Difference between path analysis and sem

A diagram of what this equation does is shown in Figure 20–1. For reasons that will be clarified soon, we have drawn a box around each of the variables and a straight arrow3 that leads from each of the predictor variables to the dependent variable (DV). This implies that we are assuming that each of the variables acts directly on the DV. But a little reflection can lead us to feel that the story is a little more complicated than this. A person's height can act directly on the coach's assessment, but it can also influence jumping ability. Therefore, a more accurate image may be the one shown in Figure 20–2. One question we may want to ask is, which model is a more accurate reflection of reality? We'll start by just looking at how the variables are correlated with each other, to get an idea of what's going on; the correlation matrix is shown in Table 20–1. As we would suspect, CLAP/TRAP has a strong positive relationship with Height and Leap Capacity, and is strongly and negatively related to IQ4; Height and Leap are positively correlated with each other and are negatively related to IQ.5 Interpretation of numbers If we now run a multiple regression based on the model in Figure 20–1, we would get (among many other things) three standardized regression weights (betas, or s), one for each of the predictors. In this case, the light value is 0.548, the Jumping error is 0.199 and the IQ value is 0.245. While the relative magnitudes of the β and their signs are parallel to those of the correlations, the relationship between these two sets of parameters is not immediately obvious. The problem is that the model in Figure 20–1 shows only part of the image; it does not take into account the correlations between the predictor variables themselves. To introduce a notation that we will use a lot in this chapter, we can show that the variables are correlated with each other by joining the boxes with two-pointed curved arrows.6 In Figure 20–3, we have added the arrows and some numbers. Numbers near curved arrows are zero-order correlations between predictor variables, and those above single-headed arrows are correlations between predictors and DV. Below the arrows, and in parentheses, are the β weights. Now, believe it or not, we can see the relationship between correlations and weights β. Let's start by seeing the effect of the IQ on CLAP/TRAP. Its β weight is 0.245. In addition, IQ exerts an indirect effect on CLAP/TRAP through its correlations to jump ability and height. Through the jump ability, the magnitude of the IQ effect is the correlation between the two predictors (r -0.433) times the effect of Jump in CLAP/TRAP, which is its β weight of 0.199; and therefore 0.433 × 0.199 to Similarly, the indirect effect of the IQ through height is 0.505 × 0.548 or 0.277. The sum of these three terms gives us: that is the correlation between IQ and CLAP/TRAP. Doing the same for the jump ability shows a direct effect of 0.199; an indirect effect through the IQ of × 0.245 (which is 0.106) and an indirect effect through Height of 0.678 × 0.548 or 0.372. The sum of these three terms is 0.677, which, in this case,7 is the zero-order correlation between Jumping and CLAP/TRAP.8 Conceptually, so r is the sum of the direct and indirect effects between the two variables. FIGURE 20-1 Linear regression model predicting CLAP/TRAP scores for Height, Jump Skill, and IQ. FIGURE 20-2 Modification of Figure 20-1 to show that the Jump Skill is affected by Height. TABLE 20–1 Correlations between variables in the prediction of CLAP/TRAP FIGURE 20-3 Scores Figure 20-1 with correlations between variables and β weights in added parentheses. More formally, we can denote the correlation between pairs of predictors with the usual symbol for a correlation, r. So the correlation between IQ and Jumping is rIQ-Jump, between Height and Jumping is rHeight-Jump, and between IQ and Height is rIQ-Height. The path between each of the predictors and the dependent variable is a β weight, so between Jumping and CLAP/TRAP is Jump, between IQ and CLAP/TRAP is IQ, and between Height and CLAP/TRAP is . Height. Therefore: FIGURE 20-4 Figure 20-2 with added path coefficients. TABLE 20–2 Original correlations in the upper triangle and correlations reproduced in the lower triangle What we have just done is to break down the correlations of the predictor variables into their direct and indirect effects on the DV. We are now in a better position to see what adding the arrow in Figure 20–2 does to our model. The main effect is that it imposes directionality on indirect effects; we're saying that Height can affect the Leap Skill (which makes sense), but that the Jump Skill doesn't affect The Height (which wouldn't make sense9). In other words, we have outlined the paths through which variables exert their effects; Not surprisingly, Figure 20–2 is called a pathmap and has just been introduced to the basic elements of route analysis. In Figure 20–4, we have added the route coefficients to the second model (the one in Figure 20–2). Direct effects are changed only slightly (although they are changed), but now the indirect effects are very different. If we follow the arrows that lead to each predictor variable, we see that Skip has a direct effect of 0.200 and, strangely it may seem, an indirect effect through Height of 0.678 × 0.550 (0.373), for a total of 0.573.10 For Height, the direct effect of 0.550 is increased by its indirect effect through the jumping capacity from 0.678 × to 0.200, or 0.136, and its indirect effect through the IQ of EUR 0.505 × to EUR 0.247 (0.125), so that its total effect is EUR 0.810. The IQ is even less simple; has a direct effect (0,247), an indirect effect on from Height to CLAP/TRAP (0.505 × 0.550 to 0.278), and a very indirect effect through The Height to Jump to CLAP/TRAP (0.505 × 0.678 × 0.200 to 0.068), for an overall effect of 0.593 euros. If you were paying attention, attention, you have noticed that, in this example, these numbers are not added up to match the correlations in Table 20–2. We'll show you why they don't when we discuss how we can tell which models are better than others. Finding their way through the paths When we break down the correlations in Figures 20–3 and 20–4, there were some paths we did not travel through, and others we took seemed somewhat strange. For example, in Figure 20–3, we track the indirect contribution of IQ → Jumping → CLAP/TRAP and IQ → Height → CLAP/TRAP; we didn't have an IQ → Jumping → Height → CLAP/TRAP path. Similarly, in Figure 20–4, we had height → jump → CLAP/TRAP, and, interestingly, Jump → to Height → CLAP/TRAP, although the arrow itself points from Height to Jump. In 1934, Sewall Wright (the grandfather of road analysis) established the rules of the road: For any road, you can go through a given variable only once. Once you've walked a path with an arrow, you can't go back to the road using a different arrow. You cannot pass through a two-pointed curved arrow more than once. Kenny (1979) added a fourth rule: You can't insert a variable into one arrowhead and leave it at another arrowhead. Kenny's rule is why, in Figure 20–3, we do not trace the IQ → Jumping → Height → CLAP/TRAP or IQ → Height → Jumping → CLAP/TRAP path. These routes enter a variable at one arrowhead from the IQ, and then they would go on one arrowhead to reach the other variable, perhaps not a felony, but definitely a misdemeanor with respect to the rules. Oddly enough, the path in Figure 20–4 that begins in Leap and then goes through Height to CLAP/TRAP does not violate any of these rules. The rule prohibits only routes that advance and then go back;, this route goes backwards, and then forwards. This path makes no sense in terms of our knowledge of biology (the technical term for this is a spurio effect), but it is legitimate when it comes to the breakdown of correlation. It exists because Jumping and CLAP/TRAP have a common cause: Height. Route and causation analysis Until recently, route analysis was called causal modeling. 11 Because we can specify routes by which we think one variable affects another, the hope was that we could make statements about causation, even if the data were really correlal. In fact, if we specified a model that didn't make much sense, such as having the way to go from IQ to Jump Capacity, we would know from statistics that something was wrong; there shouldn't be a path between IQ and jumping ability. Does this mean we can actually take correlation data and get them to tell us about causation? Despite the hopes of the first developers of this technique, the answer is No. If we had the model making the path go from the jump skill to the height, rather than the height to the jump ability, as in Figure 20–5, all the stats would be the same. I mean, this technique can tell us or there should not be a path between two variables, but it cannot tell us where the arrow should point; this can only be supplied by our theory, study design, or knowledge of literature. For example, if we find a relationship between gender and spatial or verbal ability, it is quite certain to assume that gender (or factors correlated with gender, such as educational experiences or brain structure) leads to differences in capacity. No matter how much we change people's abilities, their gender will remain the same.12 Similarly, if we modify an educational program at time 1, and find that the experimental group, but not the control, improved their skills at time 2, the causality is evident from the design of the study. Route analysis can also refute a causal model that we may or may not refute. However, not rebutting the null hypothesis does not mean that we have demonstrated it, not rejecting a causal model is not the same as showing that it is correct. The bottom line is that determining the true causality of correlations is still similar to getting lead gold. Endogenous and exogenous variables The models we have discussed so far are relatively simple. To show you what more route analysis can do, let's take a look at some models you can handle while presenting you with some of the arcane vocabularies. When we look at Figure 20–1, we refer to the variables on the left as the predictors and CLAP/TRAP as the dependent variable. Now that we have a new statistical technique, we have new terms for these variables.13 In route analysis (and, as we will see, in SEM in general), the variables on the left are known as exogenous variables. Exogenous variables have arrows that emerge from them and none pointing at them. FIGURE 20-5 Figure 20-2 with jump ability that affects height. This means that exogenous variables influence or affect other variables, but what influences them is not included in the model. For example, a person's height may be influenced by genetics and diet, but our model will ignore these factors. What we had called the dependent variable, CLAP/TRAP, is an endogenous variable in SEM terms, as it has arrows pointing towards it. What about the ability to jump in Figure 20–2? It has an arrow pointing towards it, as well as one that emerges from it. As long as there is at least one arrow (a path) pointing toward a variable, it is called endogenous. Any variable that has at least one arrow pointing towards it is an endogenous variable. This illustrates why independent, predictor, and dependent variable terms can be confusing in route analysis. In Figure 20–4, Jumping is a variable dependent on Height but a predictor in terms of its relationship with This also shows one of the main strengths of route analysis compared to multiple regression; regression cannot easily address the situation in which a variable is a separate variable (IV) and a DV. Route types Figure 20–6 shows a number of different route models, some of which we have already found. Those on the left are direct models, in the sense that exogenous variables influence endogenous variables without any intermediate steps; In other words, endogenous variables have arrows pointing towards them and none pointing away from them. In Part A of the figure (called an independent route model), the two exogenous variables affect an endogenous one, and the exogenous variables are not correlated with each other; hence the independent name. This situation is not too common in research with humans, for, as Meehl said in his famous sixth law (1990), everything correlates to some extent with everything else (p. 204).14 Part B of the figure, a model of correlated path, is more common; it's equivalent to multiple regression, although we often have more than two predictors, and drawing on all those arcs between pairs of variables can actually make the image look cluttered. The image it portrays is what we usually deal with, as we assume that the predictors are correlated with each other at various degrees. In Part C, there are two exogenous variables and two endogenous variables. The interesting thing about this model is that X1 and X2 can be different variables, or they can be the same variable measured at two different times; the same applies to Y1 and Y2. For example, in a study we did (McFarlane et al. 1983), we wanted to see if stressful events in a person's life led to more illness. X1 and X2 were the amount of stress at the time 1 and 6 months later at the time 2, and Y1 and Y2 were the number of diseases in these twice. This model allowed us to take into account the fact that the best predictor of future behavior is past behavior (the path between Y1 and Y2), and then look at the added effect of stress at time 1 on the disease at the time 2.15 The two endogenous variables do not have to be the same as the exogenous ones, nor do the two sets of variables have to be measured at different times, but they can be. Such is the beauty of road analysis. FIGURE 20-6 Different route models. The main difference between the diagrams to the left of Figure 20–6 and those on the right is that the latter have endogenous variables with arrows pointing both towards and away from them. These are known as indirect or mid models, because variables with both types of arrows mediate the effect of the variable pointing at them on the variables they point to. In Part D of Figure 20–6, variables X1 and X2 influence Y. However, X2 also mediated the effect of X1 in Y. For example, Jumping (X2) directly affects CLAP/TRAP and also mediates the Height (X1) effects. That is, we would expect that, if two people can jump equally high, but one person is 6 x 3 and the other only 5-8,16 we would expect the second person gets more credit as they are trying harder. In part E of the figure, variable X1 does not directly affect the endogenous, endogenous variable, only through his influence on X2. For example, the height of the parent (X1) affects a person's CLAP/TRAP (Y) score only because of its effects on the animator height (X2), which in turn affects Y. The final model we will show (there are infinitely many others), part F of Figure 20–6, is the same as part C, except that we have added a path between X2 and Y2. What this does is convert X2 (stress at time 2) into a mediated variable. Now we are saying that the number of diseases measured at time 2 is affected by three things: disease at time 1, stress at time 1 and stress at time 2. In addition, we are saying that stress at time 1 works in two ways in disease behavior at time 2: directly (a kind of delayed reaction to stress), and affecting stress at time 2, which in turn affects the behavior of the disease immediately. The magnitude of the route coefficients tells us how strong each effect is. Disturbing The models in Figure 20–6 seem quite complete and, in Part F, relatively complex. But, in fact, something is missing. Let's add a few more terms, not to annoy you, but to disrupt endogenous variables. For reasons that (as usual) are necessary, statistical program-analyzing route model) image is to attach terms of disturbance to each endogenous variable. These are usually denoted by the letter D (for Disturbances), E for (Error), a small circle with an arrow pointing toward the variable, or a circle with one of the letters inside, as in Figure 20–7. This is similar to what we have in multiple regression, where each equation has an error term stuck at the end that captures the measurement error associated with each of the predictor variables. In path analysis (and seM in general), the term disturbance has a broader meaning; in addition to the measurement error, it also reflects all the other factors that affect the endogenous variable and that are not in our model, either because we couldn't measure them (e.g. genetic factors) or we were too dumb to think about them at the time (for example, how much the cheerleader's parents bribed the coach so that their child would get a good score). If we had drawn complete diagrams for the examples we discussed, then Figure 20–3 would have a disturbing term attached to CLAP/TRAP. In Figure 20–4, there would be two such terms: one associated with CLAP/TRAP and one for Jumping, as it has become an endogenous variable with the addition of the route from Height. Why don't we draw them? Since each endogenous variable must have a term of disturbance, they are superfluous for those of us who are us courant with trajectory analysis. We have to draw them when we use the and we can put them in diagrams for the sake of integrity, but they are optional at other times. Recursive and non-recursive models Although may not be apparent at first, all models in Figure 20–6 are similar in two important ways. First, all paths are one-way17; unidirectional17; that is, they go from cause to effect (using those terms very freely, without actually meaning causality). Secondly, if we had drawn the terms of disturbance in the models, it would be assumed that all of them are independent of each other; that is, we would not draw curved arrows between disturbances. For reasons that overcome human understanding, these models are known as recursive models. Let's go back to the example we used in Figure 20–6F and modify it a little more. In Figure 20–8, we have added a path between X1 and Y1 and between X2 and Y2 to indicate that stress could affect the disease simultaneously. It might be just as logical for the relationship to go the other way and even more logical that the relationship was reciprocal: that stress at any given time affects the disease and also that the disease affects stress, so Figure 20–8 may be a more accurate representation of what is actually happening. A route diagram with this feedback loop is called a non-recursive model.18 Note that there is a difference between connecting two variables with a two-pointed curved arrow and joining them with two straight arrows that go in opposite directions. The first means that we simply expect the two variables to be correlated, as when we use two similar measures of the same thing. Covariance may be the result of one variable affecting the other, but it may also be due to the effect of some underlying factor affecting both. One feedback loop, on the other hand, explicitly indicates that each variable directly affects the other: stress leads to disease, and the disease leads to stress. FIGURE 20-7 Adding disturbing terms to Figure 20–6F. FIGURE 20-8 A non-recursive route model. Analysis and interpretation of non-recursive models is much more difficult than with recursive models, and far beyond the scope of a chapter in a book. If you ever need to analyze a non-recursive model, our best advice to you is to do two things: (1) serving yourself a rigid drink and thinking again, and (2) if you still want to do so, read one of the books listed in To read more. K.I.S.S.19 At first glance, it may seem as if the safest strategy to use is to draw paths that connect everything with everything else, and let the route coefficients tell you what's going on. This would be a bad idea for two reasons. First, as we will emphasize again and again in this chapter (and indeed much of the book), model building must be based on theory and knowledge. One disadvantage of the most sophisticated statistical techniques is that they are too powerful in a sense; they capitalize on the variance of possibility and, if you cheerfully and blindly use a computer program to replace your brain cells, you can be diverted wonderfully, albeit with low p levels. The second reason is that there are mathematical limits to routes you can have in any diagram. The number of parameters (i.e. statistical effects, such as route coefficients, variances, covariance, and disturbances) that you can analyze is determined by the number Observations. Remarks here is a function of the number of variables and is not related to the number of subjects. That is, a given route model has the same number of observations, whether the study used 10 subjects or 10,000 subjects. If there are variables k, then: because there are covariance [k × (k to 1)/2] between the variables, plus k variances. In Figure 20–3, we have four variables observed, so we have 10 observations. This means that we can examine a maximum of 10 parameters. Now the problem is, how many parameters do we really have in this model? Another way to ask this question is, what don't we know we should know? To respond to this, we need to elaborate a little more on the purpose of route analysis and, more generally, the modeling of structural equations. We are trying to figure out what affects endogenous variables; that is, how exogenous variables work together (curved paths, which represent correlations or covariance), and which of those paths (straight arrows) are important. This is determined by the deviations of the variables. If we had a perfect model, then knowing these variances would allow us to predict person-to-person variation in endogenous variables. Because endogenous variables are, by definition, caused or influenced by the other variables, they are not free to vary or covariate on their own, but only in response to exogenous variables. Consequently, we are not interested in estimating the variances of endogenous variables, but only those variances for variables that may vary. Note that the term disturbance associated with an endogenous variable is, in SEM thinking, free to vary and thus influence the endogenous variable, so it is one of those things that we have to estimate. Where does this leave us with respect to Figure 20–3? It is obvious that we want to estimate the 3 route coefficients, so they are in the parameter list to estimate. In addition, we do not know the variances of the exogenous variables or the covariance between them, so it adds another 6 parameters (3 variances + 3 covariance). Finally, we have the variance of the term of disturbance itself (which is not in the figure, but is implicit), which means that there are 10 parameters to estimate. Now let's do the same for Figure 20–4. There are still 4 variables observed, so the limit of 10 parameters remains, but there are a different number of arrows. We have 4 route coefficients (the 3 of the variables to CLAP/TRAP, and 1 from Height to Leap); covariance between IQ and height; variances of 2 exogenous variables (IQ and Height); and 2 terms of disturbance (jumping and CLAP/TRAP), resulting in 9 parameters to be estimated. If you have followed this so far, a number of questions should arise: Why don't we count the CLAP/TRAP variance in the 20–3 as one of the parameters to be estimated? Why do we count the Leap variance between the parameters in Figure 20–3, but not Figure 20–4? Why can't we have more than 10 parameters? makes make any difference that we had 10 parameters to estimate in Figure 20–3 and 9 in Figure 20–4? Is there intelligent life on Earth? If you've really been paying attention, you'll have noticed that we've already answered the first two questions (but we'll repeat it for your sake). Stay tuned, and the remaining questions will be answered (except perhaps the last question, which has baffled scientists and philosophers for centuries). To reiterate what we have said, we do not count the variance of CLAP/TRAP in any of the models, nor in the Leap one in Figure 20–4, for the same reason: they are endogenous variables and are therefore not free to vary and covariate on their own. Because the goal of SEM is to explain variable variances and covariance between pairs of variables that may vary, we are not interested in variable parameters that are determined by external influences. Therefore, endogenous variables do not enter the count. Why can't we examine more parameters than observations? The analogy is to have an equation with terms more unknown than data. For example, if we had a simple equation: and we know that at 5, then what are the values of b and c? The problem is that there are an infinite number of possible answers: b- 0 and c- 5; b- 1 and c to 4; 19.32 and c to 24.32; and so on. We say that the model is undefined (or subidentified) that there is no single solution. If we had another observation (e.g. b 3), then we can determine that c has to be 8; the model is now defined (the technical term is just identified). If there were as many observations as parameters (i.e. we knew in advance that at 5, b 3 and c to 8), then the model would be correct, but there would be nothing left to estimate (this is known as overidentified). This is the situation in Figure 20–3, where there are 10 observations and 10 parameters. In the next section, we'll look at the implication of this. As Good as It Gets: Good fit indicators How do we know if the model we have applied for is good? In route analysis, there are three things we seek: the importance of individual roads; reproduced (or implicit) correlations; and the model as a whole. Of the three, the importance of the roads is the easiest to assess. Path coefficients are parameters and, as is the case with all parameters, are estimated with some degree of error. Again, like other parameters, the importance depends on the relationship of the parameter to its standard estimation error. In this case, we end up with a z:y statistic (once again like other z-tests), if it is 1.96 or higher, it is significant at the .05 level, using a two-queue test.20 A second criterion refers to the reproduced (or implicit) correlation matrix. You'll remember that when we reproduced the correlations of the route coefficients in Figure 20–3, we were able to correlations perfectly plotting all routes. However, when we try to reproduce correlations correlations Figure 20–4, we find differences between actual and implicit correlations. In fact, if you look at the correlations reproduced in the lower half of Table 20–2, some of them differ quite a bit from the original correlations, which are in the upper half of the matrix. This tells us two things: (1) the model in Figure 20–3 fits the data better than the model in Figure 20–4; and (2) the model in Figure 20–3 fits the data too well, in the sense that there is no discrepancy between the original and reproduced correlations. The reason is obvious when comparing the number of observations in Figure 20–3 with the number of parameters; there are 10 of each. Like the case in which we are told in advance that to b + c and that a - 5, b - 3, and c - 8, there is nothing to estimate, and there is a perfect fit between the model and the data. Life is often much more interesting when we have fewer parameters than observations; then we can compare how close the different models are to estimating the data. When we look at the model as a whole, the main statistic we use is called the goodness of fit n.o 2 (2GoF), which we will find later in different contexts. In most statistical tests, the bigger it is21: the higher the value of a t, F-test, r, or whatever, the happier we are. This situation is reversed in the case of 2GoF, where we want the result to be as small as possible. Why this sudden change of heart? Let's go over the logic of the No. 2 tests. Basically, they all come down to the difference between what we observed and what would be expected, with expected what the values we would find if the null hypothesis were true. Therefore, the greater the discrepancy between the two sets of numbers (observed and expected), the greater the number 2. Because we normally want our results (observed values) to be different from nulls, we want the value of 2 to be large. However, when we use the number 2 to test goodness of fit, we are not testing our observed findings against the null hypothesis, but rather against some hypothetical model, and we want our results to be consistent with it. Therefore, the less you deviate from the model results, the better. In the case of 2GoF, we expect to find a non-negligible result, indicating that our results match the model. The degrees of freedom (df) associated with 2GoF is the difference between the number of observations and the number of parameters. For the model in Figure 20–4, we have 10 observations and 9 parameters to estimate; therefore, df 1. In this example, the value of 2GoF is 2,044, which has a p-level of .153. Because 2GoF is not statistically significant, we have no reason to reject the model. Just to reinforce what we said earlier, the 2GoF for the model in Figure 20-5, where the arrow between Leap and Height was the wrong way, it is also 2,044; so this statistic does not tell us about the causality of the relationship between variables. For the sake of integrity, we also discussed a different model, one that was ridiculous that we were even ashamed to draw it. This was a model where there was a path between IQ and jumping ability. The number 2GoF for this model (which also has df 1) is 42,525, which is very significant, and it means that there is a very large discrepancy between the model and the data. In Figure 20–3, we cannot calculate a value of 2GoF, because the number of parameters is equal to the number of observations, which means that df is 0. This also illustrates that when a model is completely determined (an elegant way of saying that the number of parameters and observations is the same), the data fits the model perfectly. So, if we find that the 2GoF is not significant, does this prove that our model is correct? Unfortunately, the answer is a resounding No. As we just showed, the No 2 statistics associated with Figure 20–2 and Figure 20–5 are identical, even though one model makes sense and the other is obviously ridiculous.22 In addition, 2GoF depends on the sample size, as this influences the standard errors of the estimates. If the sample size is low (say, below 100), even models that defever a lot from the data may not result in statistically significant 2s. On the contrary, if the sample size is very large (more than 200 or more), it is almost impossible to find a model that does not deviate from the data to some extent. In addition, there is no guarantee that there will be another model, untested, that can be better adjusted to the data. Therefore, 2GoF can tell us if we are on the right track, but it cannot provide definitive proof that we have arrived. What we assume the route analysis makes certain assumptions, as does all statistical tests. Many of the assumptions are the same as for multiple regression, which is not surprising, since the two techniques are closely related. The first assumption is that variables are measured without errors. This is obviously impossible in most (if not all) research, but it serves to remind us that we should try to use instruments that are as reliable as possible. When this rule is violated (as it always is), it results in underestimates of the effects of mediator variables (i.e. indirect paths) and overestimates the effects of direct paths (Baron and Kenny, 1986). The second assumption, which is much harder to detect, is that all important variables are included in the model. It's not a good idea to include variables that aren't important, but they usually become obvious from their weak path coefficients. However, when left out of crucial variables, model fit can be poor or produce spur results. No statistical test can tell us which variables have been omitted; only our theory can in this sense. Third, multiple regression and route analysis assumes that variables are additive. If there are interactions between variables, an appropriate interaction term must be incorporated into the model (Klem, 1995). Finally, both techniques can handle a moderate degree of correlation between predictor predictor variables but parameter estimates become unreliable if correlations are high (Klem, 1995; Streiner, 1994). A word about the sample size The df associated with the 2GoF test is the difference between the number of observations and the number of parameters. So where does the sample size come in? Sample size affects the importance of parameter estimates: path coefficients, variances, and covariance. In all cases, the importance is the relationship of the parameter to the standard error (SE), and the standard error, as we all know, depends on the square root of N. That said, how many topics do we need? Unfortunately, there is no simple relationship between the number of parameters to be estimated and the sample size. A very approximate rule of law is that there must be at least 10 subjects per parameter (some authors advocate for 20), as long as there are at least 200 subjects. Yes, Virginia, route analysis (and SEM in general) is extremely greedy when it comes to sample sizes. MODEL OF STRUCTURAL EQUATIONS The main limitation with route analysis is that our drawings are restricted to circles and boxes, which become extremely monotonous. Wouldn't it be nice if we could add some variety, like ovals? Of course, the answer is Yes; Why else would we ask the question? This isn't as ridiculous as it first appears. In SEM drawing conventions, circles represent terms of error or disturbance, and boxes are drawn to display measured variables, that is, we observe them directly on a physical scale (for example, Height) or a paper and pencil scale (for example, CLAP/TRAP). In the chapter on factor analysis, however, we introduced ourselves into another type of variable: latent variables (which we refer to in that chapter as factors). Very briefly, latent variables are not measured directly; rather, they are infered as the glue that joins two or more observed variables (i.e. directly measured). For example, if a person gets a high score on some paper and pencil anxiety test (a measured variable), shows an increase in heart rate in enclosed spaces (another measured variable) and uses anxiolytic medications (still a third observed variable), we would say that these are all manifestations of the invisible factor (or latent variable) of anxiety. FIGURE 20-9 Relationship between three measured variables and a latent variable. The main difference between route and seM analysis is that the former can examine relationships only between measured variables, while sem can examine measured and latent variables. Because we represent latent variables with ovals, there is the added advantage that our diagrams can now become more varied and sexy. Figure 20–9 illustrates the example use in SEM terms. Notice the direction of the arrows: point from the latent variable to the measurements, which at first glance may appear backwards. In fact, this reflects our conceptualization of latent variables, which are the underlying causes of we're measuring directly. Therefore, the latent variable (or trait or factor) of anxiety is what explains the person's score on the test, their tachycardia, and is what leads to the use of pills. In addition, since the three measured variables are endogenous, they have terms of error or disturbance associated with them. If Figure 20–9 looks suspiciously like the images we drew when we were discussing factor analysis, it's not a coincidence. Exploratory Factor Analysis (EFA, the type we discussed in Chapter 19) is now seen as a subset of SEM. We didn't earn much, however, if we had to use SEM programs and techniques to do EFA. In fact, because SEM is a test model or confirmation technique, rather than an exploratory one, the output of the equipment is less useful than that of programs designed specifically to make EFA (if that's what we want to do). Despite this, we'll start with EFA to show you how the concepts we cover in route analysis apply in this relatively simple case. This will put us in a better position to understand how SEM works in more general cases.23 SEM and Factor Analysis In addition to the Height, Jump Capacity and IQ variables, Dr. Teeme also hypothesized that since the first part of the word cheerleading is cheerleading, success in this effort also depends on the candidate's personality. However, by not finding a questionnaire that directly measures Cheer, he had to resort to using three other tests that he felt was felt to be measured collectively the same: one when touching the extroversion (the seller of the used car scale, or SUCS); another positive outlook on life (Mary Poppins Inventory, or MPI); and a third party that focuses on denial of negative feelings (scale We're FINE, or WOKS). Correlations between these scales are shown in Table 20–3. As I suspected, correlations were moderate, but positive. If we were now an EFA, using a least squares method to extract the factor, we would find that the factor loads were 0.840 for SUCS, 0.797 for MPI, and 0.724 for WOKS. A drawing of this using SEM conventions is shown as Figure 20–10. There are a few points to keep in mind. First, the term disturbance or error in EFA is usually labeled with the letter U (meaning uniqueness); the terminology is different, but the concept is the same.24 Second, for each variable, the square of the factor load (which is equivalent to a path coefficient) plus the uniqueness square is equal to 1.00 (for example, 0.8402 + 0.5432 to 1.00). In English, we have divided the variable variance into two components: which is explained by the factor (or latent variable); and what is not explained by it (uniqueness, error or disturbance). Finally, the product of two-way loads equals the correlation between variables. For example, the factor loads for MPI and WOKS are 0.797 and 0.724, respectively; your product (i.e. 0.797 × 0.724) is 0.577, or its correlation. Now, if we run this like it's a structural structural model, we would find exactly the same thing! So why do a big deal about the difference between EFA and confirmatory factor analysis (CFA—the SEM approach to factor analysis)? Mathematics aside at the moment,25 the biggest difference is conceptual. In traditional EFA, in essence, we are saying: I do not know how these variables are related to each other. Let me throw them in the pot and see what comes out. That's why it's called exploratory, folks. In fact, when we don't know what's going on, it can be a very powerful tool that can help us understand the interrelationships in our data. The disadvantage is that we can end up with a set of factors that look good statistically but don't make much sense from a clinical or scientific perspective; that is, it may not be obvious why variables are grouped the way they do. The analysis of confirmatory factors, as is the case with all VARIANTS of SEM, is a model testing technique, rather than a model that builds one.26 Therefore, while we still get pages and output pages (resams, if we are not careful with what we ask), the action is really at the end, where we see how well the model actually fits the data. This is a point we have already mentioned and, as promised, we will continue to emphasize: changes in the model to better fit the data should be based on our theoretical understanding of the phenomenon we are studying, and not on moving arrows around to get the best goodness-of-fit index. We'll say a little more about CFA later in the chapter. Before going on to discuss the steps in SEM, some other points about the advantages of using latent variables (in addition to the aesthetics of having ovals in our drawings) need to be discussed. In our example of the three scales to measure Cheer, we said that none measure it exactly, but that the three together capture more or less what we want; in other words, we're doing a three-scale superscale. We can achieve the same goal using other techniques, but the route would be more roundabout. We would have to execute an EFA on these variables (and any other set of variables that we would like to combine); use the output of this analysis to calculate a factor score for each person; and then use that factor score in the next stage of the analysis. With SEM, all of this can be achieved in one step. TABLE 20–3 Correlations between three tests to measure Cheer A second advantage arises from the use of measurement theory. Every time we measure something, for example, whether it's blood pressure or pain, there's some mistake. The error can arise from a variety of sources: a person's blood pressure may change from time to time; the pressure gauge may not be perfectly calibrated; and the observer can round up or down to 5mm or make a mistake registering the number. When it comes to paper and pencil or observer testing, there are even more sources of error, such as biases in or concentration lapses. These errors result in a measure that is less than reliable. Then a problem arises when we correlate two or more variables: the observed correlation is less than what we would find if the tests were 1.0 reliability. Therefore, we will almost always underestimate the relationships between variables and, if the reliability is very attenuated, we can mistakenly conclude that there is no association when, in fact, one exists (i.e., we would be making a type II error). The solution is to discourage reliability, and to find out what the correlation would be if both tests were fully reliable.27 An important advantage of SEM is that this is done as part of the process. If we only had a scale to take advantage of some construction (that is, we would be dealing with a measured variable instead of a latent variable defined by two or more measures), sometimes we are later if we randomly split the scale in half and then built a latent variable defined by these two subscales. Next, we can calculate reliability28 and, based on this, unattended reliability.29 FIGURES 20-10 Loads of factors and uniqueness for the Cheer factor and its three measured variables. Let's see how powerful neglect can be when we're testing a theory. Let's say we're interested in the relationship between anxiety and the obligation to read each email message as soon as it appears on your computer screen. The usual way to test this theory is to give an inventory of anxiety and proof of email reading compulsion to a group of people, and see what the correlation is. But, because the researcher has diligently digested what we just said, she decides to give her 200 subjects two scales to take advantage of anxiety (A1 and A2), and two of the email habits (H1 and H2). What you find is shown in Table 20–4. The two anxiety scales are moderately correlated at 0.63, as are the two habit scales (0.64). To weigh it, however, correlations between anxiety and email habits correlate only between 0.51 and 0.59; not bad, but not the things of which great theories are made. However, after it creates a latent anxiety variable, as measured by the two A-scales, and correlates this with the latent variable of habits, exploited by its two H-scales, it finds that the correlation between traits is 0.86, and tenure (as well as shiny hair, finding their true love, and perpetual happiness) are guaranteed. In this case, the technique has calculated the reliability of the parallel shapes, used it to neglect the reliabilities and found what the correlation would be in the absence of a measurement error. TABLE 20–4 Correlations between two anxiety scales (A1 and A2) and two ccales of habits Email reading (H1 and H2) Now that we have a bit of background, let's move on to the steps involved in developing models of structural equations. We will follow the example of Schumacker and Lomax (1996) and divide the process into five steps30: Identifying the model specification Specifying the test fit respecification model A dress designer, model specification can mean, I want a woman who measures 5 x 8 height, weighs 123 pounds, and has a perfect figure of 36-24-36. For many men, it means indicating which supercharged engine to put in a sports car, and whether to get a hard or a rag top (to attract the model of the first definition). Here, however, we mean something much more mundane: explicitly indicating the theoretical model you want to test. We've already discussed a lot of this when we were looking at the road analysis. The same concepts apply, but we will now extend them to include both measured variables (which can be analyzed with route analysis) and latent variables (which cannot be). For example, all diagrams in Figure 20–6 can be drawn by replacing the measured variables with the latent ones. In addition, we can mix and match, using latent variables and measurements in the same diagram, as long as we don't do obviously ridiculous things, such as having two latent variables define a measure. Let's use these concepts to fully develop a model of success in porristas, shown in Figure 20–11. We have a latent variable, athletic ability, which is measured by the three observed variables IQ, Height and Jumping; and a second latent variable, Cheer, which is measured by our WOKS, SUCS and MPI scales. Based on what we said earlier,31 it would make sense to randomly divide CLAP/TRAP into two parallel shapes (CLAP and TRAP) and have them define the latent variable, CT. Because ODO is now an endogenous variable (it has arrows coming towards it from Cheer and Athletic Ability), we need to give it a term of disturbance, which we have called d1. This step of model construction is called the measurement model, because we are specifying how latent variables are measured. Some of the questions we could ask at this stage are: (1) How well do the variables observed actually measure the latent variable? (2) Are some observed variables better latent variable indexes than other variables? and (3) How reliable is each observed variable? (We'll look at these topics in more depth and look at this model when discussing some elaborate examples of SEM and CFA.) Identification When we were discussing route analysis, we said that the number of parameters can never be greater than the number of observations, and ideally they should be smaller. However, it is often the case that, once we start adding up all the variances and covariance in our pretty image, their number exceeds the limit of [k × (k + 1)]/2. The solution is to place constraints on some of the parameters. In the example we used earlier, we were able to solve the equation, of 5o b + c, specifying in advance (b) to 3; that is, we made b a fixed parameter. We could also have solved the equation if we had said that b á c, in which case both b and c would have to be 21/2. In this case, b and c are known as restricted Therefore, we have three types of parameters in structural models: (1) free parameters, which can assume any value and are estimated by the structural equation model; (2) fixed parameters, which are assigned a specific value in advance; and (3) restricted parameters, which are unknown (as they are free) but are limited to being the same as the values of one or more parameters. The SEM joy32 is figuring out how many parameters to leave as free and whether the remaining parameters should be fixed or constrained. FIGURE 20-11 The complete model for success in the porrist. Perhaps the easiest way to constrain parameters33 is simply not to draw a path. We said that MPI loaded into Cheer and (implicitly, not having a path) did not load into athletic ability. That's another way of saying that the parameter for the MPI path to the athletic skill is set to have a value of 0. Making the model as simple as possible is often the best way to avoid identification problems. A second method involves latent variables. For each variable not observed with a single-headed arrow emerging from it (such as the terms of error), we must set the path coefficient or variance of the error itself to some arbitrary value other than zero; otherwise, we would be trying to estimate b and c at the same time. But, because we don't actually measure (or observe) the term error, it doesn't make sense to assign a value to it, which leaves us setting the route coefficient. The easiest thing to do is to give it the value 1, and this is what is normally done. In any case, assigning a different value will not affect the overall fit of the model, only the estimate of the error variance.34 When a latent variable has two or more single-headed arrows from it (such as CLAP/TRAP, Cheer, and Athletic Ability), one of them must be set to equal 1 (that is, be a fixed parameter). It doesn't really matter which one you choose, but it's best to use the variable with the highest reliability, if this is known. Changing which measured variable has a coefficient of 1 will alter the non-standardized regressions of all variables related to that latent variable (because non-standardized regressions are relative to the fixed variable), but does not affect standardized regression weights. Note that we do not draw curved arrows between exogenous variables. This is because we assume that exogenous variables are correlated with themselves, and most programs, such as AMOS (Arbuckle, 1997), build it automatically. Finally, as mentioned above, there may be times when we believe that the error terms of two or more variables are identical (more or less some variation), and these become restricted parameters. This situation arises, for example, if we are giving the same test in two different; or we give it to mothers and fathers; or, as in the case of CLAP and TRAP, divide a single test in half. In programs like AMOS, we indicate this by assigning same name to mistakes. You might think that if you've had the hassle of counting the number of observations, simplifying the model, and restricting some of the parameters, you'll avoid identification problems. Such thinking is naive, and reflects a confidence in the inherent justice of the world that is rarely, if ever, justified. The steps we have just discussed are necessary, but not sufficient. Identification problems will almost certainly arise if you have a non-recursive model (i.e. one that has reciprocal relationships), which is another reason to avoid them whenever possible. You may also have difficulties if the range of the array35 you are analyzing is less than the number of variables. This can occur if a variable can be predicted from one or more variables, which means that the row and column of a correlation matrix that represents that variable are not unique. For example, the verbal IQ score of some intelligence tests is derived by adding the scores of a series of subscales (six in the Wechsler test family). If you include the six subscale scores as well as the verbal IQ score in your data matrix, then you will have problems, as one variable can be predicted by the sum of the others; that is, it has seven variables (the six subscales plus the verbal IQ) but only six unique ones, resulting in an array whose range is less than the number of variables. Even high correlations between variables (for example, between height and weight) can cause problems. Often, there is no easy way to know in advance if you are going to have problems with your4th section of the model. You simply do your best to set up the model, pray hard and wait for the program to run. If the output says that the model needs more constraints, then you have to go back to your theory and determine if there is any justification in, for example, restricting two variables to have equal variances. You may suspect that they may be the same scale measured at two different times, or two different scales touching the same construction. If this doesn't help, you may want to consider a different career. Estimation Now that you have specified the model, the time has come to estimate all those parameters. The easy work is all that matrix algebra: invert arrays, transform arrays, pre- and post-multiply arrays, and so on. Because it's easy, the computer does it for us. What we have left is the hard part: deciding the best method to use. This requires brain cells; a commodity that is scarce within the computer.36 The fact that there are a number of different techniques must act as a warning, which we have found in other contexts. If there was an approach that was clearly superior, then the law of the statistical jungle37 would dictate that it would survive, and all the other would exist only as historical footnotes. The continued survival of all of them, and the constant introduction of new ones, indicates there is no clearly superior solution; therefore, the unfortunate need to think about what we are doing. The least unweighted squares (ULS) method of estimating parameters has the clear advantage that it makes no assumptions about the underlying distribution of variables. However, it depends on the scale; that is, if one or more of the indexes are transformed to a different scale, the parameter estimates will change. In most of the work we do, the scales are totally arbitrary; even height can be measured in inches or centimeters (or centimeters, if you live in the UK or one of the colonial backs.38 This means that the same study conducted in Canada and the United States can come with different results, simply because of different measurement scales used. This becomes another problem when we use paper and pencil tests that don't have significant scales. This is an unfortunate property, which is one of the reasons why ULS is rarely used. Weighted least squares (WLS) are also free of distribution and do not require multivariate normality, but require a very large sample size (usually three times the number of subjects you can enroll in) to work well. Many programs have the maximum probability (ML) method of estimating parameters by default. This works well as long as the variables are multivariate, normal, and constitute interval or relationship data. But, if the data is extremely sesewn or scaled or scaled ordinationally, then the results of the ML solution are suspicious. So, which one do you use? If you have access to the SEM program called LISREL8, and your front-end program, PRELIS2, then use them. They will calculate the correct array, depending on the types of variables you have. If you use one of the other programs (for example, EQS, PROC CALIS Stata or AMOS), then it may be worth running the model with some different types of estimators. If all the results are consistent, then you can be relatively sure about what you have found; If they differ, you'll need to go back and take a closer look at the data to see if you're not normal, asymmetry, or some other problem. Then try to fix it (for example, with transformations) or choose the method that best suits the type of data you have. Adjustment Test In the previous section, in estimation procedures, we regret the fact that there were so many different approaches. While that is undoubtedly true, the problem pales in insignificance compared to the plettive of statistics used to estimate goodness of adjustment. We have already mentioned 2GoF in the context of route analysis, and that test is also used in SEM. It has a clear advantage in that, unlike all other GoF indexes, has a test of importance associated with it. A general rule of mind for a good adjustment is that the number 2GoF is not significant and that the number 2GoF/df must be less than two. Unfortunately, as we mentioned when we were discussing the route analysis, it is very sensitive to sample size and Normal. Most of the other indexes we will discuss39 are scaled to take values between 0 (no adjustment) and 1 (perfect fit), although what is considered a good fit is often arbitrary. Typically, a value of .90 is the minimum accepted value, but as we just mentioned, there are no odds associated with these tests. Let's review some of the most common and useful ones to show the types of indexes available, without trying to be exhaustive (and exhausting). A statistics class is called comparative adjustment indexes, because they test the model against some other model. The most commonly used (though not necessarily the best) index is the Normed Fit Index (NFI; Bentler and Bonett, 1980), which proves, if the model is different from the null hypothesis, that all variables are independent of each other (in statistical jargon, that covariance is all zero).40 Takes the form of: For when we discuss multiple regression, we said that the value of R2 increases with each predictor variable we add. There is a similar problem for NFI; improves as we add parameters. Just as the adjusted R2 penalizes us for adding variables, normed Fit Index 2 (NFI2, also called Incremental Fit Index, or IFI) penalizes us for adding parameters: Another statistic of such parsimony (you are rewarded for being cheap when it comes to the number of parameters) is the comparative adjustment index (CFI). Today, a more popular variant of this is the non-normed adjustment index (NFI) , commonly called the Tucker-Lewis Index (TLI): For any given model, a ratio of less than 2 of implies a better fit. Both the IFC and the TLI depend on the mean correlation between the variables. Sometimes the IFC and TLI exceed 1, in which case they are set equal to 1. The CFI and TLI are highly correlated with each other, meaning that only one must be reported. When IFC is less than 1, it is always larger than TLI, which may tempt some to prefer it, but the movement is toward using TLI. Other indices resemble R2 in that they represent the variance ratio in the covariance matrix accounted for by the model. One of these indexes is the Goodness of Adjustment Index (GFI); Fortunately for you, its formula involves a lot of matrix algebra, so we won't bother to show it.41 The Adjusted GFI (AGFI) is still a parsimony adjustment index that we can find through. There are a number of other indexes, which all of which decrease the AGI value proportionally to the number of parameters it uses. Another widely used index is the Akaike Information Criterion (AIC), which is unusual in that smaller (closer to 0) is better42: A slight variant of it is called consistent AIC, or CAIC: Because no one knows what a good AIC or CAIC value should be, these indexes are most often used to compare models , to choose from two different models of the same data. The one with the smallest AIC or CAIC is better; however, there is no statistical evidence for indices of the difference between AIC or CAICs, so we can't tell if one model is statistically better or insignificantly better than the other. A more recent index, increasingly used now, is the Half-Root Approach Square Error (RMSEA, often called Ramsey in a casual conversation). Unlike previous indices, it takes into account both df and sample size (N): Like AIC and CAIC, it follows the criterion of anorectic fashion models that are less of a better. As with most of these indexes, there is no probability level associated with it. The guidelines are that a value of 0.01 is excellent fit, 0.05 is good, and 0.08 is mediocre. (Why they seems to be? As Fiddler's song on the Roof says, tradition – nothing else.) The problem is that, as with all parameters, RMSEA is only an estimate, and when the sample size and df are low, the confidence interval can be quite wide. One even more recent than RMSEA is the Standardized Medium Root Square Residue (SRMR; yes, each new dawn announces a new index). It is the standardized difference between the observed correlations between variables and reproduced correlations (remember them from our discussion on route analysis?). The criteria for evaluating it are generally similar to those of the RMSEA. It is not a parsimony index, so it improves (smaller) as the number of parameters in the model increases. In fact, it's not unusual to find values of zero. Life is easy when all adjustment rates tell us the same thing. What do we do when they disagree? The most common situation occurs when we get high values (that is, more than .90) for GFI or NFI2 and a low value for RMSEA (below .05), but 2GoF is significant. Unfortunately, here we have to use a little judgment; Can it be the significant 2GoF due to too much power? If so, it's probably best to trust one of the other indexes. If there are approximately 10 subjects per parameter, then we should look at a series of indexes. If everyone indicates a good fit, then go with that. But, if 2GoF is significant, and the other indices disagree with each other, then we have an adjustment that is marginal, at best, and whether or not publishing depends on your level of desperation for another article. Respecification Respecification is an elegant term to play with the model for a better fit with the data. If you have been paying attention, you should realize that statistical tests play a secondary role in this; the main role should be your understanding of the area, based on theoretical and empirical considerations. Keep singing this mantra yourself as you read this section. You can improve the fit in one of two ways: remove paths that are not significant, or add paths (or variables). There are no computer packages that give you a bronx joy and say, Fool,43 you forgot to include the whole program. All your so-called colleagues will be very happy to carry out this but only when it's too late, and you're presenting the results at an international conference. The statistical tests that exist are for the opposite type of misconfiguration errors: those due to variables that do not belong to the model or have paths that lead to the wrong endogenous or exogenous variables. The easiest way to detect them is to look at the parameters. First, they must have the expected signal. If a parameter is positive and the theory says it must be negative, then something is terribly wrong with the model. The next step is to look at the importance of parameters. As we said, all parameters have standard errors associated with them, and the parameter relationship to their standard error forms a t or z test. If the test is not significant, that parameter is likely set to 0. Of course, this assumes that it has a sufficient sample size, so you are not confirming a type II error. All major SEM programs, such as LISREL, CALIS (part of SAS), AMOS, Stata, and EQS, can examine the effects of the release parameters you have corrected (the Lagrange multiplier test) and the removal of model parameters (the Wald statistic). These statistical tests should be used with caution. Whether to follow his advice should be based on your theory and reason; otherwise, you may end up with a model that fits the current data very well, but makes little sense and may not be replicable. Now that we've given you the basics, let's run a couple of examples. A confirmatory factor analysis Assume that we have seven measured variables that we postulate reflect two latent variables: a1 to a4 are associated with the latent variable f1, and b1 to b3 with the latent variable f2. We start by drawing a diagram of our model (if we are using a program like AMOS, Stata or EQS), shown in Figure 20–12. The program is relatively intelligent,44 so it automatically corrected the parameters of all error terms to variables measured to be 1. For each of the endogenous variables, it also sets the path parameter for a measured variable to 1. We didn't like the choice the program made, so we surpassed it and selected the variable in each set which, based on previous research, has the highest reliability. After right-clicking,45 we get the diagram shown in Figure 20–13 and reams of output, which we summarized in Table 20–5. First, what do all these little numbers in Figure 20–13 mean? Those on the arrows must be familiar; are route coefficients or standardized regression weights (either term will work), which are equivalent to factor loads in EFA. numbers on rectangles are multiple squared correlations, which are equivalent to community estimates in EFA. There are two other things this figure tells us. First, we work when it comes to variable a4; that's what really seem to be caused by the latent variable f1. Note that unlike EFA, we are not told whether it charges more in factor f2 or is not loaded into any of the factors; again, that's because we're testing a model, not trying to develop one. The second fact is that factors f1 and f2 are probably not correlated; the correlation coefficient is only 0.06. FIGURE 20-12 Input diagram for a confirmatory factor analysis. FIGURE 20-13 Output diagram based on Figure 20-12. Now let's move on to the output, shown in Table 20–5 and see what we can learn. First, there are 28 sample moments in other words (i.e. English), 28 observations. This is based on the fact that there are seven measured variables, so there are (7 × 8) / 2 to 28 observations. Our model specifies 15 parameters to estimate: five regression weights (two others are not estimated because we have set them to 1); one covariance (between f1 and f2); and the variances of the seven terms of error (or disturbance) and two latent variables.46 There are 13 degrees of freedom, which is the difference between the number of observations and the number of parameters. The next output block tells us that the 2GoF is 16,392, which, based on 13 degrees of freedom, has a p-level of .229. Therefore, despite the fact that variable a4 does not work too well, the model as a whole fits the data quite well. Next, we see the non-standardized and standardized regression weights. The non-standardized weights for a1 and b1 are 1.00, which is encouraging, because we set them to be equal to 1. The other five weights have standard errors associated with them, and the relationship between weight and SE is the critical relationship (CR) that is interpreted as a z-test. All of them are significant (with or more than 1.96) except a4, further confirming that it is not specified correctly. Similarly, covariance (and therefore correlation) between f1 and f2 is low and has a CR of only 0.551; that is, the two variables or latent factors are not correlated. The following sets of numbers show the variances we are estimating and the multiple correlations squared (also shown in the figure). Because we ask for them, we also get the modification rates (MI), which tell us how much the model could be improved if we specify additional paths. The largest, which appears in Regression Weights, is a3↔b2. That means that if we draw a path from b2 to a3, our fit would improve. In fact, the path coefficient between the two is 0.123.47 and 2GoF (based now on df 12, because we specify another route) falls to 10.028.48 But, because there is no theoretical reason for this route (or for the other proposed modifications), we will simply ignore them. Finally, we have given only a few of the countless GoF indices. The model represents perfection (both parameters and observations, which means there is nothing more to estimate), and the independence model is the opposite (assuming nothing correlates with anything). Fortunately, our model is perfectly; all indices are more than 0.90. Our model would fit even better if we dropped the variable a4. Again, this must be dictated by theory. If we believe that the variable is substantive, and that the non-negligible route coefficient may be due to a sampling error or a small sample size, we would maintain it; otherwise, in the trash can goes. Comparison of two-factor analysis Sometimes we are in a position where we want to compare two factor structures; for example, are the results for patients and controls or for men and women alike49 and, if not, how do they differ? This can be done with EFA, but it is difficult, and methods of comparing factor structures leave much to be desired. However, it is relatively easy to do so with CFA.50 If we do not have any hypotheses in advance regarding the structure of the factor, we can start by running an EFA with a group and then use the results to set the parameter estimates in a CFA for the second group. On the contrary, if we have any idea what the structure should look like, we can specify it for both groups and see where it fits and does not fit for each one. TABLE 20–5 Output selected for confirmatory factor analysis As an example, we will keep the problem presented in Figure 20–13 and assume that we draw another sample, one in which a4 actually loads in f1 but we do not know in advance. Instead of setting only one of the routes of the latent variables to the measurements, we will put in the non-standardized regression weights of the first sample and again, based on our previous results, we will indicate that the covariance between f1 and f2 is 0. Production will be very similar to that of Table 20-5, with some notable exceptions. First, the number of parameters to be estimated decreases from 15 to 9, because we have corrected 6 additional parameters: 5 routes of latent variables plus covariance. Second, there will be no standard errors and critical proportions given for these 6 parameters, as we are not estimating them. The 2GoF, now based on the df 19, is a huge 144.47, indicating that the model does not fit the data worth a plugged-in nickel. If you look at modification rates, larger ones involve a4 and e4 in several ways, such as suggesting that we include covariance terms between e4 and f1, or paths between a4 and the other three variables associated with f1. None of them make sense theoretically, but they all point to a bad specification involving a4. We would go back to our theory and hypothesize that the route coefficient, which we set at 0.20 to be consistent with the results of the first sample, is wrong and perhaps should be closer to 0.80. Alternatively, we can release him and see what the program does with it. Please note (once again) that our use of is tempered by our knowledge and theory. If the model actually fits the data, then we can conclude that the factor loads we found for a group would also fit the second group. In this case, case, above the suede and make the comparison stricter: are the variances of the error terms similar between the samples? This type of analysis is very useful for determining the equivalence of questionnaires in different topic groups. A complete SEM model Now let's go back to the full model of hit in porristas, shown in Figure 20–11, and add what we've learned. First, we need to set all the paths that lead to the various terms of disturbance to 1, and set a path of each latent variable to be 1. Second, since CLAP and TRAP are random halves of the same test, it makes sense to assume that their variances are similar. We indicate the fact that we have restricted these terms by giving variances the same name. The results of all this fixation and restriction are shown in Figure 20–14; the terms vct on the terms of disturbance for CLAP and TRAP tell the program that these variances should be the same. This diagram now forms the input to the program, which should run as long as the variable names of the rectangles correspond to the variable names in our data file. The program output is shown in Figure 20–15. For starters, the 2GoF is 38,225 which, based on 19 degrees of freedom, is very significant (p. . . 006). The other GoF indices are unequivocal: GFI and NFI are slightly above the cut-off point of 0.90, while AGFI is only 0.831. All this leads us to believe that the model could withstand a little improvement, but where? Let's start with the measurement aspect of the model: how well are we measuring the latent variables of Athletic Skill, Animation, and CLAP/TRAP? AGFI - Adjusted GFI; CR - critical relationship; GFI - goodness-of-fit index; MI - modification rate; NFI - Normed Fit Index; SE - standard error. FIGURE 20-14 Figure 20-11 with restricted parameters. FIGURE 20-15 Output based on Figure 20-14. The answer seems to be, not too bad, thank you, but maybe we can do better with Skill and Animation. If you look at modification rates, most of them don't make much sense from the perspective of our theory, but one has a closer look: the suggestion to add a covariance between ea2 and ea3. Because Cheer as a whole seems to add little to the image, let's put it aside for now and rerun the model by adding e2 ↔ e3. Rewardingly, the value of 2GoF (df - 18) falls to 22,557, which has an associated p-value of .208. Because this model is a subset of the original,51, the difference between their respective 2s is distributed in itself as a 2. Therefore, if we subtract the 2s and dfs, we get 2 (1) to 15,668, which means that there was a significant improvement in the goodness of the adjustment. This is also reflected in an increase in the path coefficient of The capacity to CLAP/TRAP, 0.73; a drop in the Cheer coefficient (from 0.02 to 0.01); and the fact that the other adjustment rates are in an acceptable range. Finally, since Cheer doesn't help,52 we can have a simpler model if we just drop it. Although the change in the significant this time, all parsimony-adjusted GoF indices increase. Moreover, from a research perspective, it means that we don't have to administer these three tests to everyone, which at least makes us more cheerful. REPORT INSTRUCTIONS SeM and CFA programs produce production reses and reams. This, in turn, has resulted in remnants and remnants of guidelines for reporting results. We have worked powerfully on your behalf, reducing them to a manageable size. This cannot be considered plagiarism, because we follow the pattern to avoid charges for it: we steal from many sources, including McDonald and Ho (2002), Raykov et al. (1991) and Schreiber et al. (2006). The most useful way to report is the model diagram, with standardized regression weights and correlations near the arrows. You should also indicate which ones are statistically significant. In the text, there must be a theoretical or research-based justification for each arrow in the diagram, both directional (one arrowhead) and non-directional (curved arrows with two heads). There are many programs for SEM and CFA, and they differ from their methods. This means that, unlike techniques like ANOVA or EFA, different programs can produce different results. Therefore, you must say which program you have used. You must indicate what type of matrix was analyzed (covariance, media covariance, correlation, etc.). If you don't know, you shouldn't use these techniques. Never forget that CFA and SEM are great sample techniques. You should report not only the sample size, but also the proportion of subjects to parameters (this implies that you know how many parameters you are estimating). How did you deal with the missing outliers and data: did you leave those cases, impose values, or use any technique that can handle these anomalies? What method did you use for parameter estimation (maximum probability, generalized square minimums, weighted least squares, etc.) and why? The model fits. As we said, there are many indices of this, but we recommend at least: (a) -2GoF and its df; (b) the Tucker-Lewis Index (TLI, also known as NNFI); (c) THE RMSEA; (d) SRMSR; and (e) the NFI. Indicate whether the model has changed based on the modification rates. If you have, indicate what was changed and the theoretical justification for it, as well as how much the adjustment changes improved. Only gold members can continue reading. Sign in or sign up to continue