteem as a global (i.e., unitary) factor, although the existing exploratory factor analysis (EFA) work found two factors. Early in the process of measurement development, researchers may rely entirely on theory to develop a CFA model. However, as a measure is used over time, CFA can be used to replicate EFA or other analyses that have been conducted on the measure. In the Professional Opinion Scale (POS) example discussed in Chapter 1, Abbott's (2003) initial CFA was based both on underlying theory and an earlier EFA, whereas the Greeno et al. (2007) CFA was based on Abbott's (2003) earlier CFA work. Confirmatory factor analysis may not be an appropriate analysis to use if there is no strong underlying foundation on which to base the model, and more preliminary work, such as EFA or theory development, may be needed. This chapter includes many terms that are used in CFA, which will be defined here and in the Glossary. See Figure 2.1 for a basic CFA model with variables and parameters labeled.

Observed Variables

As discussed in Chapter 1, observed variables are those items that are directly observed, such as a response to a question. In CFA models, observed variables are represented by rectangles.

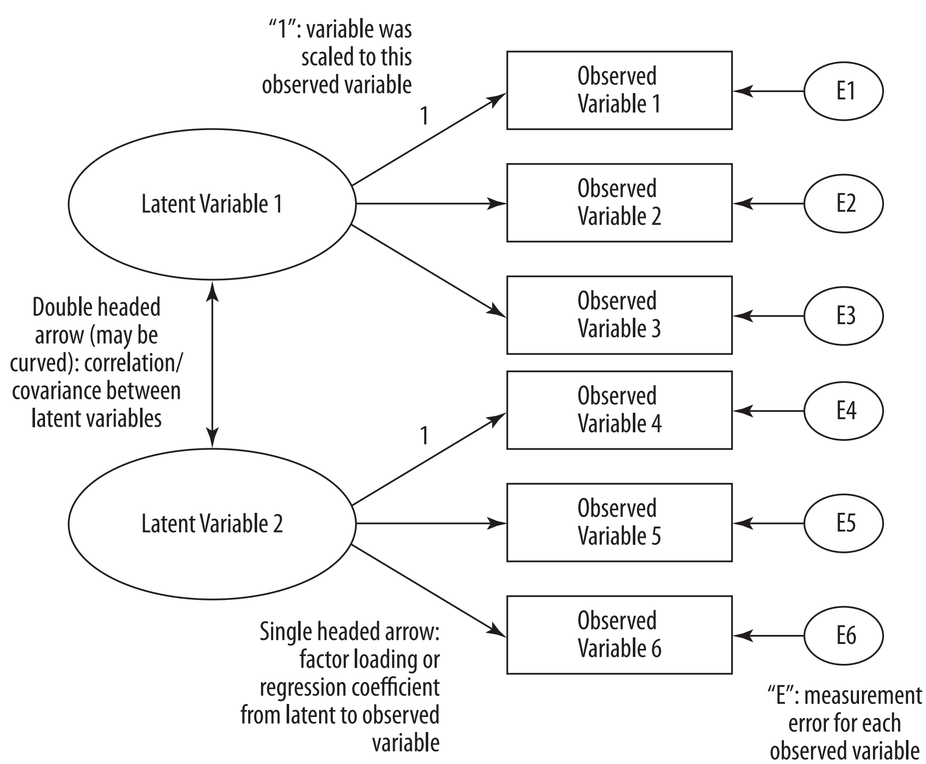

Figure 2.1 CFA Model With Parameters Labeled

Latent Variables

Latent variables are the underlying, unobserved constructs of interest. Ovals are used to represent latent variables in CFA models (sometimes circles are also used, but we will use ovals in this book). There are two types of latent variables: exogenous and endogenous. Exogenous variables are not caused by other variables in the model; they are similar to independent variables (IV), X, or predictors in regression analyses. Endogenous variables are–at least theoretically–caused by other variables, and in this sense they are similar to dependent variables (DV), Y, or outcome variables in regression analyses. In complex models, some variables may have both exogenous and endogenous functions.

CFA Model Parameters

Model parameters are the characteristics of the population that will be estimated and tested in the CFA. Relationships among observed and latent variables are indicated in CFA models by arrows going from the latent variables to the observed variables. The direction from the latent to the observed variable indicates the expectation that the underlying construct (e.g., depression) causes the observed variables (e.g., symptoms of unhappiness, feeling blue, changes in appetite, etc.). The factor loadings are the regression coefficients (i.e., slopes) for predicting the indicators from the latent factor. In general, the higher the factor loading the better, and typically loadings below 0.30 are not interpreted. As general rules of thumb, loadings above 0.71 are excellent, 0.63 very good, 0.55 good, 0.45 fair, and 0.32 poor (Tabachnick & Fidell, 2007). These rules of thumb are based on factor analyses, where factor loadings are correlations between the variable and factor, so squaring the loading yields a variance accounted for. Note that a loading of 0.71 squared would be 50% variance accounted for, whereas 0.32 squared would be 10% variance accounted for. In CFA, the interpretation of the factor loadings or regression coefficients is a little more complex if there is more than one latent variable in the model, but this basic interpretation will work for our purposes.

Whereas each indicator is believed to be caused by the latent factor, there may also be some unique variance in an indicator that is not accounted for by the latent factor(s). This unique variance is also known as measurement error, error variance, or indicator unreliability (see E1 to E6 in Figure 2.1).

Other parameters in a CFA model include factor variance, which is the variance for a factor in the sample data (in the unstandardized solution), and error covariances, which are correlated errors demonstrating that the indicators are related because of something other than the shared influence of the latent factor. Correlated errors could result from method effects (i.e., common measurement method such as self-report) or similar wording of items (e.g., positive or negative phrasing).

The relationship between two factors, or latent variables, in the model is a factor correlation in the completely standardized solution or a factor covariance in unstandardized solutions. Factor correlations represent the completely standardized solution in the same way that a Pearson's correlation is the "standardized" relationship between two variables (i.e., ranges from –1 to +1 and is unit-free—it does not include the original units of measurement). Similarly, factor covariances are unstandardized and include the original units of measurement just as variable covariances retain information about the original units of measurement and can range from negative infinity to positive infinity. Factor covariances or correlations are shown in CFA models as two-headed arrows (usually curved) between two latent variables.

Identification of the Model

Confirmatory factor analysis models must be identified to run the model and estimate the parameters. When a model is identified, it is possible to find unique estimates for each parameter with unknown values in the model, such as the factor loadings and correlations. For example, if we have an equation such as $a + b = 44$, there are an infinite number of combinations of values of a and b that could be used to solve this equation, such as $a = 3$ and $b = 41$ or $a = -8$ and $b = 52$. In this case, the model (or the equation) is underidentified because there are not enough known parameters to allow for a unique solution—in other words, there are more unknowns (*a* and *b*) than there are knowns (44) (Kline, 2005; Raykov & Marcoulides, 2006). Models must have degrees of freedom (*df*) greater than 0 (meaning we have more known than unknown parameters), and all latent variables must be scaled (which will be discussed later in this chapter) for models to be identified (Kline, 2005). When we meet these two conditions, the model can be solved and a unique set of parameters estimated. Models can be under-, just-, or overidentified.

Underidentified Models

Models are underidentified when the number of freely estimated parameters (i.e., unknowns) in the model is greater than the number of knowns. Underidentified models, such as the $a + b = 44$ example given earlier, cannot be solved because there are an infinite number of parameter estimates that will produce a perfect fit (Brown, 2006). In this situation we have negative *df*, indicating that the model cannot reach a unique solution because too many things are left to vary relative to the number of things that are known. The number of unknowns can be reduced by fixing some of the parameters to specific values. For example, if we set $b = 4$ in the aforementioned equation, then *a* can be solved because we know have more knowns (*b* and 44) than unknowns (*a*).

Just-Identified Models

Models are just-identified when the number of unknowns equals the number of knowns and $df = 0$. In this situation, there is one unique set of parameters that will perfectly fit and reproduce the data. Although this may initially sound like a great idea (What could be wrong with a perfectly fitting model?), in practice, perfectly fitting models are not very informative because they do not allow for model testing.

Overidentified Models

Models are overidentified when the number of unknowns is smaller than the number of knowns and *df* are greater than 0. Our $a + b = 44$ example stops working here because it is too simplistic to illustrate overidentified models, but Kline (2005) provides a nice example of how this works with sets of equations if you are interested in more information on identification of models. The difference between the number of knowns and unknowns is equal to the degrees of freedom (*df*) for the model. When a model is overidentified, goodness of fit can be evaluated and it is possible to test how well the model reproduces the input variance covariance matrix (Brown, 2006). Because we are interested in obtaining fit indices for CFA models, we want the models to be overidentified.

Scaling Latent Variables

As stated earlier, in addition to having *df* greater than 0, the second condition for model identification is that the latent variables have to be scaled. Scaling the latent variable creates one less unknown. Because latent variables are unobserved, they do not have a pre-defined unit of measurement; therefore, the researcher needs to set the unit of measurement. There are two ways to do this. One option is to make it the same as that of one of the indicator variables. The second option is to set the variance equal to 1 for the latent variable. In general, the first option is the more popular (Brown, 2006). Although these two options generally result in similar overall fit, they do not always do so and it is important to realize that the option chosen for scaling the latent variable may influence the standard errors and results of the CFA (Brown, 2006; Kline, 2005).

Scaling the latent variable (or setting its unit of measurement) is a little like converting currency. Imagine that you are creating a latent variable for cost of living across the United States, United Kingdom, and France, and you have three indicators—one in U.S. dollars, one in British pounds, and the other in Euros. Dollars, pounds, and Euros all have different scales of measurement, but the latent variable can be scaled (using the aforementioned option 1) to any one of these. If scaled to U.S. dollars, the latent variable will be interpretable in terms of dollars. But, the latent variable could also be scaled to either pounds or Euros—whichever will be most interpretable and meaningful for the intended audience.

Determining Whether a Model is Identified

As discussed earlier, you will want your CFA models to be overidentified so that you can test the fit of your model. Assuming that the latent variables have been properly scaled, the issue that will determine whether a model is identified is the number parameters to be estimated (i.e., the unknowns) relative to the number of known parameters. There are several rules of thumb available for testing the identification of models, such as the *t*-Rule and the Recursive Rule; however, these rules provide necessary but not sufficient guidance (Reilly, 1995), meaning that meeting the rule is necessary for identification, but the model may still be underidentified because of other issues. Fortunately for our purposes, SEM software used to conduct CFA will automatically test the identification of the model and will provide a message if the model is under- or just-identified, which should be sufficient for most situations.

Estimation Methods

"The objective of CFA is to obtain estimates for each parameter of the measurement model (i.e. factor loadings, factor variances and covariances, indicator error variances and possibly error covariances) that produce a predicted variance-covariance matrix (symbolized as $\mathbf{\Sigma}$) that represents the sample variance-covariance matrix (symbolized as $\mathbf{S}$) as closely as possible" (Brown, 2006, p. 72). In other words, in CFA we are testing whether the model fits the data. There are multiple estimation methods available for testing the fit of an overidentified model, and we briefly discuss several. The exact process of how the model is estimated using different estimation methods is beyond the scope of this book, but I will provide a general idea of how it works. Fitting a model is an iterative process that begins with an initial fit, tests how well the model fits, adjusts the model, tests the fit again, and so forth, until the model converges or fits well enough. This fitting process is done by the software used and will generally occur in a "black box" (i.e., it will not be visible to you).

This iterative fitting process is similar to having a garment, such as a wedding dress or suit, fitted. You begin with your best guess of what size should fit, and then the tailor assesses the fit and decides if adjustments are needed. If needed, the adjustments are made and then the garment is tried on again. This process continues until some fitting criteria are reached (i.e., the garment fits properly) or some external criteria (i.e., the wedding date) forces the process to stop. If the fitting criteria are reached, then the fit is said to converge and we have a well-fitting garment (or CFA model). But, if the fitting criteria are not reached, we may be forced to accept a poorly fitting garment (or CFA model) or to begin again with a new size or style (or a different CFA model). Just as there are multiple tailors available who will use slightly different fitting criteria, there are also multiple estimation methods available for CFA—each with its own advantages and disadvantages.

Some of the estimation methods that you may see in the literature include maximum likelihood (ML), weighted least squares (WLS), generalized least squares (GLS), and unweighted least squares (ULS). Although GLS and ULS are available in Amos 7.0 and may appear in the literature, both are used with multivariate normal data (Kline, 2005), and if data are multivariate normal, then ML is a better estimation procedure to use, so we will not discuss GLS and ULS. For this introductory text on CFA, we will limit our discussion to the best of the common estimation methods that are available in Amos 7.0.

Maximum Likelihood

Maximum likelihood (ML) is the most commonly used estimation method. Maximum likelihood "aims to find the parameter values that make the observed data most likely (or conversely maximize the likelihood of the parameters given the data)" (Brown, 2006, p. 73). Maximum likelihood estimation is similar (but not identical) to the ordinary least squares criterion used in multiple regression (Kline, 2005). It has several desirable statistical properties: *(1)* it provides standard errors (SEs) for each parameter estimate, which are used to calculate *p*-values (levels of significance), and confidence intervals, and *(2)* its fitting function is used to calculate many goodness-of-fit indices.

There are three key assumptions for ML estimation. First, this estimation procedure requires large sample sizes (sample size requirements will be discussed in more detail in Chapter 3). Second, indicators need to have continuous levels of measurement (i.e., no dichotomous, ordinal, or categorical indicator variables). Third, ML requires multivariate normally distributed indicators (procedures for assessing normality will be discussed in Chapter 3). ML estimation is robust to moderate violations, although extreme non-normality results in several problems: *(1)* underestimation of the SE, which inflates Type I error; *(2)* poorly behaved (inflated) χ^2 tests of overall model fit and underestimation of other fit indices (e.g., TLI and CFI, which will be discussed further in Chapter 4); and *(3)* incorrect parameter estimates. When there are severe violations of the assumptions, formulas are available for calculating robust SE estimates and the chi-square statistic as long as there are no missing data (see Gold, Bentler, & Kim, 2003). Importantly, the effects of non-normality worsen with smaller sample sizes (Brown, 2006). In addition, when the violations of the underlying assumptions are extreme, ML is prone to Heywood cases (i.e., parameter estimates with out-of-range values), such as negative error variances. In addition, minor misspecifications of the model may result in "markedly distorted solutions" (Brown, 2006, p. 75). Therefore, ML should not be used if the assumptions are violated.

Other Estimation Methods

If the model includes one or more categorical indicator variables or if there is extreme non-normality, ML is not appropriate to use and there are several alternative estimation methods available: *(1)* WLS, which is called asymptotically distribution-free (ADF) in Amos 7.0; *(2)* robust weighted least squares (WLSMV); and *(3)* ULS (Brown, 2006). However, each of these estimation methods has limitations, as discussed below. For non-normal continuous indicators, ML with robust SE and χ^2 (MLM) can be used. At this time, the Mplus program has the best options for handling categorical data because of the availability of the WLSMV estimator (Brown, 2006).

Of the estimation methods that are broadly available, including in Amos, ADF "estimates the degree of both skew and kurtosis in the raw data" and therefore makes no assumptions about the distribution of the data (Kline, 2005, p. 196). Although this addresses the problem of non-normality in the data, a drawback of this approach is that it generally requires very large sample sizes of 200 to 500 for simple models and thousands of cases for complex models (Kline, 2005, p. 196). In addition to the sample size requirements, Brown (2006) notes that ADF or WLS does not perform well with categorical data, especially when samples are not sufficiently large.

Gold et al. (2003) compared ML and ADF estimation methods for non-normal incomplete data and found that direct ML (the form of ML that can handle missing data, which is available in Amos and other software packages) performs better than ADF with pairwise deletion, regardless of missing data mechanism (p. 73). Gold et al. (2003) concluded that ADF should not be used with missing data, and if there are missing data, even when there is non-normality, "ML methods are still preferable, although they should be used with robust standard errors and rescaled chi-square statistics" (p. 74). Savalei and Bentler (2005) also concluded that direct ML is generally recommended when there are missing data and non-normality. Missing data and normality will be discussed further in Chapter 3.

In Amos Graphics 7.0, the available estimation methods are ML, GLS, ULS, scale-free least squares, and ADF. Only ML can be used if there are missing data. If there are missing data and one of the other estimation methods is needed, then some form of data imputation needs to be done before the other estimation method can be used in Amos. Readers who are likely to have problematic data may want to consider using a software package other than Amos.

Testing a Confirmatory Factor Analysis Model Example

In this section, we will use Amos 7.0 to test a CFA model using the Maslach Burnout Inventory (MBI; Maslach, Jackson, & Leiter, 1996). Brief instructions for using Amos 7.0 to conduct this analysis are provided in Appendix A. The data for this example are from a study of U.S. Air Force Family Advocacy Program (FAP) workers (Bean, Harrington, & Pintello, 1998; Harrington, Bean, Pintello, & Mathews, 2001). The sample includes 139 FAP workers and the response rate for the survey was 74%. Before continuing, it is important to note that this sample size is considered medium (Kline, 2005) for this analysis (although one can find published CFA articles with similar and even smaller size samples). Therefore, it is offered only as an example data set that readers can play with, not one from which conclusions should be drawn. Ideally, the sample size would be larger, as will be discussed in Chapter 3.

Specifying the Model

The MBI was developed in the late 1970s by Maslach and Jackson to measure burnout in human service workers. It is considered the most widely accepted and often used self-report index of burnout in research studies and employee assessment. This 22-item self-report scale treats burnout as a continuous variable that can be divided into three components: emotional exhaustion (EE), depersonalization (DP), and personal accomplishment (PA). Each item is measured on a seven-point Likert-type scale assessing the frequency of occurrence (ranging from 0 = never to 6 = a few times a day). For EE and DP, higher scores indicate higher levels of burnout, with higher levels of emotional exhaustion and depersonalization, respectively. For PA, higher scores indicate lower levels of burnout and higher levels of personal accomplishment. Maslach suggests that each subscale be scored separately rather than as a composite because this provides the best representation of the multidimensional nature of burnout as a construct (Schaufeli & Van Dierendonck, 1995).

As discussed earlier, specifying the model to be tested should be based on theory and prior research. There has been extensive work on the MBI, including a CFA on the three-factor MBI in a sample of child welfare workers (Drake & Yadama, 1995). There has also been extensive debate about how the three factors are related to burnout, whether they are all components of burnout, or whether EE is really the indicator of burnout, with PA and DP being related but separate constructs.

Like Drake and Yadama (1995), we will begin by testing the three-factor structure of the MBI as defined by Maslach et al. (1996). The observed variables for the model are the 22 items that participants responded to, and the latent variables are the three factors identified by Maslach et al. (1996). The indicators for each latent variable were chosen based on scoring instructions provided by Maslach and colleagues (1996). Because the three factors are believed to be related to each other, covariances (or correlations) among the latent variables are included in the model (shown as the two headed curved arrows in Figure 2.2 below).

Identification of the Model

The MBI CFA model is overidentified with 206 *df*, which means that there are fewer parameters to be estimated than there are known parameters. Each latent variable is scaled, with the path coefficient for one observed variable being set to "1" for each latent variable.

Estimation Method

Maximum likelihood (ML) estimation was used for this model. The MBI observed variables can be treated as continuous and the data are approximately normally distributed (data considerations will be discussed in Chapter 3), making ML a reasonable estimation method to use. It should be noted that the sample size for this example is smaller (n = 139) than desired for this or any other CFA estimation procedure, but these data are used for example purposes only.

Model Fit

All the observed variables are significantly related to the latent variables and EE and DP are significantly correlated as expected. However, contrary to

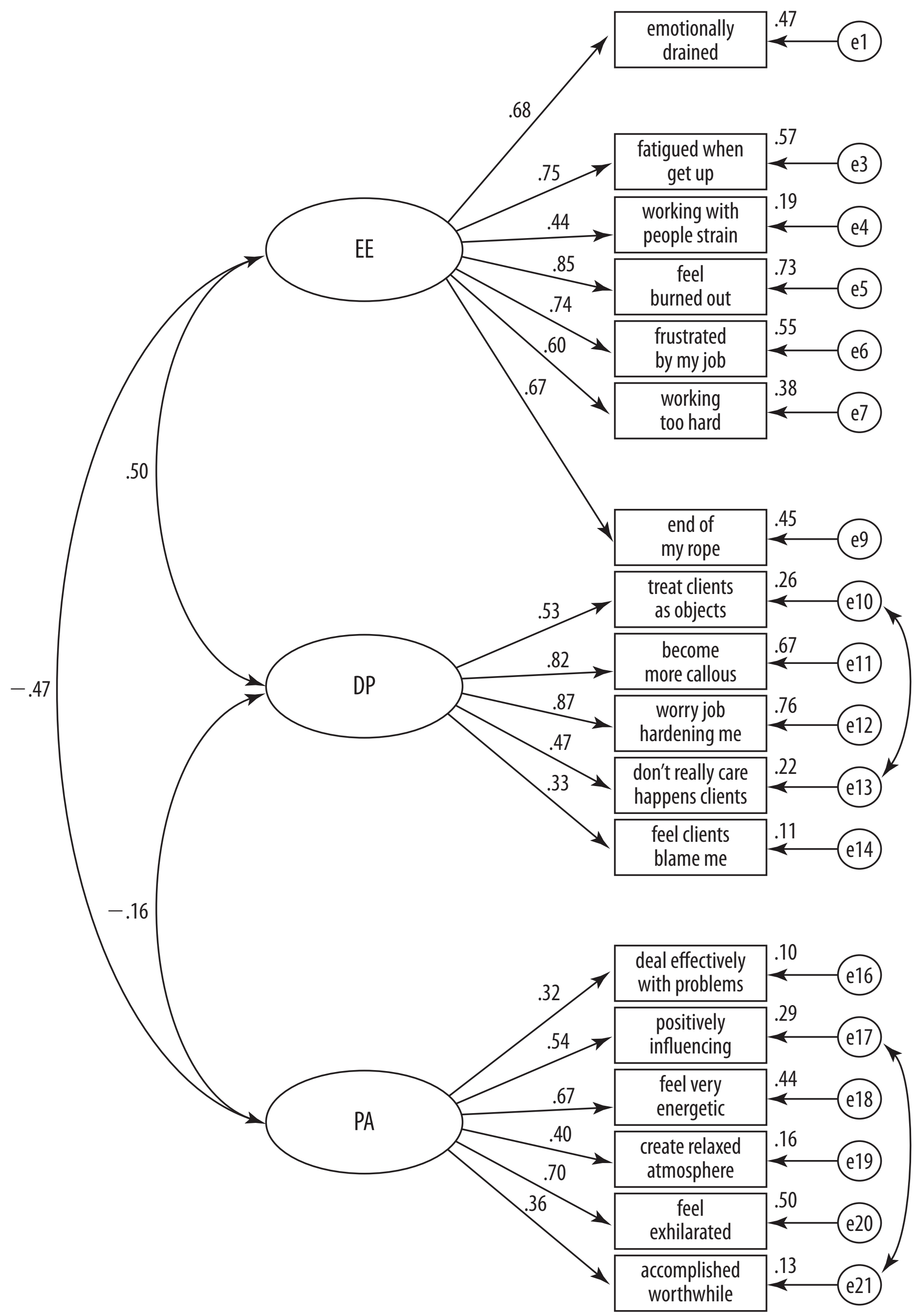

Figure 2.2 Respecified MBI CFA Model Standardized Output (Amos 7.0 Graphics)

expectations, PA and DP are not correlated ($r = -0.127$; $p = 0.261$), and PA and EE are not correlated ($r = -0.226$; $p = 0.056$). Although the model tested was based on a well-developed and tested measure, the model does not fit as well as desired. We will discuss assessing model fit in detail in Chapter 4.

Model Respecification

Drake and Yadama (1995) also found that the 22-item, three-factor MBI CFA did not fit well. To respecify their model (respecification will be discussed in more detail in Chapter 4, but briefly to respecify means to revise), they examined intercorrelations among items and deleted two items (2 and 16) that were very similar in content and highly correlated with other items. Drake and Yadama (1995) also examined squared multiple correlations and dropped two items (4 and 21) with very low squared multiple correlations. Finally, modification indices suggested that allowing the error terms to be correlated for items 5 and 15 on the DP scale and items 9 and 19 on the PA scale would improve model fit; because both of these changes seemed reasonable, the error terms were allowed to covary. Drake and Yadama's final model fit well and indicators loaded on latent variables as expected.

Using Drake and Yadama's (1996) prior work as guidance, the MBI CFA was respecified according to their final model (i.e., 18 items, three factors, and adding two error covariances). Similarly to Drake and Yadama's (1996) findings, the respecified model fits much better than the original 22-item, three-factor model. Figure 2.2 shows the standardized output for the final model. All regression weights in the model are significant and indicators load on the expected latent variables, EE and DP are significantly correlated ($p < 0.0005$), and EE and PA are significantly correlated ($p = 0.013$); the correlation between PA and DP is nonsignificant ($p = 0.193$).

Conclusion

The respecified model fit the data adequately, supporting the modified structure reported by Drake and Yadama (1995). The changes made to the model by Drake and Yadama (1995) were data-driven, and they noted that their findings should be considered preliminary with further CFA work needed with other samples. The findings in this example cautiously (because of the small sample size) suggest support for the 18-item model Drake and Yadama (1995) reported, rather than the original Maslach et al. (1996) 22-item model.

Chapter Summary

This chapter focused on creating and specifying a CFA model, including the use of theory and prior research. Observed and latent variables, CFA model parameters, model identification, and scaling the latent variables were defined, and conventions for drawing CFA models were presented. Estimation methods used in the CFA literature were briefly discussed, and ML estimation was discussed in detail. Finally, a detailed example of a CFA on the MBI was presented.

Suggestions for Further Reading

See Arbuckle (2006) for much more information on using Amos 7.0 Graphics. Byrne's (1998, 2001a, 2006) books on structural equation modeling with LISREL, Amos, and EQS (respectively) provide a number of CFA examples using these software packages. Reilly (1995) provides instructions and examples using the Rank Rule to determine whether a CFA model will be identified. Brown (2006) provides more information on how the estimation methods and fitting functions work. See Drake and Yadama (1995) for more detail on how they conducted the CFA that was replicated in this chapter. See Gold, Bentler, and Kim (2003) and Savalei and Bentler (2005) for more information on the Monte Carlo studies they conducted to compare ML and ADF estimation methods with missing data and non-normality.