2

Understanding the Latent Variable

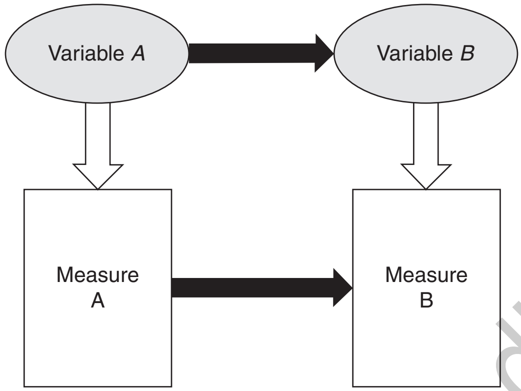
This chapter presents a conceptual schema for understanding the relationship between measures and the constructs they represent, though it is not the only framework available. Item response theory is an alternative measurement perspective that we will examine in Chapter 8. Because of its relative conceptual and computational accessibility and wide usage, we emphasize the classical measurement model, which assumes that individual items are comparable indicators of the underlying construct.

Constructs Versus Measures

Typically, researchers are interested in constructs rather than items or scales per se. For example, a market researcher measuring parents' aspirations for their children would be more interested in intangible parental sentiments and hopes about what their children will accomplish than in where those parents place marks on a questionnaire. However, recording responses to a questionnaire may, in many cases, be the best method of assessing those sentiments and hopes. Scale items are usually a means to the end of construct assessment. In other words, they are necessary because many constructs cannot be assessed directly. In a sense, measures are proxies for variables that we cannot directly observe. By assessing the relationships between measures, we indirectly infer the relationships between constructs. In Figure 2.1, for example, although our primary interest is the relationship between Variables *A* and *B*, we estimate that relationship on the basis of the connection between measures corresponding to those variables.

The underlying phenomenon or construct that a scale is intended to reflect is often called the *latent variable*. As we use the terms in this text, all scales (and

FIGURE 2.1 ● Relationships between instruments correspond with relationships between latent variables only when each measure corresponds to its latent variable



some indices) involve a latent variable. In this chapter, unless otherwise noted, our discussion is limited to scale items. Exactly what is a latent variable? Its name reveals two chief features. Consider the example of parents' aspirations for children's achievement. First, it is *latent* rather than manifest. Parents' aspirations for their children's achievement are not directly observable. In addition, the construct is *variable* rather than constant—that is, some aspect of it, such as its strength or magnitude, changes. Parents' aspirations for their children's achievement may vary according to time (e.g., during the child's infancy versus adolescence), place (e.g., on an athletic field versus a classroom), people (e.g., parents whose own backgrounds or careers differ), or any combination of these and other dimensions. The latent variable is the actual phenomenon that is of interest—in this case, child achievement aspirations.

Another noteworthy aspect of the latent variable in the case of a scale is that it is typically a characteristic of the individual who is the source of data. Thus, in our present example, parental aspirations are a characteristic of the parents and not of the children. Accordingly, we assess it by collecting data about the parents' beliefs from the parents themselves. While there may be circumstances in which some form of proxy reporting (e.g., asking parents to report some characteristic of their children) is appropriate, in general, we will ask respondents to self-report information pertaining to themselves. When this is not the case, as in a study involving parents describing the aspirations their children have for themselves, care must be taken in interpreting the resulting information. Arguably, in this hypothetical instance, the latent variable might more accurately be described as *parents' perceptions of their children's aspirations* than as *children's aspirations* per se. Likewise, if we ask a group of shoppers to evaluate characteristics of a particular store, we are assessing *shoppers' perceptions* rather than aspects of the store itself (which might be more easily assessed by direct observation). How important the distinction is between assessing the perceptions of a respondent with regard to some external stimulus (e.g., perceptions of the store), as opposed to characteristics of the external stimulus (e.g., the store itself), will depend on the specific circumstances and goals of the assessment; however, in all cases, it is important to be mindful of the distinction and to make appropriate interpretations of the resultant data.

Although we cannot observe or quantify it directly, the latent variable presumably takes on a specific value under some specified set of conditions. A scale developed to measure a latent variable is intended to estimate its actual magnitude at the time and place of measurement for each thing measured. This unobservable actual magnitude is the *true score*.

Latent Variable as the Presumed Cause of Scale Item Values

The notion of a latent variable implies a certain relationship between it and the items that tap it. The latent variable is regarded as a *cause* of the scale item score—that is, the strength or quantity of the latent variable (i.e., the value of its true score) is presumed to cause an item (or set of items) to take on a certain value.

An example may reinforce this point: The following are hypothetical items for assessing parents' aspirations for children's achievement:

1. My child's achievements determine my own success.
2. I will do almost anything to ensure my child's success.
3. No sacrifice is too great if it helps my child achieve success.
4. My child's accomplishments are more important to me than just about anything else I can think of.

If parents were given an opportunity to express how strongly they agree with each of these items, their underlying aspirations for childhood achievement should influence their responses. In other words, each item should give an indication of how strong the latent variable (aspirations for children's achievement) is. The score obtained on the item is caused by the strength or quantity of the latent variable for that person at that particular time.

A causal relationship between a latent variable and a measure implies certain empirical relationships. For example, if an item value is caused by a latent variable, then there should be a correlation between that value and the true score of the latent variable. As a consequence of each of the indicators correlating with the latent variable, they should also correlate with each other. Because we cannot directly assess the true score, we cannot compute a correlation between it and the item. However, when we examine a set of items that are presumably caused by the same latent variable, we can examine their relationships to one another. So if we had several items like the ones preceding measuring parental aspirations for child achievement, we could look directly at how they correlated with one another, invoke the latent variable as the basis for the correlations among items, and use that information to infer how highly each item was correlated with the latent variable. Shortly, we will explain how all this can be learned from correlations among items. First, however, we will introduce some diagrammatic procedures to help make this explanation more clear.

Path Diagrams

Coverage of this topic will be limited to a brief review of issues pertinent to scale development. For greater depth, consult Asher (1983) or Loehlin (1998).

Diagrammatic Conventions

Path diagrams are a method for depicting *causal* relationships among variables. Although they can be used in conjunction with path analysis, which is a data analytic method, path diagrams have more general utility as a means of specifying how a set of variables are interrelated. These diagrams adhere to certain conventions. A *straight arrow* drawn from one variable label to another indicates that the two are *causally related* and that the direction of causality is as indicated by the arrow. Thus $X \rightarrow Y$ indicates explicitly that X is the cause of Y. Often, associational paths are identified by labels, such as the letter a in Figure 2.2.

The *absence* of an arrow also has an explicit meaning—namely, that two variables are *unrelated*. Thus, $A \rightarrow B \rightarrow C\ D \rightarrow E$ specifies that A causes B, B causes C, C and D are *unrelated,* and D causes E.

Another convention of path diagrams is the method of representing *error,* which is usually depicted as an additional causal variable. This error term is a *residual,* representing all sources of variation not accounted for by other causes explicitly depicted in the diagram.

Because this error term is a residual, it represents the discrepancy between the actual value of Y and what we would predict Y to be based on knowledge of X and Z (in this case; see Figure 2.3). Sometimes, the error term is assumed and, thus, not included in the diagram.

Path Diagrams in Scale Development

Path diagrams can help us see how scale items are causally related to a latent variable. They can also help us understand how certain relationships among items imply certain relationships between items and the latent variable. We

FIGURE 2.2 ● The causal pathway from *X* to *Y*

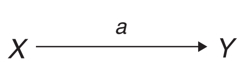


FIGURE 2.3 ● Two variables plus error determine *Y*

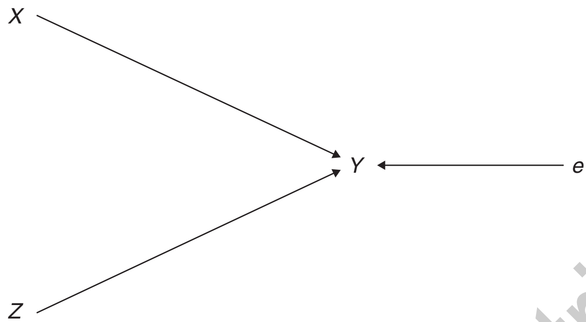
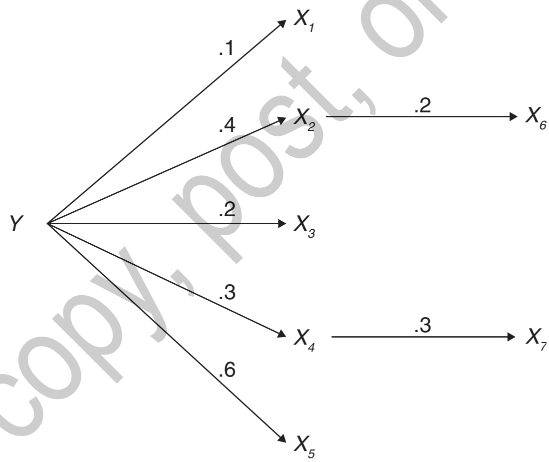


FIGURE 2.4 ● A path diagram with path coefficients, which can be used to compute correlations between variables



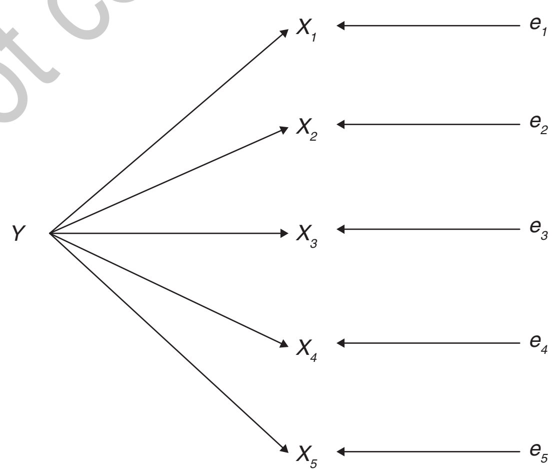
begin by examining a simple computational rule for path diagrams. Let us look at the simple path diagram in Figure 2.4.

The numbers along the paths are *standardized path coefficients*. Each one expresses the strength of the causal relationship between the variables joined by the arrow. The fact that the coefficients are standardized means that they all use the same scale to quantify the causal relationships and that their values can range from –1.0 to +1.0. In this diagram, Y is a cause of X_1 through X_5. A useful relationship exists between the values of path coefficients and the correlations between the Xs (which would represent items in the case of a scale-development–type path diagram). For diagrams like this one having only one common origin (Y in this case), the correlation between any two Xs is equal to the product of the coefficients for the arrows forming a route, through *Y*, between the *X* variables in question. For example, the correlation between X_1 and X_5 is calculated by multiplying the two standardized path coefficients that join them via *Y*. Thus, $r_{1,5} = .6 \times .1 = .06$. Variables X_6 and X_7 also share *Y* as a common source, but the route connecting them is longer. However, the rule still applies. Beginning at X_7, we can trace back to *Y* and then forward again to X_6 (or in the other direction, from X_6 to X_7). The result is $.3 \times .3 \times .4 \times .2 = .0072$. Thus, $r_{6,7} = .0072$.

This relationship between path coefficients and correlations provides a basis for estimating paths between a latent variable and the items that it influences. Even though the latent variable is hypothetical and unmeasurable, the items are real and the correlations among them can be directly computed. By using these correlations, the simple rule just discussed, and some assumptions about the relationships among items and the true score, we can come up with estimates for the paths between the items and the latent variable. We can begin with a set of correlations among variables. Then, working backward from the relationship among paths and correlations, we can determine what the values of certain paths must be if the assumptions are correct. Let us consider the example in Figure 2.5.

This diagram is similar to the example considered earlier in Figure 2.4, except that there are no path values, the variables X_6 and X_7 have been dropped, the remaining *X* variables represent scale items, and each item has a variable (error) other than *Y* influencing it. These *e* variables are unique in the case of each item and represent the residual variation in each item not explained by *Y*. This diagram indicates that all the items are influenced by *Y*. In addition, each is influenced by a unique set of variables other than *Y* that are collectively treated as error.

FIGURE 2.5 ● A path diagram with error terms



This revised diagram represents how five individual items are related to a single latent variable, Y. The numerical subscripts given to the *e*s and *X*s indicate that the five items are different and that the five sources of error, one for each item, are also different. The diagram has no arrows going directly from one X to another X or going from an e to another e or from an e to an X other than the one with which it is associated. These aspects of the diagram represent assumptions that will be discussed later.

If we had five actual items that a group of people had completed, we would have item scores that we could then correlate with one another. The rule examined earlier allowed the computations of correlations from path coefficients. With the addition of some assumptions, it also lets us compute path coefficients from correlations—that is, correlations computed from actual items can be used to determine how each item relates to the latent variable. If, for example, X_1 and X_4 have a correlation of .49, then we know that the product of the values for the path leading from Y to X_1 and the path leading from Y to X_4 is equal to .49. We know this because our rule established that the correlation of two variables equals the product of the path coefficients along the route that joins them. If we also assume that the *two path values are equal,* then they both must be .70.[1]

Further Elaboration of the Measurement Model

Classical Measurement Assumptions

The classical measurement model—which asserts that an observed score, X, results from the summation of a true score, T, plus error, e—starts with common assumptions about items and their relationships to the latent variable and sources of error:

1. The amount of error associated with individual items varies randomly. The error associated with individual items has a mean of zero when aggregated across a large number of people. Thus, items' means tend to be unaffected by error when a large number of respondents complete the items.
2. One item's error term is *not* correlated with another item's error term; the only routes linking items always pass through the latent variable, never through any error term.
3. Error terms are *not* correlated with the true score of the latent variable. Note that the paths emanating from the latent variable do not extend outward to the error terms. The arrow between an item and its error term aims the other way.

The first two assumptions above are common statistical assumptions that underlie many analytic procedures. The third amounts to defining "error" as the residual remaining after considering all the relationships between a set of predictors and an outcome or in this case, a set of items and their latent variable.

Parallel Tests

Classical measurement theory, in its most orthodox form, is based on the assumption of parallel tests. The term *parallel tests* stems from the fact that one can view each individual item as a "test" for the value of the latent variable. For our purposes, referring to parallel items would be more accurate. However, we will defer to convention and use the traditional name.

A virtue of the parallel tests model is that its assumptions make it quite easy to reach useful conclusions about how individual items relate to the latent variable based on our observations of how the items relate to one another. Earlier, we suggested that, with knowledge of the correlations among items and with certain assumptions, one could make inferences about the paths leading from a causal variable to an item. As will be shown in the next chapter, being able to assign a numerical value to the relationships between the latent variable and the items themselves is quite important. Thus, in this section, I will examine in some detail how the assumptions of parallel tests lead to certain conclusions that make this possible.

The rationale underlying the model of parallel tests is that each item of a scale is precisely as good a measure of the latent variable as any other of the scale items. The individual items are thus *strictly parallel,* which is to say that each item's relationship to the latent variable is presumed identical to every other item's relationship to that variable *and* the amount of error present in each item is also presumed to be identical. Diagrammatically, this model can be represented as shown in Figure 2.6.

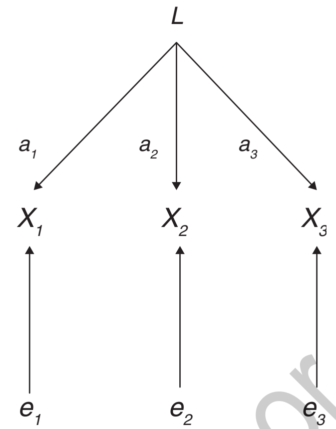
This model adds two assumptions to those listed earlier:

1. The amount of influence from the latent variable to each item is assumed to be the same for all items.
2. Each item is assumed to have the same amount of error as any other item, meaning that the influence of factors *other* than the latent variable is equal for all items.

These added assumptions mean that the correlations of each item with the true score are identical. Being able to assert that these correlations are *equal* is important because it leads to a means of determining the *value* for each of these identical correlations. This, in turn, leads to a means of quantifying reliability, which will be discussed in the next chapter.

Asserting that correlations between the true score and each item are equal requires *both* of the preceding assumptions. A squared correlation is the proportion of variance shared between two variables. So if correlations between the true score and each of two items are equal, the proportions of variance shared

FIGURE 2.6 ● A diagram of a parallel tests model, in which all pathways from the latent variable (*L*) to the items (X_1, X_2, X_3) are equal in value to one another, as are all pathways from the error terms to the items



between the true score and each item also must be equal. Assume that a true score contributes the same *amount* of variance to each of two items. This amount can be an equal *proportion* of total variance for each item only if the items have identical total variances. In order for the total variances to be equal for the two items, the amount of variance each item receives from sources other than the true score must also be equal. As all variation sources other than the true score are lumped together as error, this means that the two items must have equal error variances. For example, if X_1 got 9 arbitrary units of variation from its true score and 1 from error, the true score proportion would be 90% of total variation. If X_2 also got 9 units of variation from the true score, these 9 units could be 90% of the total only if the total variation were 10. The total could equal 10 only if error contributed 1 unit to X_2 as it did to X_1. The correlation between each item and the true score then would equal the square root of the proportion of each item's variance that is attributable to the true score or roughly .95 in this case.

Thus, because the parallel tests model assumes that the amount of influence from the latent variable is the same for each item *and* that the amount from other sources (error) is the same for each item, the proportions of item variance attributable to the latent variable and to error are equal for all items. This also means that, under the assumptions of parallel tests, standardized path coefficients from the latent variable to each item are equal for all items. It was assuming that standardized path coefficients were equal that made it possible, in an earlier example, to compute path coefficients from correlations between items. The path diagram rule relating path coefficients to correlations, discussed earlier, should help us understand why these equalities hold when one accepts the preceding assumptions.

The assumptions of this model also imply that correlations among items are identical (e.g., the correlation between X_1 and X_2 is identical to the correlation between X_1 and X_3 or X_2 and X_3). How do we arrive at this conclusion from the assumptions? The correlations are all the same because the only mechanism to account for the correlation between any two items is the route through the latent variable that links those items. For example, X_1 and X_2 are linked only by the route made up of paths a_1 and a_2. The correlation can be computed by tracing the route joining the two items in question and multiplying the path values. For any two items, this entails multiplying two paths that have identical values (i.e., $a_1 = a_2 = a_3$). Correlations computed by multiplying equal values will, of course, be equal.

The assumptions also imply that each of these correlations between items equals the square of any path from the latent variable to an individual item. How do we reach this conclusion? The product of two different paths (e.g., a_1 and a_2) is identical to the square of either path because both path coefficients are identical. If $a_1 = a_2 = a_3$ and $(a_1 \times a_2) = (a_1 \times a_3) = (a_2 \times a_3)$, then each of these latter products must also equal the value of any of the paths multiplied by itself. Looking back at Figure 2.6 may make these relationships and their implications clearer.

It also follows from the assumptions of this model that the proportion of error associated with each item is the complement of the proportion of variance that is related to the latent variable. In other words, any effect on a given item that is not explained by the latent variable must be explained by error. Together, these two effects explain 100% of the variation in any given item. This is so simply because the error term (*e*) is defined as encompassing all sources of variation in the item other than the latent variable.

These assumptions support at least one other conclusion: Because each item is influenced equally by the latent variable and each error term's influence on its corresponding item is also equal, the items all have equal means and equal variances. If the only two sources that can influence the mean are identical for all items, then clearly the means for the items also will be identical. This reasoning also holds for the item variances.

In conclusion, the parallel tests model assumes the following:

1. Error is random.
2. Errors are not correlated with one another.
3. Errors are not correlated with true score.
4. The latent variable affects all items equally.
5. The amount of error for each item is equal.

These assumptions allow us to reach a variety of interesting conclusions. Furthermore, the model enables us to make inferences about the latent variable based on the items' correlations with one another. However, the model accomplishes this feat by setting forth fairly stringent assumptions.

Alternative Models

As it happens, all the narrowly restrictive assumptions associated with strictly parallel tests are not necessary in order to make useful inferences about the relationship of true scores to observed scores. A model based on what are technically called *tau-equivalent tests* makes a more liberal assumption—namely, that the amount of error variance associated with a given item need not equal the error variance of the other items (e.g., Allen & Yen, 1979). Tau-equivalent tests still require identical true scores for items, although a slight loosening of that assumption defines *essentially tau-equivalent tests* (or occasionally, *randomly parallel tests*). Any pair of items adhering to essential tau equivalence may have true scores that differ by some constant. Of course, adding a constant to one item has no effect on any correlation involving that item because correlations are standardized expressions. Consequently, the correlation between any pair of items or between an item's true score and the item's obtained score is not affected by relaxing the assumptions of strict tau equivalence to those of essential tau equivalence. So what we have said thus far about tau equivalence also applies to essential tau equivalence. In either of these cases, the *standardized* values of the paths from the latent variable to each item may not be equal. However, the *unstandardized* values of the path from the latent variable to each item (i.e., the *amount* as opposed to *proportion* of influence that the latent variable has on each item) are still presumed to be identical for all items. This means that items are parallel with respect to how much they are influenced by the latent variable but are not necessarily influenced to exactly the same extent by extraneous factors that are lumped together as error. Under strictly parallel assumptions, not only do different items tap the true score to the same degree; their error components are also the same. Tau equivalency (tau is the Greek equivalent to *t*, as in true score) is much easier to live with because it does not impose the "equal errors" condition. Because errors may vary, item means and variances may also vary. The more liberal assumptions of this model are attractive because finding equivalent measures of equal variance are rare. This model allows us to reach many of the same conclusions as with strictly parallel tests but with less restrictive assumptions. Readers may wish to compare this model with Nunnally and Bernstein's (1994) discussion of the domain sampling model.

Some scale developers consider even the essentially tau-equivalent model too restrictive. After all, how often can we assume that each item is influenced by the latent variable to the same degree? Tests developed under what is called the *congeneric model* (Jöreskog, 1971) are subject to an even more relaxed set of assumptions (see Carmines & McIver, 1981, for a discussion of congeneric tests). This model assumes (beyond the basic measurement assumptions) merely that all the items share a common latent variable. They need not bear equally strong relationships to the latent variable, and their error variances need not be equal. One must assume only that each item reflects the true score to some degree. Of course, the more strongly each item correlates with the true score, the more reliable the scale will be.

An even less constrained approach is the *general factor model,* which allows multiple latent variables to underlie a given set of items. Carmines and McIver (1981), Loehlin (1998), and Long (1983) have discussed the merits of this type of very general model, chief among them being its improved correspondence to real-world data. Structural equation modeling approaches often incorporate factor analyses into their measurement models; situations in which multiple latent variables underlie a set of indicators exemplify the general factor model (Loehlin, 1998).

The congeneric model is a special case of the factor model (i.e., a single-factor case). Likewise, an essentially tau-equivalent measure is a special case of a congeneric measure—one for which the relationships of items to their latent variable are assumed to be equal. Finally, a strictly parallel test is a special case of an essentially tau-equivalent one, adding the assumption of equal relationships between each item and its associated sources of error.

Another measurement strategy should be mentioned. This strategy is item response theory (IRT). This approach has been used primarily but not exclusively with dichotomous-response (e.g., correct versus incorrect) items in developing ability tests. IRT assumes that each individual item has its own characteristic sensitivity to the latent variable, represented by an item-characteristic curve—a plot of the relationship between the value of the latent variable (e.g., ability) and the probability of a certain response to an item (e.g., answering it correctly). Thus, the curve reveals how much ability an item demands to be answered correctly. We will consider IRT further in Chapter 8.

In Chapters 6, 7, and 8, we will look at factor analysis, indices, and item response theory respectively. In those chapters, we will necessarily go beyond the models we have discussed so far. In Chapters 1 through 5, however, we will focus primarily on parallel and essentially tau-equivalent models for several reasons. First, they exemplify "classical" measurement theory. Second, discussing the mechanisms by which other models operate can quickly complicate topics unnecessarily if those models are not necessary to a basic understanding. Finally, classical models have proven very useful for social scientists with primary interests other than measurement who, nonetheless, take careful measurement seriously. This group is the audience for whom the present text has been written. For these individuals, the scale development procedures that follow from a classical model generally yield satisfactory scales. Indeed, to my knowledge although no tally is readily available, I suspect that (outside ability testing) a substantial majority of the well-known and highly regarded scales used in social science research were developed using such procedures.

Choosing a Causal Model

Choosing the causal model that underpins a variable, when feasible, can be an important aspect of measurement. The very conceptualization of a variable can sometimes be subtly adapted at the outset of a research project to make its eventual measurement more manageable. As an example, consider a researcher who wants to assess how the physical work environment affects employee productivity. One approach might be to develop a long list of environmental factors that are thought to influence productivity—such as lighting, sense of privacy, or access to a computer—and develop an instrument that has workers rate the extent to which those factors are present in a given workplace. A problem with this approach is that the instrument may end up being an index rather than a scale or perhaps a hybrid of the two (topics we discuss in Chapter 7). That is, the indicators (e.g., good lighting, reasonable privacy, computer access) might not really share a common cause but rather a common effect, namely, an improvement in the work environment. If, instead, the investigator considered the eventual measurement problem early on in the research process, he or she may have decided to conceptualize the variable somewhat differently. For example, had the investigator defined the variable of interest as employees' perceptions of the work environment, that definition may have led to a more tractable set of items. For example, employees could be asked to endorse items such as, "My workplace environment provides the basic equipment I need to do my job effectively." Here, the latent variable is not a feature of the environment per se but the employees' perceptions. How the employees perceive the environment is the common cause driving their responses to individual items. It may be easier to assume that an employee has a sense of the work environment that will give rise to answers across a set of questions about its adequacy than to imagine the environment itself as a cause of employee responses. Moreover, the psychological nature of employee perceptions may actually be closer to what the investigator considered relevant to productivity than the mere presence or absence of specific environmental features. That is, whether a given worker perceives the environment as conducive to productivity may be a more relevant variable than someone else's judgment regarding the adequacy of the work environment. So conceptualizing the variable of interest in this way may serve the underlying research question well while also potentially facilitating the eventual measurement of the variable.

Of course, if the variable simply does not lend itself to a causal conceptualization consistent with a straightforward measurement strategy, the integrity of the variable of interest should not be compromised. Chapter 7 offers ways to proceed in those instances. Certain approaches may help the investigator work around the limitations inherent in the variable and the way in which it is operationalized. But if an acceptable alternative conceptualization of the variable and the model relating it to its indicators can be simplified, it well may be possible to develop a measurement tool that meets a simpler set of assumptions and thus can be explored using less complex analytic tools. Having the tools to handle the more complex situations is certainly a good thing, but avoiding those complexities and precluding the need for those more advanced tools may be even better, assuming that it does justice to the construct.

Exercises

1. How can we infer the relationship between the latent variable and two items related to it based on the correlations between the two items?
2. What is the chief difference in assumptions between the parallel tests and essentially tau-equivalent models?
3. Which measurement model assumes, beyond the basic assumptions common to all measurement approaches, only that the items share a common latent variable?
4. Assume an essentially tau-equivalent model with true score *T* and indicators A, B, and C. In such a model, any two indicators (e.g., A and B) that share a common true score must have a covariance identical to the covariance between any other two indicators (e.g., B and C) sharing that true score. However, the correlations between different pairs of indicators need not be equal. Explain why this is so.

Note

1. Although -.70 is also an allowable square root of .49, deciding between the positive or negative root is typically of less concern than one would think. As long as all the items can be made to correlate positively with one another (if necessary, by reverse scoring certain items, as discussed in Chapter 5), then the signs of the path coefficients from the latent variable to the individual items will be the same and are arbitrary. Note, however, that giving positive signs to these paths implies that the items indicate more of the construct, whereas negative coefficients would imply the opposite.