Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). Applied multivariate reserach: Design and interpretation. Thousand Oaks, CA: Sage.

CHAPTER 13A

# Confirmatory Factor Analysis

## Exploratory Versus Confirmatory Strategies

The previous chapter devoted to principal components and factor analysis represented an approach that is generically labeled *exploratory factor analysis.* It is exploratory in the sense that researchers adopt the inductive strategy of determining the factor structure empirically. Simply put, researchers allow the statistical procedure to examine the correlations between the variables and to generate a factor structure based on those relationships. From the perspective of the researchers at the start of the analysis, any variable may be associated with any component or factor. Confirmatory factor analysis, by contrast, requires researchers to use a deductive strategy. Within this strategy, the factors and the variables that are held to represent them are postulated at the beginning of the procedure rather than emerging from the analysis. The statistical procedure is then performed to determine how well this hypothesized theoretical structure fits the empirical data.

## The Transition From Exploratory to Confirmatory Analysis

The transition that we are about to make between exploratory and confirmatory analysis is a fairly substantial one. The topics we have covered thus far, including exploratory factor analysis, are based on the general linear model. In this model, the solution to many of the statistical procedures we have studied involves determining the weights that a set of variables will be assigned when we form them into a linear composite or variate. One example is the multiple regression equation, a weighted linear composite of the predictor (independent variables) that best predicts the criterion (dependent) variable. A second example can be drawn from our discussion of MANOVA (multivariate analysis of variance) where a composite variable is computed as a vector representing the weighted sum of the dependent variables. And a third example is principal components analysis in which the component is a weighted linear composite of the variables in the analysis.

Understanding the nature of the variate is a primary goal of the general linear model. Accordingly, the general linear model determines the weights of the variables in the linear combination that produces the variate. As an alternative, you could potentially deduce from a theory or hypothesize from a less developed theory the weights that the variables should take on well in advance of the statistical analysis. You could use the general linear model to confirm or validate the weights for you because you would have determined them for yourself before you engaged in the statistical analysis. Specifically, the general linear model evaluates how well your hypothesized structure fits the data. This is the objective of structural equation modeling.

Confirmatory factor analysis is a specialized case of structural equation modeling. This statistical procedure allows (requires) researchers to hypothesize a particular model or factor structure that they believe underlies the variables measured in the study. Confirmatory factor analysis will then estimate the value of the parameters that tie the variables together (e.g., the pattern/structure coefficients), thus completing the description of the model, and will provide indexes that assess the quality of the fit between the model and the data.

## A Structural Approach to Exploratory and Confirmatory Factor Analysis

A useful way to transition to confirmatory factor analysis is to use a structural or path analysis framework to characterize the exploratory approach because this is the framework within which we need to understand the confirmatory strategy. This has the advantage that we can also talk about the differences between principal components and factor analysis within the general domain of exploratory factor analysis. We will start with a quick overview of structural models; a more detailed account will be given in Chapter 14A "Causal Modeling: Path Analysis and Structural Equation Modeling." Once we can see the general structure of such a model, we will discuss principal components analysis. Then we will talk about factor analysis using principle axis factoring as our exemplar and contrast it with principal components analysis. Finally, we will present the confirmatory approach and contrast it with principle axis factoring.

## Very Brief Overview of Structural Models

Many readers are somewhat familiar with these sorts of models. They are presented in the form of "flow" diagrams and are commonly published in the research literature in the context of path analysis. Path analysis and confirmatory factor analysis can be thought of as two applications of the more general approach known as structural equation modeling (Hoyle, 1995). The remaining chapters in the book, including this one, are designed to tie these types of designs together.

At a global level, causal models have three kinds of elements: text inside geometric shapes, lines with arrows pointing to or away from these shapes, and coefficients on the lines. The lines with arrows show the direction of hypothesized cause or inference, and we may refer to them occasionally as paths (in the sense of inferential flow and not to suggest that we are performing a path analysis per se). For example, an arrow pointing from *X* to *Y* can have several related interpretations, depending on the context. Among the ways to verbalize the meaning conveyed by the arrow are the following:

- *X* is hypothesized to cause *Y*
- *Y* follows (in some sense or context) from *X*
- One way that *X* can be made apparent is by measuring *Y*
- *X* in combination with other *X*s combine to form or define *Y*

The geometric shapes enclosing text represent variables of some kind. Rounded geometric shapes—that is, circles and ovals—represent latent variables. These are variables that are not directly measured in the study but are either constructed by the statistical procedure or hypothesized by the researchers (and specified in the model). The components in a principal components analysis and the factors in a factor analysis are examples of latent variables.

The other types of geometric shapes used in these sorts of models are rectangles or squares. Such squared-off shapes are used to denote measured variables—that is, variables that have been directly observed in the study, such as items on an inventory. These variables are also called *observed variables* (Raykov & Marcoulides, 2000).

## The Structure of Principal Components Analysis

We have presented a portion of a principal components analysis in a structural equation model format in Figure 13a.1. To keep the figure simple enough to use for this discussion, we show only one component and only two measured variables associated with that component. We will

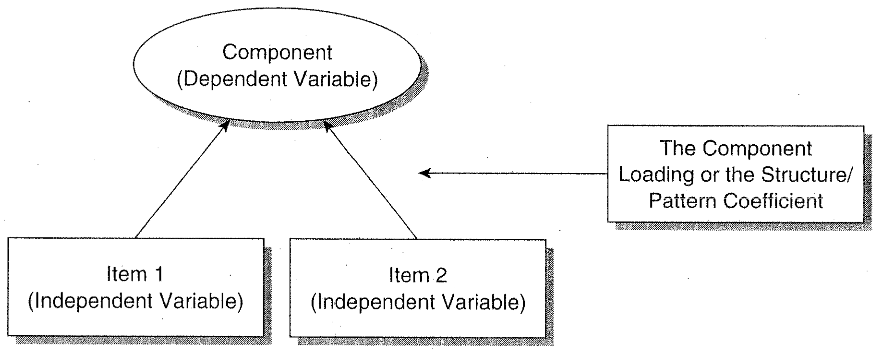


**Figure 13a.1** Diagram of a Principle Component

show more of this analysis in a moment, but certain features of this model are noteworthy.

The principle component is contained in a circle, telling us that it is a latent variable. The inventory items are contained in squares, telling us that they are measured variables in the study. But now we have an additional ingredient. The arrows or paths point from the items toward the component, telling us that the causal or inferential flow is inductive. This treatment is in accord with Pedhazur and Schmelkin (1991) who state that "components in *PCA* [principal components analysis] may be conceived of as dependent variables whose values are obtained by differentially weighting the indicators, which are treated as the independent variables" (p. 598). Within this framework, the items, taken together, cause or define the factor. Recall that variables have weights (or correlations) with the factors quantified by pattern or structure coefficients. If we were to include these coefficients, they would be placed on the connectors between the measured variables and the latent variable. This model makes clear the inductive nature of principal components analysis.

The flow of causal inference also defines the roles that these variables play in the analysis. Independent variables cause, predict, or infer dependent variables. Because the causal inference flows from the measured variables, the measured variables are the independent variables, whereas the component is the dependent variable (Hauser, 1972). The measured variables are called different names by different authors: "formative indicators" by Pedhazur and Schmelkin (1991), "producers" by Costner (1969), and "caused indicators" by Blalock (1971). We will tend to refer to them as *indicators* or *indicator variables*. Because the causal inference flows to the component, it is considered to be the dependent variable in the model.

A more complete but still simplified model of a principal components model is shown in Figure 13a.2. In this figure, we show three components drawn within circles to represent the idea that they are latent variables, and six indicators (measured variables or inventory items). Recall from the previous chapter that the factor matrix shows the weights for all indicator variables on all components. Thus, we have drawn arrows from each indicator to each component.

We also hope or expect that some indicators will correlate more strongly with one component and other indicators will correlate more strongly with another component. In Figure 13a.2, we have placed Items 1 and 2 under **Component 1**, Items 3 and 4 under **Component 2**, and Items 5 and 6 under **Component 3** to show which items are best indicators of each component. The interpretation of the components would be based on what items were most strongly associated with them. For example, **Component 1** would be interpreted as whatever Items 1 and 2 had in common.

In this more complete model, we have included curved double-pointed arrows between pairs of components. In structural equation models, double-pointed lined arrows depict the notion that we are addressing the issue of correlations. In principal components analysis, your rotation strategy is going to be either varimax (which keeps factors uncorrelated) or oblique (which allows factors to be correlated). This specification, which we took for granted in the last chapter, clearly applies to all the components. That is, in a varimax rotation, all the components remain uncorrelated, and in an oblique rotation, all the factors are allowed a certain amount of correlation. We have visually represented this situation by using the double curved-arrow lines between all possible pairs of components.

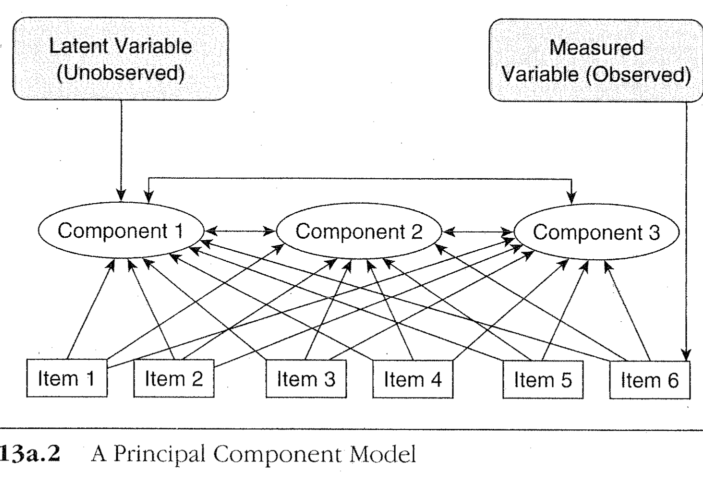


**Figure 13a.2** A Principal Component Model

## The Structure of Principal Axis Factor Analysis

We can contrast principal components analysis with principle axis factor analysis, one of the more popular factor analytic techniques. We have presented a portion of a principle axis analysis in a structural model format in Figure 13a.3. Again to keep the figure simple enough to use for this discussion, we show only one factor and only two measured variables associated with that component.

Figure 13a.3 is analogous to our model for a principal components analysis but there are some very important differences between that one and this one. Here, the factor, still a latent variable as shown by the circle in which it is drawn, is conceived of as the independent variable, and the indicator variables are the dependent variables. Note that the direction of the causal flow is from the latent variable or factor to the measured variables. This is one of the major differences between principal components and factor analysis. In factor analysis, the factor is seen as the focus point, and the measured variables are some of the ways that aspects of the factor can be measured. The term *indicator variable* is very descriptive of how the measured variables are viewed in the analysis; they are the filter through which the factor is known to us.

A more complete but still simplified factor model is drawn in Figure 13a.4. Again, there are six indicators and three factors pictured in a structure similar to what we did with the principal components analysis. Inference still flows from the latent factors to the indicators, and double arrow lines connect each pair of factors for the same reason: Factors will all be uncorrelated (by means of a varimax rotation) or will all be allowed some correlation (by means of an oblique rotation). And again, each factor shows six

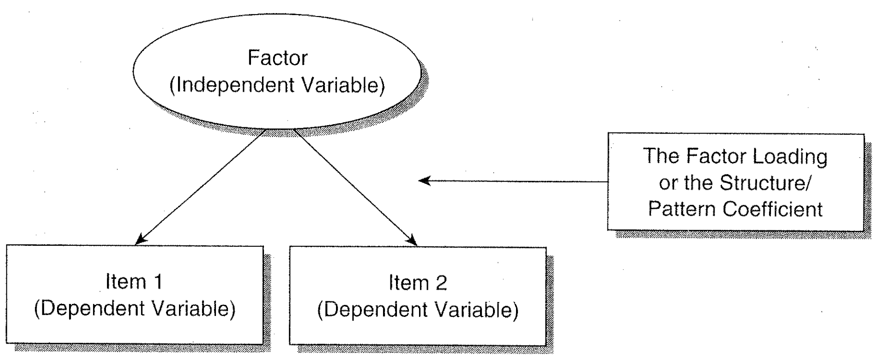


**Figure 13a.3** Diagram of a Principal Axis Factor

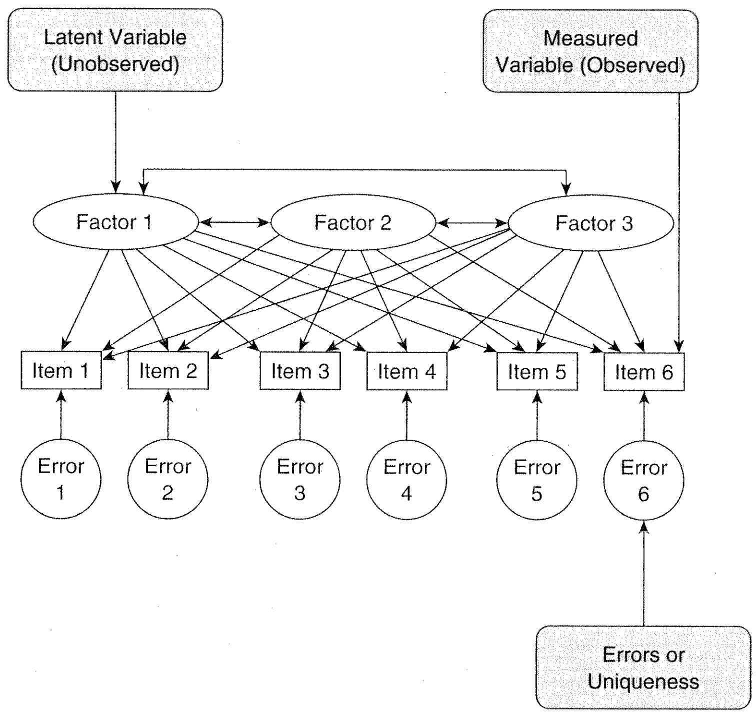


**Figure 13a.4** A Principal Axis Factor Model

paths leading from it—one to each indicator—because each indicator will receive a certain weight (ideally, most small and two higher) for each factor. The model also depicts the higher weights of certain variables with certain factors (e.g., Items 1 and 2 are more strongly correlated with Factor 1).

One major difference between the models for principal components and principle axis, besides the direction of the arrows and the reversal of independent and dependent variable roles, is the presence here of error elements associated with the indicators. Essentially, this means that principle axis factoring explicitly recognizes that not all the variance of the indicators will be associated with the factors. Recall from our discussion of the difference between principal components and factor analysis in Chapter 12A that principal components attempts to account for the total variance of the measured variables but that factor analysis attempts to account for only the variance common to the factors.

Variance that is not common or shared is said to be unique to each measured variable, and this unique variance is called *error, residual,* or *unique* *variance* in the structural model (Maruyama, 1998). The unique variance is usually some combination of systematic variance (e.g., another construct may actually be represented to a certain extent) and unsystematic (measurement error). Calling this whole term "error variance" just means that it is not part of the variance common to the set of variables in the analysis and therefore unsystematic with respect to that common variance; it does not mean that all that variance is, in any absolute sense, completely unsystematic.

The errors terms are enclosed in circles because they, too, are latent. The causal flow is from the error to the indicator, telling us what is probably obvious to most researchers: Our measurement of the indicators contains some variance not captured in the factor structure.

Note that there are no double arrow lines connecting the error variables. That is because these error terms are assumed to be uncorrelated in principle axis factoring. There is some real risk, however, that such an assumption may not hold in all situations. The assumption that the errors are uncorrelated is certainly one restriction or limitation we face when using exploratory factor analysis. In addition to this restriction is the requirement that the factors must be presumed either to all be uncorrelated or to all be correlated. Add to that the atheoretical nature of exploratory analysis in general and you have the seeds of concern that have led some writers (e.g., Long, 1983) to urge great caution in using exploratory procedures lest we wind up with a "garbage in–garbage out" model.

## The Structure of Confirmatory Factor Analysis

We display the more complete model for confirmatory factor analysis in Figure 13a.5. Two features of this model may be quickly apparent: (a) It is more similar to principle axis than to principal components, but (b) it is also simpler than the principle axis model because instead of having each measured variable relating to every factor, the researchers now clearly specify which measured variable "belongs" to which factor. Let's treat each in turn.

The similarity between principle axis factor analysis and confirmatory factor analysis occurs because—this may seem pretty obvious—they are both addressing factor analysis. Therefore, the causal flow is from the factors to the observed variables, and one thinks of these latter variables as indicators of the factors. Furthermore, there are error terms associated with each measured variable because factor analysis, whether in its exploratory or confirmatory form, partitions the variance of the indicators into common variance and unique variance. The analysis attempts to account for the common variance but treats the unique variance as residual or error variance.

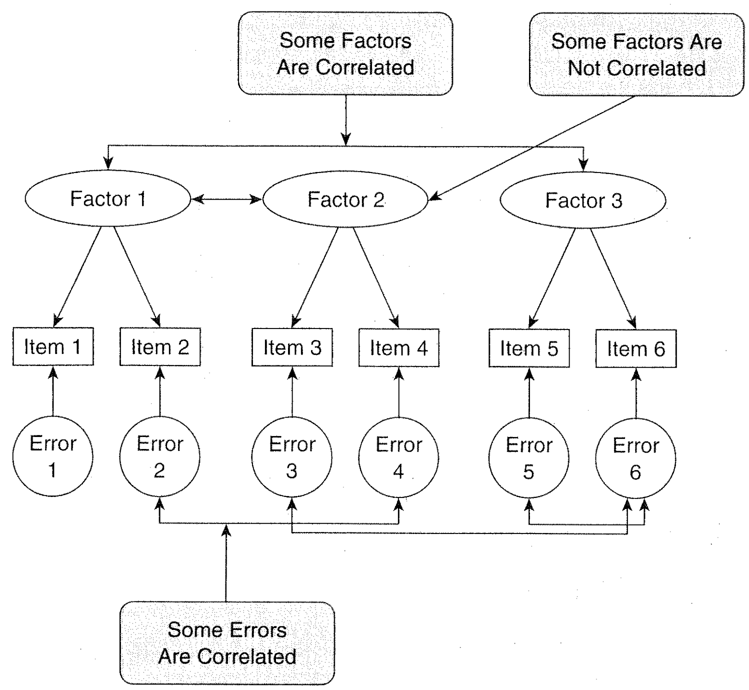


**Figure 13a.5** An Example of a Confirmatory Factor Analysis (CFA)

The differences between the principle axis and confirmatory structures can be seen in the positioning of the arrows connecting (a) the factors to the indicators, (b) the factors to the factors, and (c) the errors to the errors. The first of these may be the most striking because the structure looks a lot simpler here than in principle axis factoring. In principle axis factoring, each factor is connected to all the indicators. This is done because the factors are a weighted composite of all the measured variables. Each factor is represented by a different set of weights, but all the indicators are in the equation. For Factor 1, for instance, observed Variables 1 and 2 should have, according to what the model shows, much greater weights than observed Variables 3 through 6. However, for Factor 2, Indicators 3 and 4 would have the greatest weights.

Another difference between the principle axis and confirmatory structures is the connections between pairs of factors. Exploratory analysis decrees that what will be true for one pair will be true for all pairs. If we invoke a varimax rotation, we decree that all factors will be uncorrelated; if we invoke an oblique rotation, we decree that all factors will be correlated.

That restriction is lifted in confirmatory factor analysis. Researchers are able to specify the relationships they believe holds between any two factors. In Figure 13a.5, we have specified that Factor 1 is correlated with Factors 2 and 3 but that Factors 2 and 3 are uncorrelated.

In an analogous manner, researchers can also specify the associations between pairs of error terms. Because these error terms reflect unique variance, the possibility exists that this variance may be systematic and thus interpretable. It may challenge the theoretical structure from which these hypotheses are derived (or the creativity or insightfulness of the researchers), but if the theory or researchers are up to the task, it may be possible to hypothesize that some pairs of unique variance are related. In Figure 13a.5, we have specified that Errors 2 and 4, Errors 3 and 6, and Errors 5 and 6 are correlated.

## Confirmatory Analysis Is Theory Based

Exploratory factor analysis represents an inductive approach in that researchers employ a bottom-up strategy by developing a conclusion from specific observations. This conclusion is the interpretation of the factor based on those measured variables most strongly associated with it. These measured variables thus become indicators of the factor as a result of the statistical analysis. Although this strategy does appear to be empirically (inductively) based, research is not conducted in a vacuum, and researchers always have some ideas about the underlying structure of the variables that they have placed together in advance. For example, researchers interested in studying a particular construct will not select items haphazardly to write for their new inventory; rather, they will guide the content of the items to conform to what they know (or believe they know) about the construct. In this sense, no exploratory analysis is devoid of some researcher-based rationale (Gorsuch, 2003; Thompson, 2004).

Confirmatory factor analysis represents a deductive approach in that researchers employ a top-down approach by predicting an outcome from a theoretical framework. This outcome is the specification in advance of the statistical analysis of which measured variables are indicators of the factor. Further specification regarding the relationships between the factors and relationships between the error terms must also be forwarded by the researchers.

We have talked about the theoretical framework within which one engages in confirmatory analysis and the atheoretical environment where one must of necessity engage in exploratory analysis as though the two were easily distinguishable and compartmentalized. Statistical tests that analyze exploratory research are known as "first-generation" procedures, whereas statistical tests designed to confirm a theory are known as "second-generation" procedures (Fornell, 1987). Reality, however, has a way of complicating the decisions that researchers need to make regarding which procedure is to be used in a given situation.

Theory development in the social and behavioral sciences is an ongoing process, and very little research is conducted completely outside a theoretical context (Chin, 1998). The issue of contention is usually about the degree to which a theory has been articulated rather than whether or not there is one underlying the research. Sometimes it is not clear if an exploratory or confirmatory approach is more appropriate given that a particular theory may still be in the formative stage (Chin, Marcolin, & Newsted, 1996).

Furthermore, it is possible and often productive to make use of exploratory analysis as one of the theory-generating techniques open to researchers (Stevens, 2002). Thus, researchers having conducted a principal components analysis and interpreted the factors may be able to formulate some notions regarding the general construct that they were studying and begin to hypothesize what should occur under certain experimental or observational conditions. If they are able to fill in some of the details of their speculation, they may be ready to engage in confirmatory analysis the next time around.

## Overview of Confirmatory Factor Analysis

The major objective in confirmatory factor analysis is determining if the relationships between the variables in the hypothesized model resemble the relationships between the variables in the observed data set. More formally expressed, the analysis determines the extent to which the proposed covariance matches the observed covariance.

Once a model is proposed (i.e., relationships between the variables have been hypothesized), a correlation/covariance matrix is created. The development and evaluation of a confirmatory analysis typically involves five steps (Bollen & Long, 1993): (a) model specification, (b) model identification, (c) model estimation, (d) model evaluation, and (e) model respecification.

We will illustrate these steps using a scale developed by Gamst et al. (2004) in a study designed to assess the degree to which mental health practitioners demonstrated the ability to work with and understand clients from a diverse set of cultural backgrounds. They labeled the construct as Multicultural Competence and called their scale the California Brief Multicultural Competence Scale (CBMCS). These researchers also had identified four indicator variables of sensitivity to consumers, non-ethnic ability, awareness of cultural barriers, and multicultural knowledge. For the sake of keeping this example manageable, these indicators will be treated as though they are measured variables by placing them within rectangles in the structural model. As the study was conducted, the inventory was composed of 21 items, each measured on a 4-point summative response scale. These indicators, then, are really themselves latent variables (variates) computed by averaging the scores on a subset of inventory items.

We will apply confirmatory factor analysis by following the above five steps to assess the construct validity of the model. Construct validity refers to the degree to which a measure actually assesses the theoretical construct it is supposed to assess and is often assessed through confirmatory factor analysis. Evidence for construct validity is achieved if the model is a good fit of the data.

## Model Specification

The hypothesized model is diagrammed in Figure 13a.6. The researchers have specified their model (i.e., the four measured variables are hypothesized to be indicators of one factor). This model is an explanation of why these four variables relate. In the Bentler-Weeks method (Bentler & Weeks, 1980) of model specification, all variables (latent or measured) are assigned the role of either independent variables or dependent variables. Independent variables have arrows pointing away from the variable, whereas dependent variables have arrows pointing to the variable. In Figure 13a.6, there are five independent variables, the latent variable of multicultural competence and the four error or unique terms. The dependent variables are the four measured variables.

In this proposed model, the researchers are hypothesizing that multicultural competence causes sensitivity to consumers, non-ethnic ability, awareness of cultural barriers, and multicultural knowledge. Also, the researchers are specifying that there are other influences on the indicator variables besides multicultural competence. These other influences are represented as the error or unique-variance terms.

## Model Identification

To assess whether the proposed model fits the data, a necessary but not sufficient condition must be met: The model must be identified (Bollen, 1989). Model identification has to do with the difference between, at a very general level, the number of variables in the analysis, and the number of parameters that need to be estimated by the model. These parameters, or at

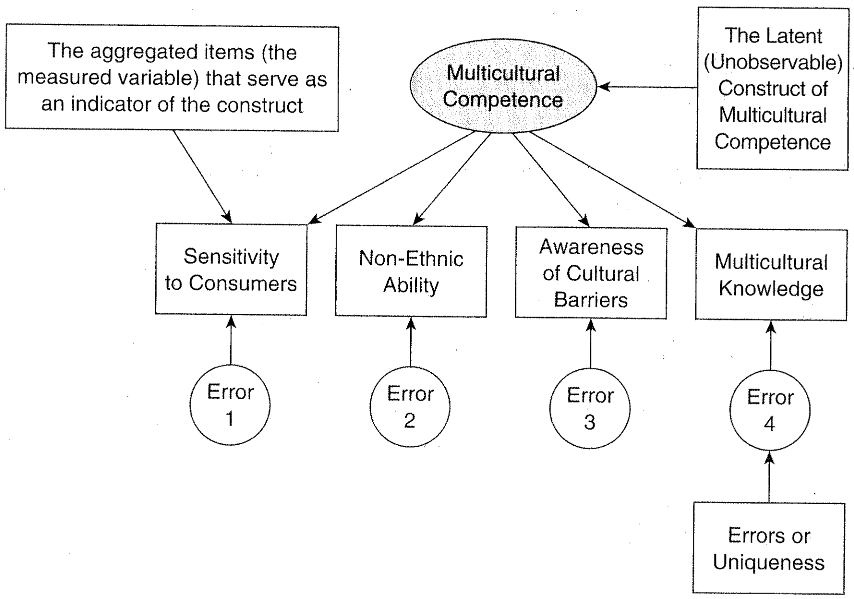


**Figure 13a.6** The Hypothesized Model for Multicultural Competence

least estimates of them, are what the structural equation model is designed to generate and are therefore unknown at the start of the analysis. The parameters in the model are the pattern or structure coefficients relating the independent to the dependent variables, correlation coefficients relating the independent variables to each other, and the variance of the independent variables. If we subtract the number of unknown parameters from the number of known or nonredundant elements, we obtain the degrees of freedom for the analysis:

$$df = \frac{\text{number of}}{\text{nonredundant elements}} - \frac{\text{number of}}{\text{unknown parameters}}$$

The situation that must be met in the model identification stage is to have more known information elements than unknown parameters (Bentler & Chou, 1987). If there are more unknown elements than known ones, the value for the degrees of freedom is negative. The model is then said to be *underidentified* and cannot be processed meaningfully in the analysis. If the numbers of known and unknown elements are equal, the model is said to be saturated or *just-defined.* It will perfectly but, unfortunately, artificially fit the data, and again no meaningful solution will be obtained by running the analysis. Only when the degrees of freedom are positive (we have more known than unknown elements) can a meaningful analysis be performed. We then say that the model is *identified* (technically, positive values for degrees of freedom indicate that the model is overidentified) and that it is therefore ready to be processed.

The task in this model identification stage is to count the number of known and unknown elements and, if the initial count is insufficient, to do whatever is necessary to make sure that we wind up with more of the former than the latter. So the obvious question that can be asked at this juncture concerns what kind of elements can be counted. The answer is (a) entries in a covariance or correlation matrix for the measured (dependent) variables and (b) variances and coefficients of various sorts (correlation coefficients, pattern coefficients, structure coefficients) for the paths of the independent variables.

The number of known elements is equal to the number of unique or nonredundant entries (allowing the pairing of the variable with itself) in a matrix that represents the covariances or correlations of the indicator variables. Consider a set of three variables labeled A, B, and C. How many unique entries are there in the matrix? There are six: AA, BB, CC, AB, AC, and BC. If you imagine a square matrix and count the entries in its upper half plus the diagonal, you can see the six positions. Anything in the lower half is a redundancy.

Rather than physically counting this every time we start a confirmatory analysis, we can apply a simple formula to obtain the count of nonredundant elements (Raykov & Marcoulides, 2000). The formula is as follows:

$$\text{Number of nonredundant elements} = V\,(V + 1)\,/\,2$$

where $V$ is the number of the measured variables in the study. If $V$ is equal to 3 in our above illustration, then we have $3 \times (3 + 1)$ or 12 divided by 2, which gives us 6.

We can now return to our multicultural competence example shown in Figure 13a.6. The means and standard deviations of the four measured variables are shown in Table 13a.1, and the correlations between the measured variables are shown in Table 13a.2.

Counting the nonredundant elements in this study is done as follows. There are four indicator variables. Plugging that value into our simple formula gives us $4 \times (4 + 1)$ divided by 2 or 10 pieces of nonredundant or known information. This value of 10 is immutable—we cannot increase this number. If there turn out to be more than 10 unknown parameters, thus yielding a negative value for degrees of freedom, we must take some action to reduce that count of unknowns because we cannot raise the number of nonredundant elements.

**Table 13a.1** Means and Standard Deviations of the Four Measured Variables

| | *N* | *M* | *SD* |
|---|---|---|---|
| Sensitivity to consumers | 80 | 53.97 | 6.910 |
| Non-ethnic ability | 80 | 52.22 | 7.323 |
| Awareness | 80 | 47.94 | 9.168 |
| Multicultural knowledge | 80 | 53.79 | 4.890 |

**Table 13a.2** Correlations Among the Four Measured Variables

| | *2* | *3* | *4* |
|---|---|---|---|
| 1. Sensitivity to consumers | .462* | .244 | .546* |
| 2. Non-ethnic ability | — | .400* | .525* |
| 3. Awareness | | — | .261 |
| 4. Multicultural knowledge | | | — |

*$p < .01$.

Model identification involves one other piece of business: The latent variable needs to be scaled. That is, to conduct the analysis, the latent variable must be assigned a metric of some kind. Now, we already pretty much know how we would be doing this if we had to—provided that the model is viable. We would compute a multicultural competence score in a way analogous to the way in which we computed the four indicator scores—namely, by averaging the scores on all the inventory items or averaging the scores on the four indicators themselves. Either way we did it, multicultural competence would then be assessed on the same 4-point scale that we used for the items and the indicators.

Now, in the empirical world of the researchers, if their model is workable, they would be working with the four factors and not with the more complex larger construct (that really is the whole point of factor analysis). But the mathematical world of statistics requires that the scale of the latent variable be specified in the model or the analysis is a no-go, so we very much want to specify the construct's metric.

Note that the indicators are already measured in the course of conducting the study and so have a measurement scale already tied to them. However, in structural equation modeling where we have specified, for example, factors, each of these (unobserved) latent variables must explicitly be assigned a metric within the context of the analysis. This is normally done by constraining one of the paths from the latent variable to one of its indicator (reference) variables, such as assigning the value of 1.0 to the pattern/structure coefficient for this path. Given this constraint, the remaining paths can then be estimated. The indicator selected to be constrained to 1.0 is called the *reference item.* Typically, we select as the reference item the one that in factor analysis loads most heavily on the dimension represented by the latent variable, thereby allowing it to anchor the meaning of that dimension. Alternatively, one may set the factor variances to 1, thereby effectively obtaining a standardized solution, or one may select the path associated with the indicator having the best reliability (because we are tying down the measurement of the construct).

Here, we will set the coefficient to a value of 1 (actually, any positive value will suffice) on the selected path. Such an act informs AMOS, the statistical software we use for structural equation modeling, that the construct is to be assessed on the same metric as that indicator. In Figure 13a.7, we have identified sensitivity to consumers as the indicator to which we will initially scale multicultural competence.

Here is why our action might seem a bit strange. One of the goals of the statistical analysis is to estimate parameters that were left unspecified by the model. The pattern or structure coefficients linking the construct and the indicators are among those parameters. But we have just filled in

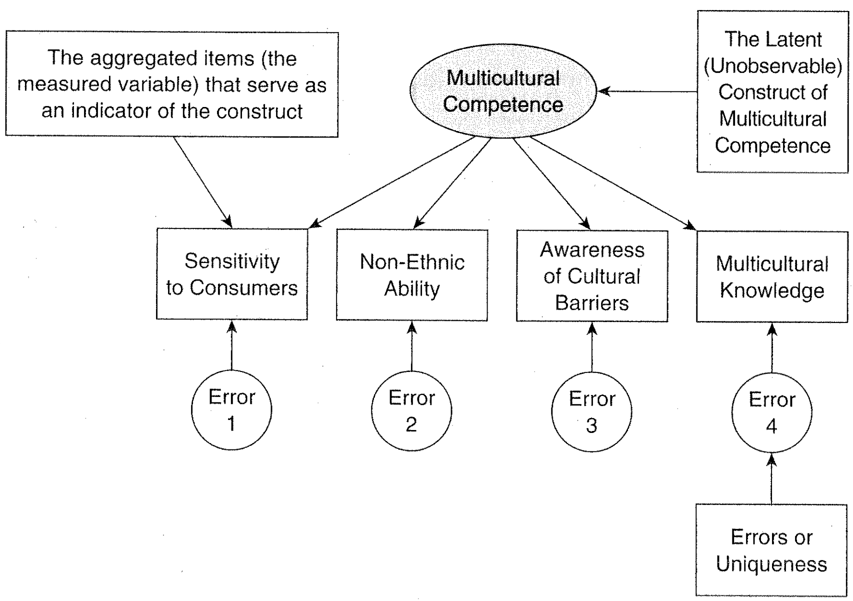


**Figure 13a.7** Scaling Sensitivity to Consumers by Constraining It to a Value of 1

one of them, which sounds as though it might defeat the purpose of performing the analysis. The bad news here is that we had to make this specification in order to run the analysis in the first place. But the good news is that structural equation modeling uses a maximum likelihood procedure that iterates its solution until it reaches a stable result. The upshot of this is that even though we have specified a value of 1 as the coefficient, the statistical procedure will use that only as a starting place and will eventually produce its own estimate. There is even more good news. Yes, we specified the coefficient leading to one of the indicators, but if we had chosen a different indicator, our solution would still be essentially the same.

Having done our preparation work, we now need to determine how many unknowns there are. These unknown elements will be the parameters estimated by the statistical procedure. In this model, we will need to estimate (a) the variance of the multicultural competence latent factor, (b) the variances of the four unique error terms, (c) the remaining three pattern/structure coefficients relating the latent variable to the measured variables whose coefficients are not yet specified (we have already removed from the realm of the unknown one parameter by scaling the latent factor), and (d) the four pattern/structure coefficients relating the unique variables (the error terms) to the measured variables. This gives us a total of 13 parameters (1 + 4 + 3 + 4) that we are asking the model to estimate.

A quick calculation to figure out if this model is identified tells us that we are (temporarily only) in a bit of trouble. Subtracting the unknowns (13) from the known elements (10) yields –3 degrees of freedom. Thus, the model is not yet identified, and we are not ready to go forward into the analysis.

The portions of the model most susceptible to specification are the paths leading from the error terms to the indicator variables. What we do here is assign them an initial value of 1 as shown in Figure 13a.8, recognizing that they will be estimated in the solution. This removes these four parameters from the list of unknowns. It also serves the added advantage of scaling these latent error terms to their respective measured variables. Because the error variables represent unique variance, the model will actually have an easy job filling in these values. Once we learn of the correlation between the indicators and the latent factor (the common variance), these values may be calculated as the residual or what remains of the total variance once the common variance is accounted for.

Changing the status of these parameters from unknown to known in the model now puts us in the positive range for degrees of freedom. The known elements still count to 10. The unknown parameters after all our work are (a) the variance of the latent factor, (b) the variances of the four latent error variables, and (c) the remaining three pattern/structure coefficients. That

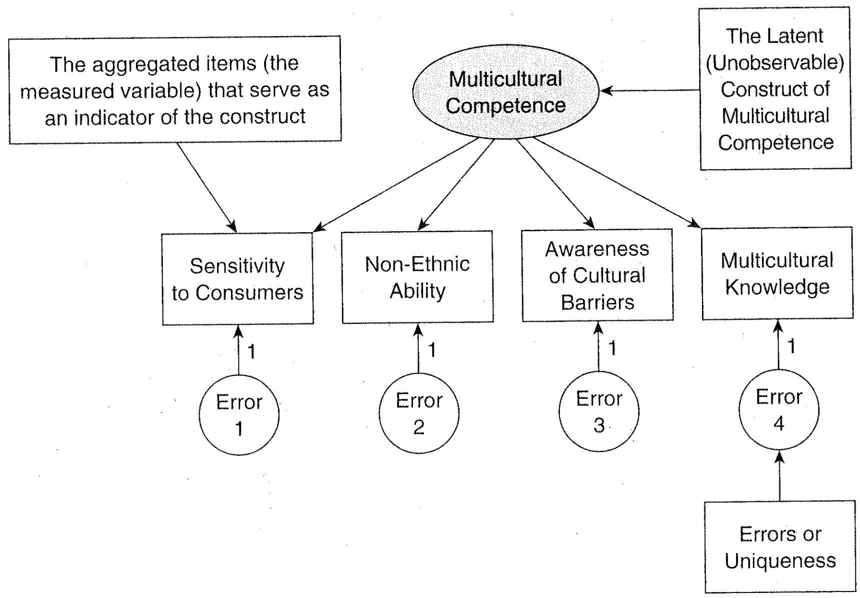


**Figure 13a.8** Constraining the Error Variables

gives us a total of 8. By subtracting the number of unknown parameters from the number of known elements (10 − 8 = 2), we now have positive degrees of freedom. This model is *overidentified* (which is what we wanted to happen) and can thus be meaningfully solved.

## Model Estimation

Once we have completed the identification stage, we are then ready to conduct the statistical analysis. This involves mathematically building the model and estimating the relationships between the variables in the model. We calculate these estimates using the maximum likelihood estimation (MLE) procedure, one of several methods available to researchers through the AMOS program. Maximum likelihood attempts to estimate the values of the parameters that would result in the highest likelihood of the actual data matching with the proposed model. These methods always require iterative solutions.

We will present the results of the data analysis shortly because to really justify examining and interpreting the outcome of the analysis, it is a good idea to determine how well the model is able to fit the data. As part of the procedure of running a confirmatory factor analysis, we will want to evaluate the model by evaluating these fit indexes.

## Model Evaluation

The proposed or hypothesized model is assessed by producing estimates of the unknown parameters. These findings are then compared with the relationships (the correlation/covariance matrices) existent in the actual or observed data. Confirmatory factor analysis assesses how well the predicted interrelationships between the variables match the interrelationships between the actual or observed interrelationships. If the two matrices (the proposed and the actual) are consistent with one another, then the model can be considered a credible explanation for the hypothesized relationships.

Stevens (2002) divides the assessment of a model into two categories: "those that measure the overall fit of the model and those that are concerned with individual model parameters" (pp. 402–403). All the fit measures are analogous to the omnibus test in ANOVA in that they provide an overall assessment of the model. As in the ANOVA, post hoc tests need to be conducted to provide further interpretation of the analysis. Similarly, in confirmatory factor analysis, the overall fit of a model to the data may appear acceptable, yet some relationships in the model may not be supported by the data (Bollen, 1990).

The chi-square statistic is used to test the difference between the predicted and the observed relationships (correlations/covariances). Because the researcher is predicting a close fit, a nonsignificant chi square is desired. In the multicultural competence example, the chi-square value was 3.44. At two degrees of freedom, the probability of that value occurring by chance was .179 (it was greater than our default alpha value of .05 and is therefore not statistically significant), indicating an acceptable match between the proposed model and the observed data.

Jöreskog and Sörbom (1989) and Bentler (1990), however, advised against the sole use of the chi-square value in judging the overall fit of the model because of the sensitivity of the chi square to sample size. As sample size increases, power increases. Therefore, the chi-square test can detect small discrepancies between the observed and predicted covariances and suggest that the model does not fit the data. A good-fitting model could be rejected because of trivial but statistically significant differences between the observed and predicted values. Because of these limitations, many other fit indexes were developed as alternatives or supplements to chi square.

### Assessing Fit of Hypothesized Models

How to assess the fit between the hypothesized model and the observed data continues to develop. Over the past 20 years, at least 24 fit indexes have been proposed (Klem, 2000). All of these fit indexes were developed to diminish the Type II error (i.e., concluding that the data do not support the proposed model when in fact they do). For the 24 fit measures available through statistical software programs, there is presently no general agreement on which measures are preferred (Maruyama, 1998). As Hair Anderson, Tatham, and Black (1998) state, "SEM [structural equation modeling] has no single statistical test that best describes the 'strength' of the model's predictions" (p. 653). None of the measures has a related statistical test, except for the chi-square test. The confusing consequence of these competing fit indexes is that different research studies report different fit indexes.

Complicating this matter of competing fit indexes is the lack of consensus among structural equation modeling writers as to how we may categorize or organize this array of fit indexes, although most agree that some organizational schema is necessary (because then we might be able to report fit indexes from each category). Further complicating the decision of what fit index to report, there is disagreement about which individual fit measures might best be classified together.

Researchers have proposed various classification schemas to organize the fit indexes. For example, Arbuckle (1999) devised an eight-category scheme (parsimony, sample discrepancy, population discrepancy, information-theoretic, baseline model, parsimony adjusted, goodness of fit, and miscellaneous), whereas Tabachnick and Fidell (2001b) suggest a five-category system (comparative, absolute, proportion of variance, parsimony, and residual based). Other authors (e.g., Hair et al., 1998; Jaccard & Wan, 1996) promote a three-group scheme (absolute, relative, and parsimonious). And finally, Maruyama (1998) adopts Hu and Bentler's (1999a) two-type scheme (absolute and relative but with the latter divided into four subtypes). The most cited organization system appears to be the three-classification scheme (absolute, relative, and parsimonious). Brief descriptions of the three classification schemes with their respective fit measures are presented below and summarized in Table 13a.3.

## Absolute Fit Measures

Absolute fit measures indicate how well the proposed interrelationships between the variables match the interrelationships between the actual or observed interrelationships. This means how well the correlation/covariance of the hypothesized model fits the correlation/covariance of the actual or observed data. The four most common absolute fit measures assessing this general feature are the chi square, the goodness-of-fit index (GFI), the root mean square residual (RMSR), and the root mean square error of approximation (RMSEA). Because the researcher is predicting a close fit, a nonsignificant chi square is preferred. As we indicated, the chi-square test is too

**Table 13a.3** Absolute, Relative, and Parsimonious Fit Measures

| *Fit Measures* | | | | | |
|---|---|---|---|---|---|
| *Absolute* | | *Relative* | | *Parsimonious* | |
| *Test* | *Value* | *Test* | *Value* | *Test* | *Value* |
| $\chi^2$ | p >.05 | CFI | >.95 | PNFI | >.50 |
| GFI | >.90 | NFI> | >.90 | PCFI > | >.50 |
| RMSR | <.05 | IFI | >.90 | | |
| RMSEA | <.10 | RFI | >.90 | | |

Note: $\chi^2$ = Chi-square test; GFI = Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; 2/df = Chi square divided by degrees of freedom test; PGFI = Parsimony Goodness of Fit Index, AGFI = Adjusted Goodness of Fit Index; CFI = Comparative Fit Index; NFI = Normed Fit Index; IFI = Incremental Fit Index; RFI = Relative Fit Index; PNFI = Parsimony Normed Fit Index; PCFI = Parsimony Comparative Fit Index.

powerful. As sample size increases, power increases, and the chi-square test can return a statistically significant outcome even when the model fits the data reasonably well.

The GFI is conceptually similar to the $R^2$ in multiple regression (Kline, 1998). It is the proportion of variance in the sample correlation/covariance accounted for by the predicted model, with values ranging from 0 (no fit) to 1 (a perfect fit). Although the GFI can vary from 0 to 1, theoretically, it can yield meaningless negative values. By convention, GFI should be equal to or greater than .90 as indicative of an acceptable model.

The RMSR is a measure of the average size of the residuals between actual covariance and the proposed model covariance. The smaller the RMSR, the better the fit (e.g., < .05). The RMSEA is the average of the residuals between the observed correlation/covariance from the sample and the expected model estimated for the population. Byrne (1998) states that the RMSEA "has only recently been recognized as one of the most informative criteria in covariance structure modeling" (p. 112). Values less than .08 are deemed acceptable, whereas values greater than .10 are generally unacceptable.

## Relative Fit Measures

Relative fit measures are also known as comparisons with baseline measures or incremental fit measures. These are measures of fit relative to the independence model, which assumes that there are no relationships in the data (thus a poor fit) and the saturated model, which assumes a perfect fit. The incremental fit measures indicate the relative position on this continuum between worst fit to perfect fit, with values greater than .90 suggesting an acceptable fit between the model and the data. Incremental fit measures are also referred to as comparisons with baseline measures or relative fit measures.

Byrne (1998) suggests that the comparative fit index (CFI) should be the fit statistic of choice in structural equation modeling research. Knight, Virdin, Ocampo, and Roosa (1994) have suggested the following guidelines to evaluate the CFI: good fit > .90; adequate but marginal fit = .80 to .89; poor fit = .60 to .79; very poor fit < .60. Hu and Bentler (1999a, 1999b) revised the value representing a good fit to .95. Other common incremental fit measures are the normed fit index (NFI), non-normed fit index (NNFI), incremental fit index (IFI), and the relative fit index (RFI), with values between .90 and .95 indicating an acceptable fit.

### Parsimonious Fit Measures

Parsimonious fit measures are sometimes called adjusted fit measures. These fit statistics are similar to the adjusted $R^2$ in multiple regression analysis; the parsimonious fit statistics penalize larger models with more estimated parameters. Here is why: Recall that MLE maximizes the likelihood that the data will support the proposed model. The more paths a researcher is estimating, the more likely the fit will be acceptable no matter how nonsensical the model may be. Researchers could "stack the deck" in their favor (unintentionally, of course) by increasing the complexity of their model. However, one of the goals of research is to develop more parsimonious models. These parsimonious fit measures can be used to compare models with differing number of parameters to determine the impact of adding additional parameters to the model.

Common parsimonious fit measures are the parsimonious adjusted goodness of fit (AGFI) and the parsimonious goodness of fit (PGFI). The AGFI corresponds to the GFI in replacing the total sum of squares by the mean sum of squares. PGFI adjusts for degrees of freedom in the baseline model. It is a variant of GFI that penalizes GFI by multiplying it by the ratio formed by the degrees of freedom in the tested model and degrees of freedom in the independence model. Ideally, values greater than .90 indicate an acceptable model; however, typically, parsimony-based measures have lower acceptable values (e.g., .50 or greater is deemed acceptable; Mulaik et al., 1989).

### Where We Stand Now

The goal of researchers who use structural equation modeling techniques is to evaluate the plausibility of their hypothesized model (i.e., the relationships between the variables). The proposed model is either guided through prior research or based on some theoretical framework. The model is then compared with the actual or observed data. If the model and observed data resemble each other, then the model is said to fit the data.

Historically, chi square was used to assess the fit of the model, with the expectation that the $p$ value would be nonsignificant. Because a chi square assesses the difference between expected values and observed results, a nonsignificant value would indicate a close fit. For models with relatively small cases (75 to 200), chi square is an adequate measure of fit (Kenny, 2003). However, for models with more cases, chi square will usually be statistically significant, suggesting a poor fit when there really are only trivial differences between the model and the data. Chi square is also affected by the size of the correlations in the model; larger correlations generally cause a poorer fit (Kenny, 2003). Because of these problems with chi square, alternative measures of fit are constantly being developed and studied.

Over the past 20 years, at least 24 fit indexes have been proposed (Klem, 2000). All these fit indexes were developed to diminish the chances of committing a Type II error (i.e., concluding that the data does not support the proposed model when in fact it does). One of the problems with chi-square tests was that with large samples, the results would suggest rejecting a model when, in fact, the model supported the data. Some of the alternative fit measures, however, have suffered the same fate as some technologies in the music industry: What was once considered "leading edge" is now considered inadequate and obsolete. What emerged with the development of alternative fit measures is the realization that no single fit index is powerful enough to assess the adequacy of the model. According to Maruyama (1998), "The different fit indexes differ with respect to dimensions such as susceptibility to sample size differences, variability in the range of fit possible for any particular data set, and valuing simplicity of model specification needed to attain an improved fit" (p. 239). Until recently, most texts on structural equation modeling presented a three-classification scheme: absolute, relative, and parsimonious.

Absolute fit measures judge how well the proposed interrelationships between the variables match the interrelationships between the actual or observed interrelationships. This means how well the correlation/covariance of the hypothesized model fits the correlation/covariance of the actual or observed data.

Relative fit measures are also known as comparisons to baseline measures or incremental fit measures. These are measures of fit relative to the independence model, which assumes that there are no relationships in the data (thus a poor fit) and the saturated model, which assumes a perfect fit.

The incremental fit measures indicate the relative position on this continuum between worst fit to perfect fit, with values greater than .90 suggesting an acceptable fit between the model and the data.

Both the absolute and the relative measures will report better fit measures if the models being evaluated have more parameter estimations. This is one of the by-products of the maximum likelihood method employed to estimate the parameters. Recall that the maximum likelihood method is an iterative process that compares the parameter estimates at each phase of the estimation procedure and reports the estimates that provide the best fit. With more parameters available to estimate, the greater the likelihood of developing a better fit.

To adjust for this inflated fit bias, the third classification, known as parsimonious fit measures, was developed. These fit measures adjust for the number of estimations by penalizing models with greater parameter estimations. Kelloway (1998) warns that unlike the absolute or relative fit measures that have conventional values of .90 or .95 as acceptable models, parsimonious fit measure have no generally acceptable cutoff. Parsimonious fit measures are recommended to compare two competing models, with the model with the higher fit measure as superior.

### Which Fit Measures to Report?

There is disagreement among structural equation modeling researchers on just which fit indexes to report. One often-cited recommendation is from Jaccard and Wan (1996) who suggests reporting at least three fit tests—one absolute, one relative, and one parsimonious—to reflect diverse criteria. Recently, with the advances of more realistic simulation studies, Kline (2005) and Thompson (2004) recommend fit measures without reference to their classification. We suggest reporting chi square, the NFK, the CFI, and the RMSEA as fit measures. Although chi square is less informative as an assessment of a single model, it is useful in comparing nested models (i.e., where one model is a subset of another model). The model with the lower chi-square value is considered to be the preferable model. Both the NFI and the CFI should achieve a value of .95 for the model to be deemed acceptable. The RMSEA is the average of the residuals between the observed correlation/covariance from the sample and the expected model estimated from the population. Byrne (1994) stated, "[RMSEA] has only recently been recognized as one of the most informative criteria in covariance structure modeling" (p. 112). Loehlin (2004) proposes the following criteria for evaluating this index: (a) less than .08 indicates good fit, (b) .08 to .1 indicates a moderate fit, (c) greater than .1 indicates poor fit.

## Model Estimation: Assessing Pattern/Structure Coefficients

The next step in assessing the model is to see if the factor pattern/structure coefficients are statistically significant and meaningful. Figure 13a.9 presents the pattern/structure coefficients of the four measured variables (indicators) to the factor Multicultural Competence.

All the factor pattern/structure coefficients achieved statistical significance at an a priori alpha level of $p < .05$. The statistically significant pattern/structure coefficients indicate that Multicultural Competence is composed of the four measured variables named in the model. The pattern/structure coefficients also achieved meaningful (practical) significance with coefficients greater than .3, although the inference strength from the latent factor to awareness of cultural barriers is a bit more tenuous than the inferences to the other three indicators. Have the researchers proven their model? The answer is no. Confirmatory factor analysis is not an arena for proof, but it can be a source of support for a model and can supply enough information to researchers to cause them to reject a model if it does not fit the data very well. Even in this latter case, researchers may be motivated to modify their existing model rather than completely discarding it if they can determine what might need to be remedied so that the model may be improved.

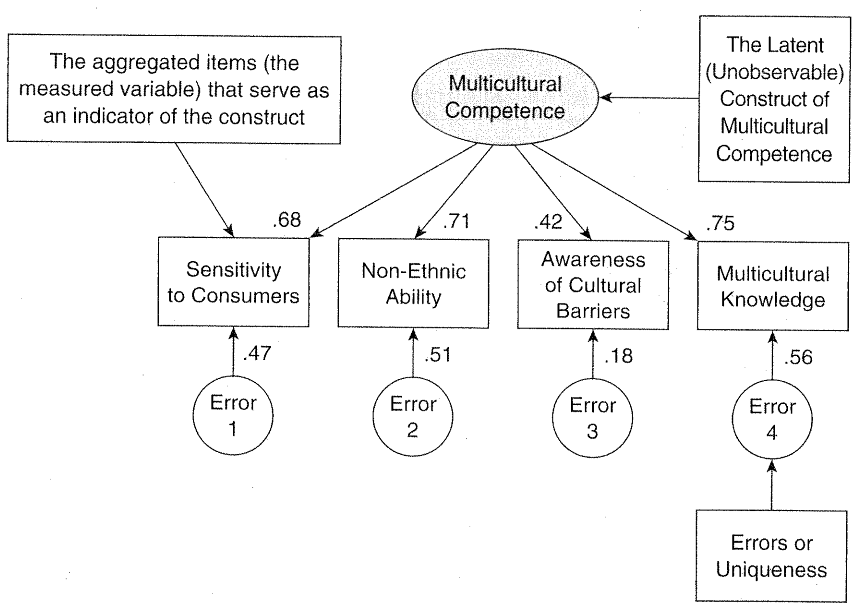


**Figure 13a.9** The Model With the Estimated Coefficients

## Model Respecification

It is not uncommon that the initial proposed model fails to achieve an adequate fit and that the researchers consider modifying the model (i.e., respecifying). In respecifying the model, the researchers may delete the nonsignificant coefficients in an attempt to trim their models. Researchers could also add coefficients between factors and indicator variables that were previously ignored as an approach to develop their models. In either situation, the analyses are now exploratory rather than confirmatory. Byrne (2001) explains:

> In other words, once an hypothesized CFA model, for example, has been rejected, this spells the end of the confirmatory factor-analytic approach, in its truest sense. Although CFA procedures continue to be used in any respecification and reestimation of the model, these analyses are exploratory in the sense that they focus on the detection of misfitting parameters in the originally hypothesized model. (p. 91)

The researchers need to consider if the respecification is theoretically justifiable. If deleting or adding a coefficient lacks any theoretical justification, the researchers need to avoid this temptation in an attempt to improve the model fit. Steiger (1990) warns researchers of their ability to justify new parameters when he stated, "What percentage of researchers would find themselves unable to think up a 'theoretical justification' for freeing a parameter? . . . I assume that the answer . . . is near zero" (p. 175).

If a model is respecified and achieves acceptable fit with the data, this new model needs to be retested on an independent sample. This new independent sample can either be a holdout sample from the original study (provided that the sample is large enough, perhaps 400 or more) or a new sample.

To illustrate an example of a respecified model, the following heuristic will be used. Researchers are assessing the construct validity of a new instrument used to determine aptitude for graduate school. They have proposed two latent factors, Verbal Ability and Math Ability. Each factor has three measured (indicator) variables. The verbal factor is composed of scores from three subscales: spelling, grammar, and comprehension. The math factor is composed of scores from three subscales: word problems, calculations, and conceptual understanding. The model is presented in Figure 13a.10.

The results of this hypothetical study are shown in Figure 13a.11. We display the pattern coefficients and the correlation between the two latent variables. As can be seen, the pattern coefficients are not especially high, although they are statistically significant and certainly meaningful (i.e., they do exceed .3).

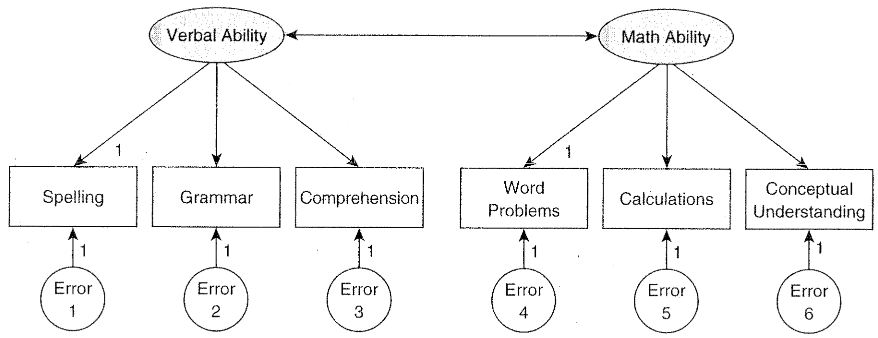


**Figure 13a.10** Proposed Two-Factor Model With Six Indicator Variables

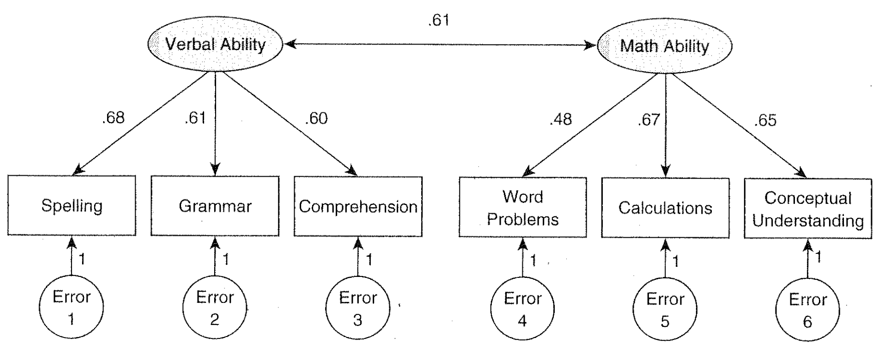


**Figure 13a.11** Pattern Coefficients of the Six Indicator Variables for the Two Factors

The results of the overall 'model, however, indicated a moderate fit between the proposed model and the observed data. The chi square was statistically significant, indicating a lack of fit. Although the absolute fit measure of the GFI indicated a good fit (.975), the RMSEA indicated a marginal fit with .085. The incremental measure, the CFI, was also marginal (.944) because it did not quite meet the .95 criterion.

Perhaps a respecified model may better account for the observed data. As will be illustrated in Chapter 13B, confirmatory factor analysis output will provide modification indexes. These modification indexes are suggestions to improve the fit of a model. Recall that the researchers need to consider if the respecification is theoretically justifiable. Because all the coefficients achieved both statistical and practical significance, deleting any of the coefficients would be counterproductive.

However, it is possible to suggest additions to the model that may enhance its fit to the data. One plausible addition is adding a path (coefficient) from Verbal Ability to Word Problems. This path makes theoretical sense because verbal ability is necessary to comprehend math word problems. Another modification suggestion is that Math Ability could be assessed by the verbal indicator variable Comprehension. Cognitive psychologists can make the case that proficient mathematical ability aids in verbal comprehension. Thus, these "cross-loadings" make theoretical sense. Such a revised model is drawn in Figure 13a.12.

The results of the new model are shown in Figure 13a.13. Note that the correlation between Verbal Ability and Math Ability has decreased from .61 to .44. This drop occurred because some of the variance that was "exclusive" to Verbal Ability has been "released" to the math side of the model; that is, Verbal Ability now has a path to Word Problems. At the same time, some of the variance that was exclusive to Math Ability is now reassigned to the verbal side of the model; that is, Math Ability now has a path to Comprehension.

In terms of the pattern/structure coefficients, we see some gains here as well. The paths from the original model now show values that are a bit higher. For example, Spelling now "loads" on Verbal Ability at .71 compared with its prior value of .68, and Calculations now has a coefficient of .69 compared with its prior value of .67. The new coefficients, the ones that are cross loading, although not large in an absolute sense, do add explanatory power. Thus, by permitting word problems to be an indicator of verbal ability in addition to indicating math ability, and by permitting comprehension to be an indicator

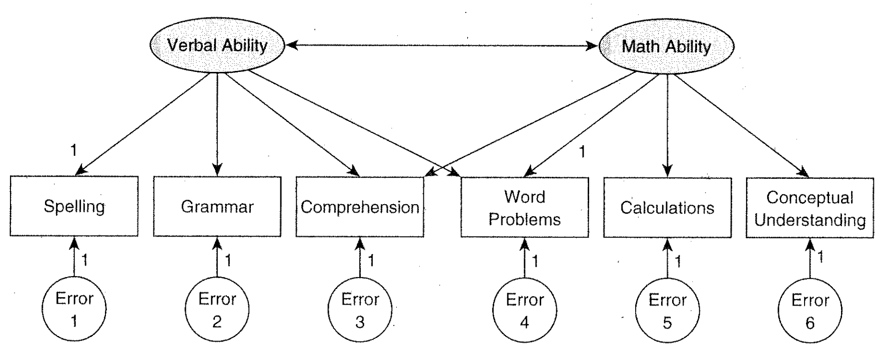


**Figure 13a.12** A Respecified Model

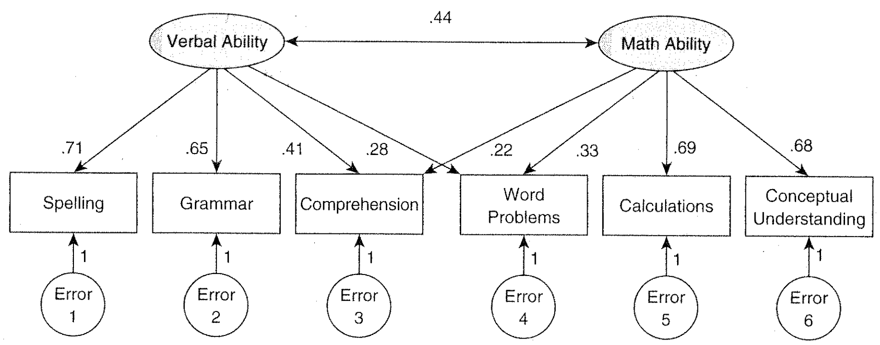


**Figure 13a.13** The Results of the Respecified Model

of math ability in addition to indicating verbal ability, we have apparently reached a better understanding of the dynamics underlying these variables.

In terms of the statistical results, all the coefficients achieved statistical significance ($p < .05$). The chi-square value of .334 is nonsignificant, indicating a fit between the model and data. Improvements in the other fit indexes were also obtained. The absolute fit measure of the GFI increased to .996 and the RMSEA attenuated to .017. The incremental measure, CFI, increased to .998.

## General Considerations

Structural equation modeling is able to accommodate some departures from normality. The AMOS program also performs state-of-the-art estimation of missing data by full information maximum likelihood instead of relying on ad hoc methods such as listwise or pairwise deletion or mean imputation.

Sample size is also something that researchers should take into consideration when performing confirmatory factor analysis or structural equation modeling in general. In the literature, sample sizes for these sorts of studies commonly run in the 200 to 400 range for models with 10 to 15 indicators. Loehlin (1992) recommends at least 100 cases and preferably 200. With more than 10 variables, sample sizes under 200 generally cause parameter estimates to be unstable and the tests of statistical significance tests lack a bit of power. One rule of thumb is that sample size should be at least 50 more than 8 times the number of variables in the model. Another rule of thumb, based on Stevens (2002), is to have at least 15 cases per measured variable or indicator. Bentler and Chou (1987) recommend at least 5 cases per parameter estimate (including error terms as well as path coefficients).

## Recommended Readings

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6,* 56–83.

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

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Muliak, S. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Muliak, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwill, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430–445.

Reilly, T. (1995). A necessary and sufficient condition for identification of confirmatory factor analysis models of complexity one. *Sociological Methods & Research, 23,* 421–441.

Steiger, J. H. (1998). A note on multisample extensions of the RMSEA fit index. *Structural Equation Modeling, 5,* 411–419.

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261–283). Washington, DC: American Psychological Association.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.