believe best measure self-esteem and locus of control and subject them to confirmatory factor analysis. First use SPSS (or another general statistical program) to create a matrix for analysis in Amos (or one of the other programs). Then analyze your model using this matrix. I recommend using the matrix for analysis in order to temporarily avoid dealing with missing data in Amos.

3. The files "DAS 5–8 simulated 6.sav" and "DAS 5–8 simulated 6.xls" include 500 cases of simulated data for the DAS-II.
   a. Conduct the first-order factor analyses from this chapter using the simulated data. Interpret the findings. How do the results compare with those in this chapter (and in Exercise 1)? Would you come to different conclusions following these analyses than we did in the chapter?
   b. Note the fit indexes. Which changed the most from the analyses in the chapter? Why do you think this may be?
   c. As you examine your analyses, are any other hypotheses or models suggested by the findings? If so, conduct these analyses and interpret the findings.

Notes

1 In higher-order intelligence models, Heywood cases often show up in connection with Fluid Reasoning factors (Gf, in the DAS-II represented by the Nonverbal Reasoning factor). When this happens, the *g* to Gf path may approach or exceed 1 and the associated unique factor variance become negative. Note in Figure 15.11 that the *g* to Nonverbal Reasoning loading approached 1. One implication of such a finding is that *g* and Gf factors are not separable. Some researchers use this not-uncommon finding to argue that the Gf factor is redundant with *g*, whereas others argue that this shows that *g* is redundant. As noted, one common method for dealing with negative variances is to set the value to zero. This makes sense if the value is fairly close to zero but is less defensible if it is a large negative value (which likely indicates problems with the model). There are also other possible ways to deal with negative variances, including constraining the value to be positive.

2 Here is an interesting conundrum. When factors are correlated, it is possible (although not desirable) to have factors referenced by only two measured variables each. So, for example, a correlated two-factor, four-measured variable model would have one degree of freedom. But when factors are uncorrelated, each factor requires a minimum of three measured variables for identification, and with three measured variables each factor is just-identified (as in the present bifactor example). That means that if a bifactor model includes fewer than three variables for a factor, the researchers will need to either make additional constraints (e.g., constrain the two factor loadings to be equal) or, counter-intuitively, relax constraints (e.g., allow that factor to be correlated with another factor). As you are reading research using the bifactor model and you notice only two measured variables on a factor, make sure the researchers tell you what they have done to solve this problem! This conundrum of identification also occasionally leads to a phenomenon known as "empirical underidentification" in which a model allows factors to be correlated, but that correlation is small and nonsignificant. If one of the offending factors involves fewer than three measured variables, it will thus be underidentified. The phenomenon of empirical underidentification applies to first-order factor models as well (Kenny, 1979).

16 Putting It All Together
Introduction to Latent Variable SEM

Let's review our progress in our adventures beyond MR. You know how to conduct path analysis using MR. This experience includes the estimation of standardized and unstandardized paths, the calculation of disturbances ($\sqrt{1-R^2}$), and the calculation and comparison of direct, indirect, and total effects using two different methods. We transitioned into estimating path models using Amos and other SEM programs and focused again on the estimation of both standardized and unstandardized effects and direct, indirect, and total effects. With Amos, we switched from the estimation of the paths from disturbances to estimating the variances of the disturbances, although either is possible. We have defined just-identified, overidentified, and underidentified models, and I suggested that you use a SEM program to estimate overidentified models but use either MR or an SEM program if your models are just-identified. We have examined fit indexes for overidentified models and have highlighted a few that are useful for evaluating a single model and those that are useful for comparing competing models. We briefly focused on equivalent models, nonrecursive models, and longitudinal data. We focused on the effects of measurement error on path analysis, MR, nonexperimental research, and research in general and began considering the use of latent variables as a method of obviating this threat. We expanded our knowledge of latent variables, their meaning, and estimation via confirmatory factor analysis.

PUTTING THE PIECES TOGETHER

In this chapter, we will begin putting all these pieces together in latent variable structural equation modeling. As noted in Chapter 14, you can consider latent variable SEM as a

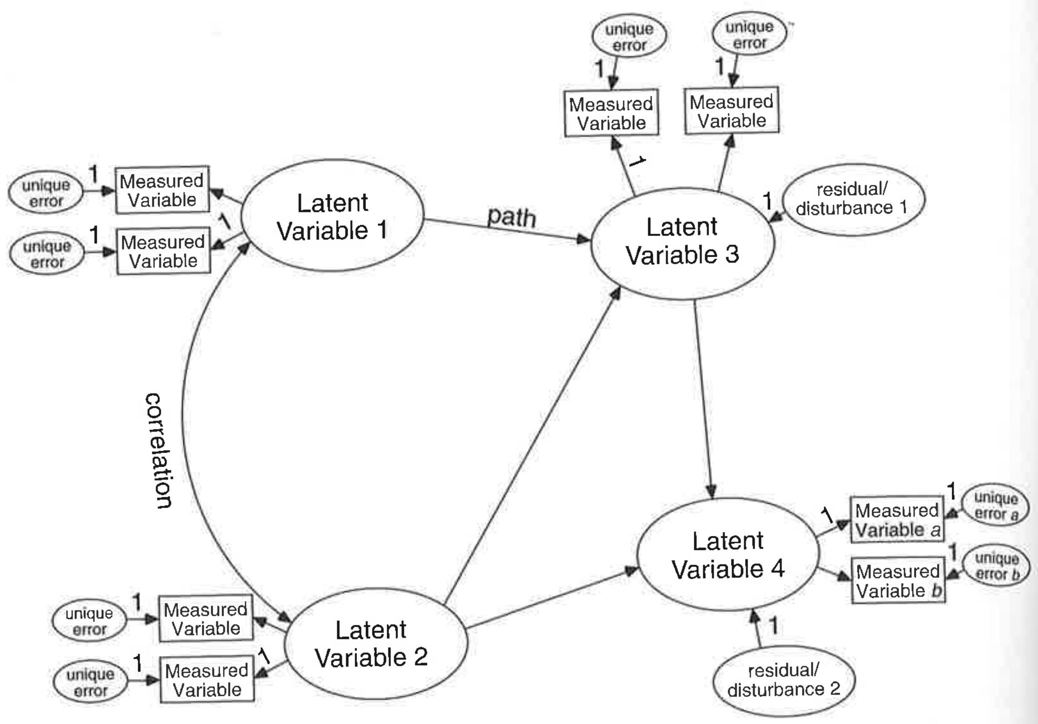


Figure 16.1 Full latent variable SEM model

confirmatory factor analysis of the constructs involved in the research project, along with a path analysis of the effects of these constructs on each other. For this reason, many writers refer to these as the measurement model and the structural model, respectively (e.g., Mulaik & Millsap, 2000), to denote the conceptual distinctions between components of latent variable SEMs. Although this separation of measurement and structural portions is not necessary statistically, it can be very useful conceptually, especially at this stage of learning.

Figure 16.1 displays, for review, the components of a latent variable SEM. The measurement model consists of the estimation of the four latent variables from eight measured variables. The structural model consists of four paths and one correlation among the four latent variables. Note that each variable that has a path pointing to it also has a residual–disturbance–error term pointing to it, representing all other influences on the variable other than the variables pointing to it. Some of these residuals represent the unique and error variances of measured variables, the remaining influences on these measured variables other than the latent variable underlying it. Some residuals represent disturbance terms for latent variables, meaning all remaining influences on these latent variables other than the other latent variables. Although I refer to some of these as unique–error variances and others as disturbances, the terms error and residual are used fairly interchangeably.

Why, you may wonder, doesn't Latent Variable 1 have a disturbance pointing to it? Because Latent Variable 1 has no paths pointing to it; it is exogenous. Note also that each latent variable (including the unique–error variances and the disturbances) has its scale set by fixing a single path from it to another variable to 1. So, for example, the latent variable labeled residual/disturbance 2 has its scale set to the same value as the latent variable labeled Latent Variable 4, which in turn is set to the same value as Measured Variable *a*. Note that the biggest difference between this model and the CFA models from the last chapter is that some correlations among latent variables are replaced by paths. As a result, the latent variables

with paths pointing to them also have disturbances pointing to them. Of course, this is akin to the difference between a correlation matrix of variables and a path model specifying that one variable influences another. Take some time studying the model to make sure you understand it.

AN EXAMPLE: EFFECTS OF PEER REJECTION

Overview, Data, and Model

Eric Buhs and Gary Ladd used SEM to examine the effects of peer rejection on Kindergarten students' academic and emotional adjustment (2001). A portion of the model they analyzed is shown in Figure 16.2. The latent variables in the model, along with the measured variables used to estimate them, were these:

1. Rejection was indexed by averaged sociometric ratings for each child by the other children in the class (Averaged Rating; the scale of this variable was reversed to make it consistent with the negative [Rejection] name of the latent variable) and by the number of times each child was nominated negatively (as someone other children did not want to play with; Negative Nominations).
2. Change, from a previous rating, in Classroom Participation. This variable was estimated from teacher ratings of Cooperative Participation (e.g., accepts responsibility) and Autonomous Participation (e.g., self-directive).

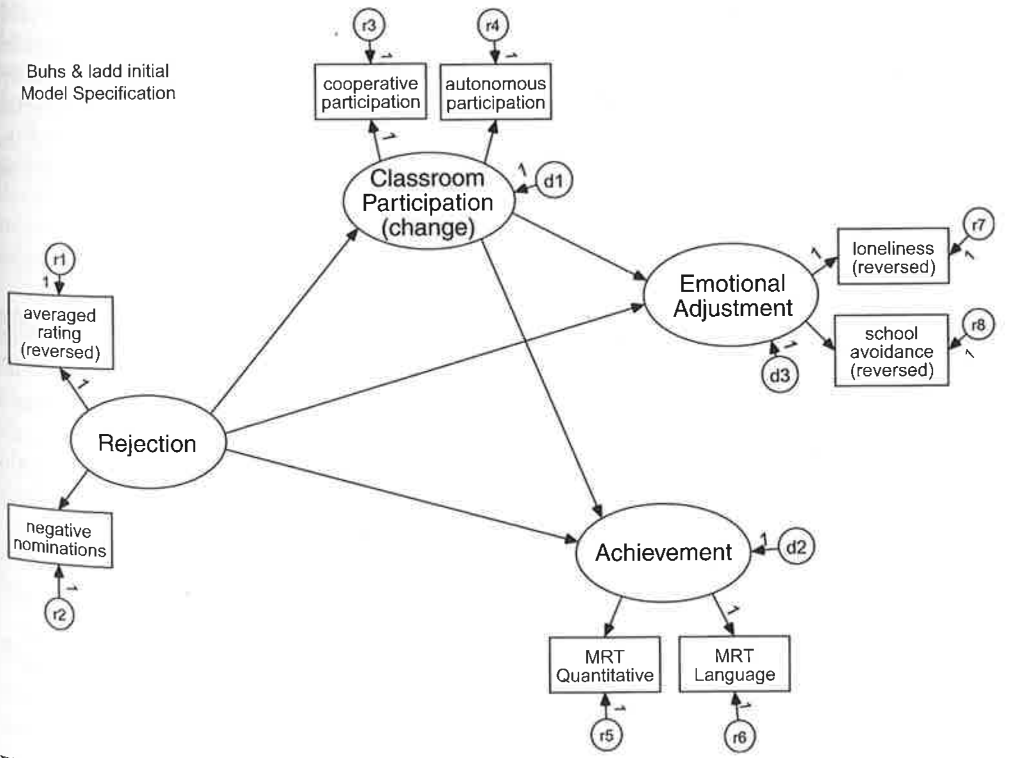


Figure 16.2 Effects of peer rejection on Academic and Emotional Adjustment, initial model. The model was derived from Buhs and Ladd, 2001.

3. Achievement, which the authors considered one aspect of adjustment, was estimated from the Language and Quantitative subtests from a standardized school readiness test (the Metropolitan Readiness Test, Nurss & McGauvran, 1986).
4. Emotional Adjustment, as indexed by self-ratings of students' Loneliness at school and their desire to avoid school (School Avoidance). These two variables were reversed to make the latent variable consistent with the positive name (Adjustment).

Buhs and Ladd's article included an additional intervening variable (Negative Peer Treatment) and an additional indicator of Rejection. These variables were not included here to simplify the model. The model is longitudinal; the Rejection variables were collected in the fall, the other variables in the spring (for more detail, see Buhs & Ladd, 2001).

Recall that with our earlier path models (e.g., the homework models in Chapter 13) many of the variables in the model were composites (e.g., Achievement was a composite of four scores). Buhs and Ladd (2001) could have done the same thing here, but instead of adding Quantitative and Language into an achievement *composite* variable, for example, the authors used these two measures as indicators of an Achievement latent variable. Recall our discussion in Part 1 about multiple regression predicting an outcome variable from an *optimally weighted combination* of the independent variables. Conceptually, the latent variables in SEM are similar: they are optimally weighted combinations of the measured variables.

The model will be estimated from the *measured* variables. A portion of the data is shown in Table 16.1 (and is saved as data files on the Web site under the label "buhs & ladd data.sav" and "buhs & ladd data.xls"). Note there are no variables in the data file corresponding to the latent variables. This is because the latent variables, or factors, are estimated from the measured variables. If this is still confusing, think of the latent variables as *imaginary variables* that we estimate from the measured variables. (In the actual data file, the variable names are shortened versions of the variable labels used in the table and the Amos model, but they should be self-explanatory. Note that the data included here and on the Web site are not the actual data but rather simulated data created to be consistent with the correlation matrix, means, and standard deviations reported in the actual article. $N = 399$. Three of the measured variables were reversed to make them consistent with the variable names and thus more easily interpretable.)

Table 16.1 Sample Data: Measured Variables for the Peer Rejection Example

Child	*Averaged Rating*	*Negative Nominations*	*Cooperative Participation*	*Autonomous Participation*	*Quantitative*	*Language*	*Loneliness*	*School Avoidance*
1	−1.33	−1.09	1.19	.69	7.47	6.30	2.09	2.48
2	1.32	.55	−.13	−.07	2.72	2.76	1.42	2.16
3	−.64	−1.09	−.29	−1.26	6.40	5.39	1.59	1.38
4	1.42	−.36	−.19	−.56	.99	1.05	.94	2.13
5	.58	−.01	−.36	−.13	2.80	3.56	.36	2.20
6	−1.20	−1.51	.04	.07	7.07	7.79	1.37	2.03
7	.42	.39	−.25	.40	3.68	3.47	2.08	3.00
8	−.40	−.81	.78	1.03	7.03	4.94	2.03	2.61
9	1.99	1.89	−.45	−.66	1.51	5.08	.66	.99

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
ave_rat	399	-2.33	3.10	.1200	.95000	.101	.122	.011	.244
neg_nom	399	-2.37	2.64	-.1000	.90000	.186	.122	-.140	.244
coop	399	-1.38	1.88	.0000	.60000	.145	.122	-.256	.244
auto	399	-1.71	1.96	.0000	.62000	.135	.122	.043	.244
quant	399	-.86	11.91	5.3800	1.98000	-.150	.122	.472	.244
lang	399	.51	10.47	5.3600	1.78000	-.010	.122	-.302	.244
lone	399	-.02	3.12	1.5100	.56000	-.047	.122	-.182	.244
schavoid	399	.12	4.11	2.0500	.67000	.133	.122	-.128	.244
Valid N (listwise)	399								

Figure 16.3 Descriptive statistics for the simulated rejection data.

Just because our analyses have gotten more sophisticated does not mean we should ignore the mandate from Part 1: Always, always, always, always, always, always check your data prior to conducting analyses! This command is just as important—maybe even more so—as our analyses become more complex. So before conducting the SEMs here, make sure you check means, *SDs*, minimums and maximums of the variables in this file. As we conduct SEM, you should also get in the habit of examining skew and kurtosis. Note that with the current data, few of the measured variables had meaningful scales, and many had both positive and negative values. The averaged ratings, for example, were standardized within classroom. The descriptive data are shown in Figure 16.3.

For the current model, I have symbolized the unique–error variances of the measured variables as r1 through r8 and the disturbances of the latent variables as d1 through d3. Recall that we can consider the unique–error variances as all other influences on the measured variables beyond the influence of the latent variable, just as the disturbances are all other influences on a latent variable beyond those of the other latent variables.

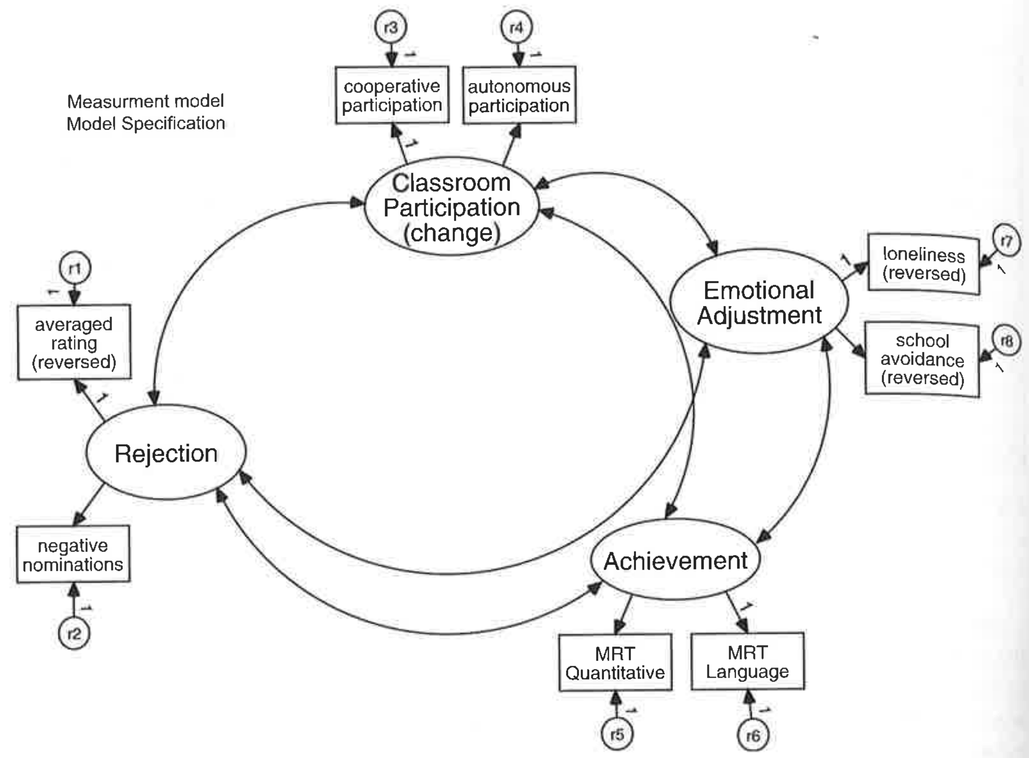
Measurement Model

For the sake of clarity, the measurement model, without the structural model, is shown in Figure 16.4. Except for its placement of variables (in a circular fashion instead of in a line), the model is similar to the confirmatory factor models from the last chapter. The model simply delineates the estimation of the four latent variables (Rejection, Adjustment, etc.) from the eight measured variables (Averaged Rating, Negative Nominations, etc.).

Note that each latent variable had its scale set by a single factor loading (path from the latent to measured variable) set to 1. Each error–unique (residual) variable had its scale set by setting the path from it to its corresponding measured variable to 1.

Structural Model

The structural portion of the model is shown in Figure 16.5, a figural representation of the hypotheses of the effects of one latent variable on another, and includes the disturbances for the endogenous latent variables in the model. The model examines the effect



Figures 16.4 Measurement model portion of the initial peer rejection model.

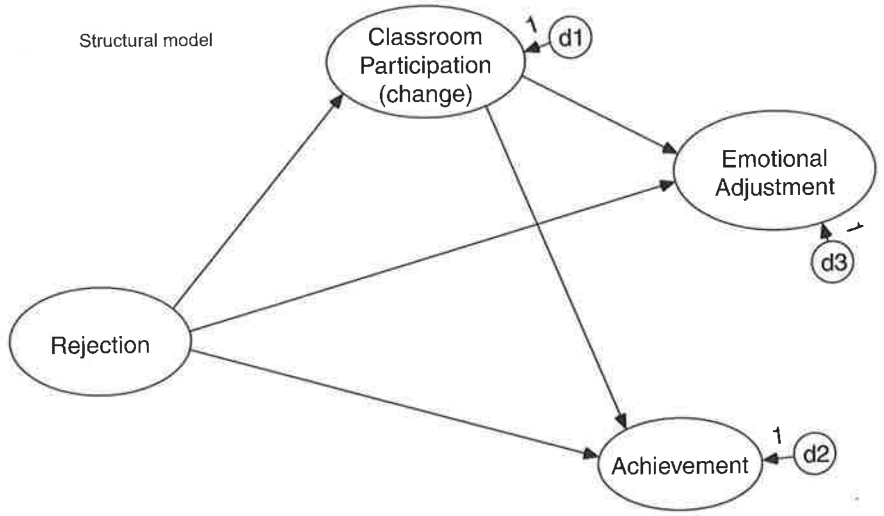


Figure 16.5 Structural model portion of the initial peer rejection model.

of Rejection on Adjustment, both directly and indirectly, through the class participation of the students.

The full SEM model (Figure 16.2) has 15 degrees of freedom. Fourteen degrees of freedom are from the measurement portion of the model (Figure 16.4). Note that all the factor loadings that could be included in the model (e.g., a path from Rejection to Cooperative Participation or Loneliness) are not included; these constraints are the source of this 14 *df*. The structural model (Figure 16.5) includes one additional *df*, resulting from the omission of a path between Achievement and Adjustment. The model is saved on the Web site (www.tzkeith.com) in the file "Buhs & Ladd model 1.amw." Note 1 at the end of the chapter shows the calculation of the degrees of freedom.[1]

Results: The Initial Model

The model (Figure 16.2) was analyzed using the raw data (Table 16.1 and the file "buhs & ladd data.sav" or "buhs & ladd data.xls") via Amos. Figure 16.6 shows relevant fit indexes, along with the standardized output. The model shows an adequate, but not good, fit to the data. The RMSEA was above .05 (.067, 90% confidence interval = .043 to .092), but was below .08. The SRMR was below the cutoff of .08 or .06 (.046). The CFI was above .95, but the TLI was below our informal cutoff for a good fit of .95. Although not shown in the figure the χ^2 was also statistically significant ($p < .01$), further suggesting a lack of fit. Again, the model shows an adequate, but not good, fit. The full array of fit indexes is shown in Figure 16.7. Because the model had an adequate fit, we'll first interpret these results. Later in the chapter we'll take a look at the more detailed fit information and consider how the model might be modified.

Figure 16.8 shows more detail concerning the paths and factor loadings, including the unstandardized coefficients, their standard errors, and critical ranges (*z* statistics). All the parameters that were estimated were statistically significant (*z* greater than approximately 2).

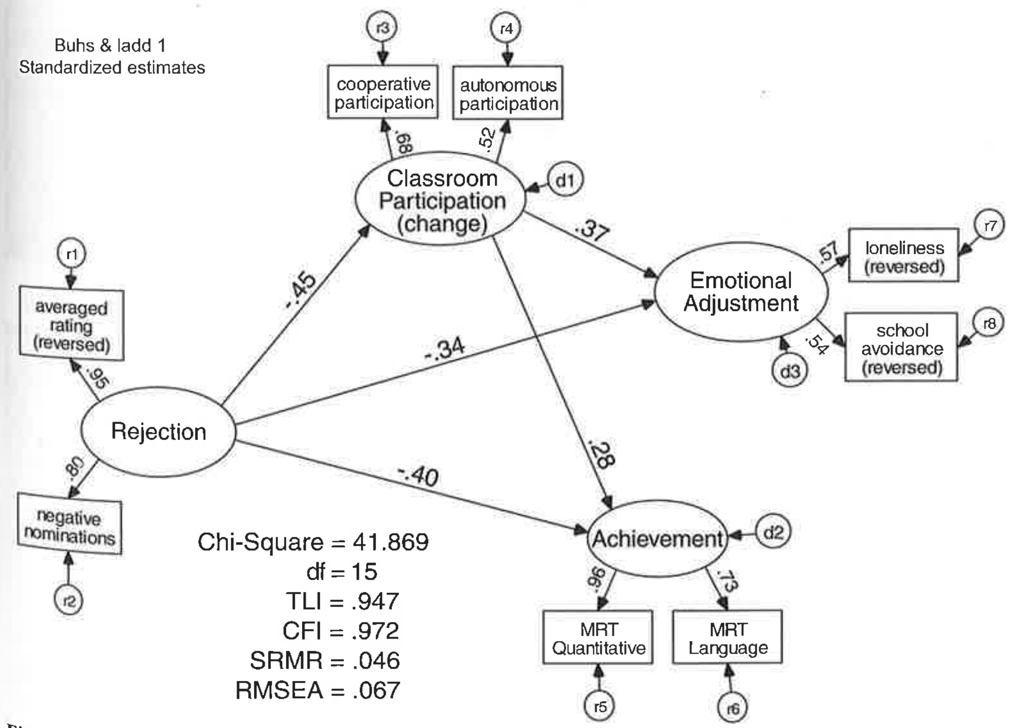


Figure 16.6 Standardized estimates from the initial peer rejection model. The model has an adequate, but not good, fit to the data.

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	21	41.869	15	.000	2.791
Saturated model	36	.000	0		
Independence model	8	972.032	28	.000	34.715

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.047	.974	.938	.406
Saturated model	.000	1.000		
Independence model	.504	.574	.453	.447

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.957	.920	.972	.947	.972
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.536	.513	.520
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.105	.068	.028	.126
Saturated model	.000	.000	.000	.000
Independence model	2.442	2.372	2.125	2.637

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.067	.043	.092	.110
Independence model	.291	.276	.307	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	83.869	84.841	167.637	188.637
Saturated model	72.000	73.666	215.603	251.603
Independence model	988.032	988.403	1019.944	1027.944

Figure 16.7 Fit indexes for the initial rejection model.

Standardized Results

Let's now focus on the meaning of the results (Figure 16.6). Our primary interest was in the effects of Rejection on kindergarten students' academic Achievement and Emotional Adjustment. The standardized direct effect of Rejection on Achievement was −.40, whereas the direct effect on Emotional Adjustment was −.34. Both effects were statistically significant and

large. Given the adequacy of the model, for each *SD* change in the latent Rejection variable, Emotional Adjustment should decrease by .34 of a standard deviation, and Achievement should decrease by .40 of a *SD*, other things being equal. These findings, in turn, suggest strong effects for Rejection on kindergarteners' subsequent Adjustment, both academically and emotionally. Obviously, Rejection can have deleterious effects.

Unstandardized Findings

Focus on the unstandardized coefficients (Figure 16.8). The unstandardized direct effect of Rejection on Adjustment was −.118, meaning that for each 1-unit change in the latent Rejection variable Emotional Adjustment decreased by .118 points. To understand the meaning of this statement, we need to understand the scales involved. The Rejection latent variable was set to have the same scale as the measured Averaged Ratings variable, whereas the Emotional Adjustment latent variable was set to the same scale as the Loneliness scale. The Averaged Ratings variable was originally based on a 3-point scale but was each child's average rating on this 3-point scale by all of his or her classmates. In addition, these ratings were standardized separately by classroom (Buhs & Ladd, 2001). This seems a good approach, but it means

Regression Weights

			Estimate	S.E.	C.R.	P
Classroom_Participation_(change)	<---	Rejection	-.205	.034	-6.055	***
Emotional_Adjustment	<---	Classroom_Participation_(change)	.289	.098	2.944	.003
Achievement	<---	Rejection	-.578	.105	-5.526	***
Achievement	<---	Classroom_Participation_(change)	.886	.274	3.236	.001
Emotional_Adjustment	<---	Rejection	-.118	.034	-3.430	***
NEG_NOM	<---	Rejection	.802	.057	14.157	***
AVE_RAT	<---	Rejection	1.000			
COOP	<---	Classroom_Participation_(change)	1.000			
AUTO	<---	Classroom_Participation_(change)	.788	.141	5.596	***
LONE	<---	Emotional_Adjustment	1.000			
SCHAVOID	<---	Emotional_Adjustment	1.140	.223	5.104	***
LANG	<---	Achievement	1.000			
QUANT	<---	Achievement	1.465	.134	10.901	***

Standardized Regression Weights

			Estimate
Classroom_Participation_(change)	<---	Rejection	-.451
Emotional_Adjustment	<---	Classroom_Participation_(change)	.372
Achievement	<---	Rejection	-.403
Achievement	<---	Classroom_Participation_(change)	.281
Emotional_Adjustment	<---	Rejection	-.335
NEG_NOM	<---	Rejection	.803
AVE_RAT	<---	Rejection	.949
COOP	<---	Classroom_Participation_(change)	.682
AUTO	<---	Classroom_Participation_(change)	.520
LONE	<---	Emotional_Adjustment	.567
SCHAVOID	<---	Emotional_Adjustment	.540
LANG	<---	Achievement	.726
QUANT	<---	Achievement	.957

Figure 16.8 Unstandardized and standardized paths and loadings, standard errors, and critical ratios.

that the Averaged Ratings unstandardized metric and thus the metric of the Rejection latent variable are not readily interpretable. According to the authors, the Loneliness scale is a five-item composite (Buhs & Ladd). Although not explained further, it appears from the means and standard deviations that this scale is also a mean of the item scores. Without further detail, the unstandardized metric of this variable and thus the Emotional Adjustment latent variable are also not interpretable. The unstandardized coefficients, although useful for other purposes (e.g., comparisons with other research), are not readily interpretable, and thus the previous interpretation of the standardized paths is probably our best approach.

Mediation

Many more interesting findings are contained in the model beyond the direct effects. One primary interest of the researchers was to determine whether classroom participation *mediated* the effect of Rejection on Adjustment. In other words, what were the *indirect* effects of Rejection on Adjustment through Classroom Participation? Note in Figure 16.8 that Rejection had a powerful effect on Participation (–.45): rejected children showed less participation than did their nonrejected peers. Classroom Participation, in turn, had a strong effect on both Achievement (.28) and on Emotional Adjustment (.37); children who participated evidenced higher achievement and better adjustment. Thus, it certainly seems that the indirect effects of Rejection on the two adjustment variables were also substantial and that Classroom Participation partially mediates the effects of Rejection on Adjustment.

Indirect and Total Effects

Figure 16.9 shows the standardized direct, indirect, and total effects of the latent variables on each other. Rejection had moderate and negative indirect effects on Achievement (–.126) and Emotional Adjustment (–.168). Although not shown in the figure, these effects were also

Standardized Total Effects

	Rejection	Classroom_Participation _(change)	Achievement	Emotional_ Adjustment
Classroom_Participation_(change)	-.451	.000	.000	.000
Achievement	-.529	.281	.000	.000
Emotional_Adjustment	-.503	.372	.000	.000

Standardized Direct Effects

	Rejection	Classroom_Participation _(change)	Achievement	Emotional_ Adjustment
Classroom_Participation_(change)	-.451	.000	.000	.000
Achievement	-.403	.281	.000	.000
Emotional_Adjustment	-.335	.372	.000	.000

Standardized Indirect Effects

	Rejection	Classroom_Participation _(change)	Achievement	Emotional_ Adjustment
Classroom_Participation_(change)	.000	.000	.000	.000
Achievement	-.126	.000	.000	.000
Emotional_Adjustment	-.168	.000	.000	.000

Figure 16.9 Standardized total, direct, and indirect effects for the initial rejection model.

statistically significant. Although these effects are smaller than the direct effect of Rejection on each variable, they are meaningful and show that students' participation in class partially mediates the effects of rejection on adjustment. Children who are rejected by their peers show less participation, which, in turn, results in lower levels of school emotional adjustment and achievement. Because the direct and indirect effects of Rejection on the academic (Achievement) and Emotional Adjustment variables were both negative, the total effects were even larger (–.529 on Achievement; –.503 on Emotional Adjustment). (Of course we could have calculated these indirect and total effects by hand. For example, the indirect effect of Rejection on Achievement via Participation = –.451 × .281 = –.127. The total effect = –.127—.403 = –.530 [the same value as the figure, within errors of rounding]. With more complex figures, of course, such calculations become considerably more complex.)

COMPETING MODELS

We may wonder if the model, as drawn, is correctly specified. Is it reasonable, for example, to assume that the *only way* Achievement and Adjustment are related to each other is by their both being affected by Rejection and Participation? Or does Achievement affect Adjustment, as well (or Adjustment affect Achievement)?

Figure 16.10 shows an alternative model in which Achievement affects Adjustment. The logic behind this competing model is simple: children who are successful academically, a major component of the orientation of kindergarten, will, as a result, be better emotionally

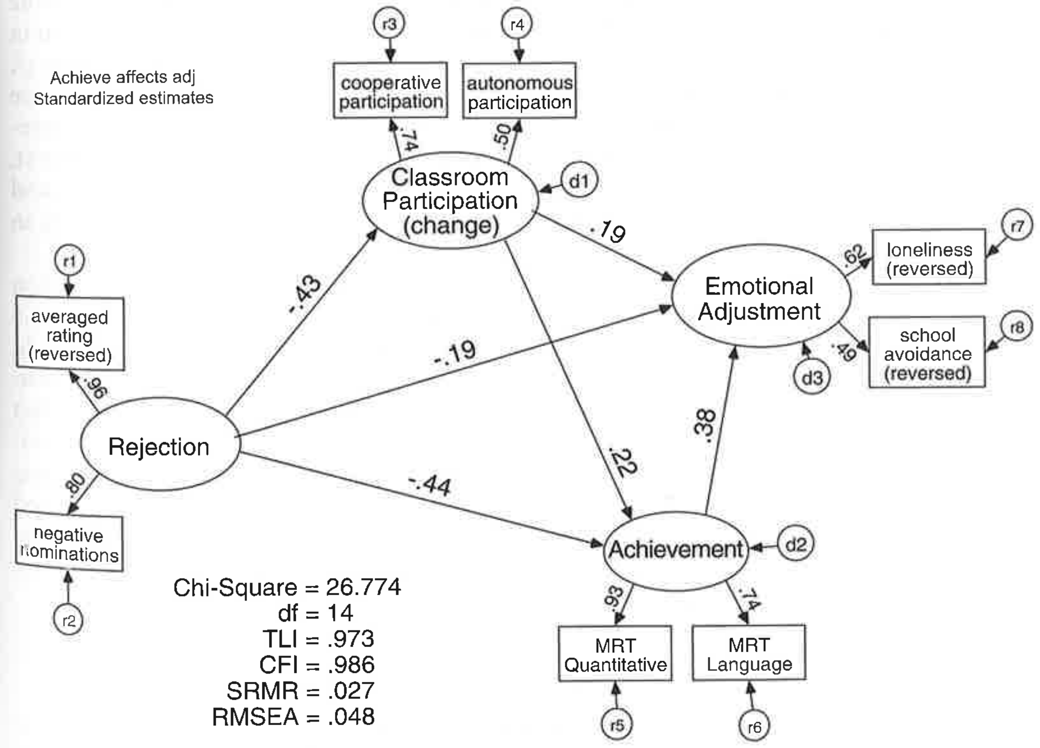


Figure 16.10 Alternative Achievement Effect model of the effects of rejection on educational and emotional adjustment. The model includes a path from Achievement (educational adjustment) to Emotional Adjustment.

Table 16.2 Comparison of the Fit of Alternative Peer Rejection Models

Model	χ^2	*df*	$\Delta\chi^2$	*df*	*p*	*AIC*	*TLI*	*CFI*	*SRMR*	*RMSEA* (90% CI)
1. Initial	41.869	15				83.869	.947	.972	.046	.067 (043-.092)
2. Achievement Effects	26.774	14	15.095	1	< .001	70.774	.973	.986	.027	.048 (.018-.075)

adjusted than will children who have difficulty with the academic aspects of kindergarten. As shown in the figure, this model had a good fit to the data. In particular, the RMSEA was .048, and the TLI and CFI were above .95.

More directly, we can compare the fit of this model with the initial model. Because the two models are nested, we can use $\Delta\chi^2$ to compare the two models. The fit statistics for this Achievement Effect model are shown in Table 16.2, along with those from the initial model. As can be seen in the table, the model in which Achievement was allowed to affect Adjustment resulted in a smaller χ^2 than did the initial model, and this $\Delta\chi^2$ was statistically significant ($\Delta\chi^2$ [1 *df*] = 15.095, $p < .001$). Although the initial model was more parsimonious, our rule of thumb is that when $\Delta\chi^2$ is statistically significant we will reject the more parsimonious model in favor of the better fitting model. In this case, the model shown in Figure 16.10 is the better fitting model; the decrease in parsimony is worth the decrease in χ^2.

Given our acceptance of the Achievement Effect model over the Initial Model, what are the implications for this new model? The results shown in Figure 16.10 suggest that Achievement has a powerful effect on Emotional Adjustment ($\beta = .38$). If this model is correct, then it appears that Achievement is an important mediating variable between Rejection and Adjustment: children who are rejected suffer academically, and this academic difficulty, in turn, results in lower levels of adjustment in school.

This change in the model also substantially reduced the direct effect of both Rejection and Participation on Emotional Adjustment (compare the models shown in Figures 16.6 and 16.10). If you compare the total effects for Rejection on Adjustment in the two models, however, you will find them to be similar. Take a few minutes to consider why this is the case. As long as you are pondering models, it is also worth noting that with the Achievement Effect model (Figure 16.10), the structural portion of the SEM (the paths among the latent variables) is just-identified. That is, for a measurement model there are six correlations among the latent variable; for the model shown in Figure 16.10, all six of those correlations are used to estimate the six paths among the latent variables. Finally, please note that these results are with simulated data. I do not know if the addition of this path would have led to such an improvement in fit in the actual data.

Other Possible Models

You may question why I drew the path from Achievement to Adjustment rather than the reverse. The decision was based primarily on logic. I reasoned that the types of skills and abilities assessed by the Achievement measured variables are more stable than the ratings of loneliness and school avoidance assessed by the Adjustment latent variable. Given what

is meant by these two latent variables, it seemed to me that it was more likely that Achievement would affect Adjustment than it was that Adjustment would affect Achievement. What do you think? Should the path go in this direction or the reverse? It is interesting to conduct this exercise, but if we examine this model as more than an exercise, we will need to examine relevant theory and previous research to see which of these possibilities is more likely. We would use such theory and research to design the study and to draw the path in the appropriate direction.

Why not, you may wonder, just estimate a model with the path drawn in the opposite direction and see how that model fits? Recall the rules for equivalent models in Chapter 13. Unfortunately, these two models are statistically equivalent; their fit is identical. Although this alternative Adjustment Effect model will have very different implications for interpretation, the data cannot tell us which model is correct. It is also inappropriate to run this alternative model, interpret it, and then decide which interpretation we like more. Perhaps, then, we can draw the paths in both directions, a nonrecursive model, and see which path is stronger? This solution will not work either; the structural model will be underidentified. If differentiation between these two models is one of the purposes of the research, the researchers could build in noncommon causes of the two outcome variables and thus test nonequivalent or nonrecursive models; likewise, longitudinal data will help. With the current model and data, we must rely on theory and previous research to make this decision.

What if theory and previous research do not inform this decision; what if you cannot decide in which direction to draw the path? One option, an agnostic option, is shown in Figure 16.11. In this model, we have allowed the disturbances of Achievement and Adjustment

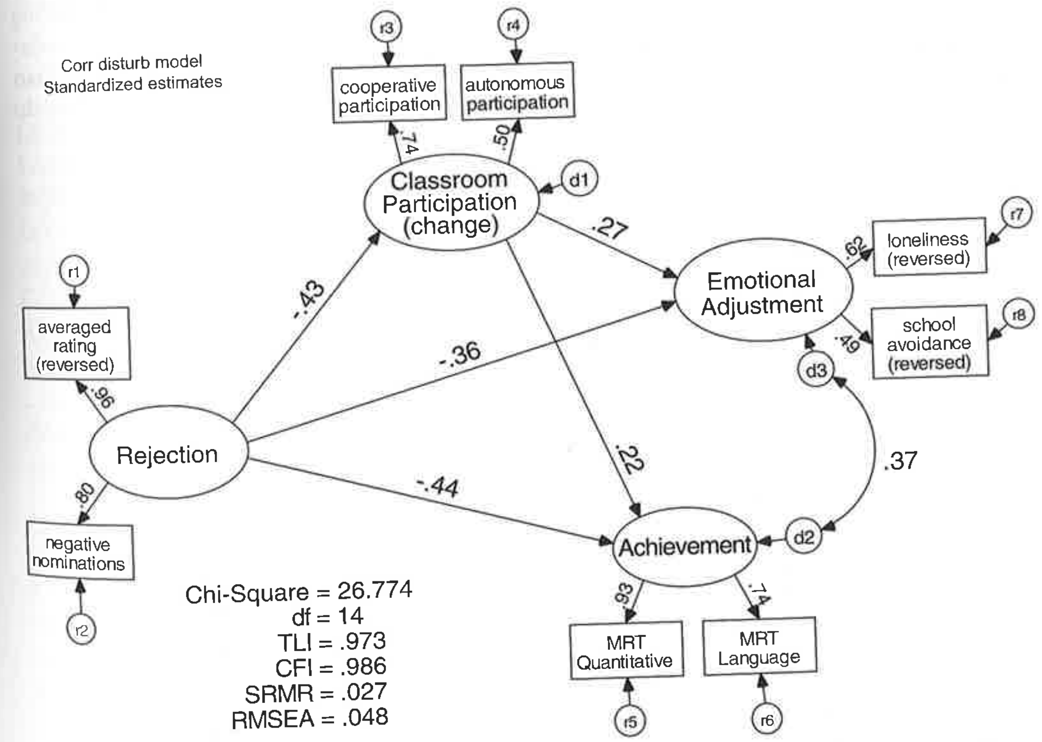


Figure 16.11 Another alternative model of the effects of rejection. This agnostic model specifies an unknown causal relation between Emotional Adjustment and Achievement. The model is equivalent to and statistically indistinguishable from the previous Achievement Effect model.

to be correlated. Note that this model is also equivalent to the model in Figure 16.10; the fit indexes are therefore the same, and the data cannot tell us which of the two models is correct. But consider what this model with the correlated disturbances says about our assumptions of the causal process underlying these variables. The disturbances represent all other influences on the latent variables other than the variables in the model that are pointing to the latent variable. To allow the disturbances to be correlated means that we recognize that these other causes may be related. In other words, we recognize that Emotional Adjustment and Achievement may be related in other ways beyond the paths shown in the model, but we're not really sure what these other relations may be. Practically, these correlated disturbances may mean that the two variables are causally related, but we don't know the direction. The correlated disturbances may also mean that there is some other variable, not included in the model, that affects both Adjustment and Achievement (an unmeasured common cause). If you think about it, this correlation means what any correlation may mean: *a* may cause *b*, *b* may cause *a*, or there may be a third variable, *c*, that causes both *a* and *b*. Again, the models are equivalent, so we can't decide which is correct based on the data. As a general rule, however, I prefer to make the causal statement (Figure 16.10) than to be noncommittal (Figure 16.11), but I want a more solid grounding in relevant theory and research than I now have before making the decision of causal direction. We will return to the topic of causal direction in the next chapter.

MODEL MODIFICATIONS

The competing model discussed above was developed based on logic rather than analysis of the detailed fit information. You may wonder, if we had not thought of this competing model, would the modification indexes (MIs) or the standardized residuals (or the correlation residuals) have hinted at it? Figure 16.12 shows the modification indexes greater than 4.0 for the initial model (from Figure 16.6). Although many of the modification indexes do

Modification Indices

Covariances

	M.I.	Par Change
d3 <-->d2	12.075	.086
r7 <-->d2	16.425	.120
r7 <-->r5	4.590	.073
r4 <-->r5	7.147	.101
r4 <-->r6	10.334	-.117

Regression Weights

		M.I.	Par Change
Achievement	<--- Emotional_Adjustment	4.714	.522
Emotional_Adjustment	<--- Achievement	7.159	.046
QUANT	<--- AUTO	5.448	.250
LANG	<--- AUTO	8.107	-.294
LONE	<--- Achievement	9.785	.065
LONE	<--- QUANT	9.925	.041
LONE	<--- LANG	9.972	.046
AUTO	<--- LANG	4.345	-.033

Figure 16.12 Modification indexes for the initial rejection model.

not make a lot of sense, several are worth noting. The largest index suggests that χ^2 could be reduced by at least 16.425 by freeing the correlation–covariance between the residual for Loneliness (r7) and the disturbance for Achievement (d2). This modification makes little sense. The next largest modification index (12.075 for the covariance between d3 and d2) does, however. This MI suggests that the model will fit statistically significantly better if this covariance is freed. Focus on the MIs for the regression weights (the paths). Although they are not the largest MIs, the first two listed also suggest that the fit of the model could be improved by focusing on the relation between Achievement and Emotional Adjustment. Thus, although the modification indexes do not point directly to our Achievement Effect model, they certainly hint in that direction.

Table 16.3 shows the standardized residual covariances and the correlation residuals among the variables. These residuals show that the Initial Model did not adequately account for the correlations between Loneliness and the MRT Quantitative and Language scores and also between Language and School Avoidance. The table of residual correlations also shows that these residuals are substantial. The model predicts a correlation of .144 between the MRT Language test and the Loneliness scale, whereas the actual correlation between these measured variables was .301, a difference of .157 (the actual correlation and the implied correlation are not shown in the table but are easily accessible in the text output from Amos or

Table 16.3 Standardized residual covariances and residuals correlations for the initial rejection model.

Standardized Residual Covariances

	QUANT	*LANG*	*SCHAVOID*	*LONE*	*AUTO*	*COOP*	*AVE_RAT*	*NEG_NOM*
QUANT	0							
LANG	0	0						
SCHAVOID	.682	1.260	0					
LONE	2.968	3.096	0	0				
AUTO	.321	−1.821	−.008	−1.089	0			
COOP	−.608	−.488	−.128	−.444	.313	0		
AVE_RAT	.006	−.235	.236	−.053	.760	−.286	0	
NEG_NOM	.433	−.104	−.536	.741	.333	−.983	.014	0

Residual Correlations

	QUANT	*LANG*	*SCHAVOID*	*LONE*	*AUTO*	*COOP*	*AVE_RAT*	*NEG_NOM*
QUANT	0							
LANG	0	0						
SCHAVOID	.035	.064	0					
LONE	.152	.157	0	0				
AUTO	.016	−.093	−.001	−.055	0			
COOP	−.031	−.025	−.007	−.023	.016	0		
AVE_RAT	.001	−.013	.012	−.003	.039	−.015	0	
NEG_NOM	.024	−.005	−.028	.038	.017	−.051	.001	0

other SEM programs). Again, the residuals *might* lead you in the direction of the Achievement Effect model if you had not thought of it previously.

As long as we are cleaning up our models, we might reexamine the statistical significance of the various parameter estimates to see if all paths are statistically significant, with the idea that if any are not it will be okay to remove them. As shown earlier in Figure 16.8, all paths were statistically significant. Although not shown here, all paths are also statistically significant in the Achievement Effect model. It is worth reiterating a previous point: models that are extensively modified based on modification indexes and other tools for model modification should be considered exploratory, tentative models until tested against new data.

SUMMARY

This chapter is the first to focus on latent variable structural equation models. Such SEM models may be considered as a confirmatory factor analysis of the various constructs involved in the research, with a simultaneous path analysis of the effects of these constructs on each other. The chapter reviewed the components of latent variable SEMs and illustrated the methodology with an extended example from the research literature.

Conceptually, you may consider latent variable SEM as a confirmatory factor analysis of the constructs underlying the measured variables in the research, along with a path analysis of the latent variables. The measurement model includes the latent variables, constructs, or factors that underlie the measured variables in the research as causes of these measured variables. The measurement model also includes latent variables, one per measured variable, representing the unique and error variances of each variable, or all other causes of that measured variable other than the construct/latent variable. The structural model includes the paths and covariances among the latent variables, along with the disturbances for the endogenous latent variables (all other causes of the latent variables other than those with arrows pointing to the latent variables). It is often confusing to those new to the SEM methodology to know which variables require latent variables representing disturbances or unique/error variances. At the most mechanical level, any variable that has an arrow pointing to it must also include a latent variable representing all other influences on this variable. For measured variables, these other influences are unique and error variances. For latent variables, these other influences generally represent disturbances along the lines of the disturbances from path analysis or the residuals from multiple regression analysis. In fact, you can, and some methodologists recommend that you do, analyze the model separately as a measurement (confirmatory factor) model, and then add the structural model. We have not used this process here, but it can be useful, especially for complex models or in the beginning stages of research.

The research example used in the chapter was based on research on the effects of peer rejection on kindergarten students' academic and emotional adjustment (Buhs & Ladd, 2001). The example analyzed models similar to (but smaller than) those analyzed in the actual research, with data simulated to mimic the actual data. The initial model included four latent variables with two measured variables indexing each latent variable (more good measures per latent variable are preferable in practice, but our interest was in a smaller, more manageable example). We split apart the measurement model from the structural model for conceptual purposes but not for analysis. The initial model was fairly parsimonious (15 *df*), with most of the degrees of freedom a result of constraints in the measurement model (undrawn factor loadings from latent to measured variables).

The initial model had an adequate fit to the data and suggested that Rejection by peers resulted in lower subsequent Achievement and school-related Emotional Adjustment. A portion of these effects were indirect, or mediated, through Class Participation: rejected students had lower rates of participation, which resulted in lower achievement and adjustment. Thus, all three types of effects—direct, indirect, and total—were interesting and interpretable.

An alternative model, which included an additional path from Achievement to Emotional Adjustment, was also estimated. This change resulted in a statistically significant improvement in χ^2, which we interpreted as meaning that the alternative Achievement Effect model was a better explanation of the data than the initial model. The alternative model led to different interpretations of direct, indirect, and total effects. As an aside, this change (in the structural portion of the model) used up the 1 degree of freedom that was due to the structural portion of the model.

Any complacency we may have garnered that we had now found the correct model was quickly shattered, however. The chapter discussed two alternative models that are equivalent to our preferred Achievement Effect model. Although these two models are statistically indistinguishable from the Achievement Effect model, they have very different interpretations and implications. The chapter included the standardized figural output from one of these alternative models to demonstrate its statistical, but not conceptual, equivalence to the Achievement Effect model. This fuzziness served as another reminder of the importance of theory, logic, and previous research in the construction of models. The equivalent models also served as a reminder of the importance of planning the research so that you can indeed answer the questions of interest.

In the final section of the chapter we examined some of the more detailed fit statistics from the SEM program output. The modification indexes and the standardized residual covariances and correlations for the initial model hinted at the change we made in the Achievement Effect model (although they also suggested the other equivalent, indistinguishable models). Although we might have arrived at the same place had we constructed the alternative Achievement Effect model based on these hints, alternative models devised prior to the examination of the data and results should generally be given more credence than models derived from extensive data-driven model modifications. There were no statistically not-significant paths or factor loadings that we might have constrained in subsequent models.

Although not discussed in detail, there are always equivalent possible models, and their veracity must be tested against these (theory, etc.) standards, not through complex statistical analysis. We can test and reject some models, but we can rarely (maybe never) test and evaluate all possible models that would result in alternative interpretations. Some we don't think of, and some are indistinguishable. At the most basic level, our models always come back to this need for theory, thought, and previous research. "The study of structural equation models can be divided into two parts: the easy part and the hard part" (Duncan, 1975, p. 149). The hard part is developing sound, theory-grounded models. Again, welcome to the dangerous world of SEM.

EXERCISES

1. Analyze the simulated Buhs and Ladd data ("Buhs & Ladd data.sav" or Buhs & Ladd data.xls") using a structural equation modeling program (if you are using Amos, the initial model is saved as "Buhs & Ladd 1.amw"; the Mplus script is also online).
   a. Estimate the models discussed in this chapter. Study the parameter estimates and standard errors, the fit statistics, modification indexes, and standardized residuals.
   b. Interpret the model. Be sure to interpret the indirect and total effects in addition to the direct effects.

c. Compare the initial model with the competing model discussed in this chapter (the Achievement Effect model). Do you agree that this model is a better alternative? What theoretical, logical, and research evidence can you offer in support of this model? What evidence argues against this model?
d. Are there other alternative models that you are interested in testing? Do so; be sure to evaluate the relative fit of the model and to interpret your findings.
e. Are there any common causes that the authors may have neglected? How could you investigate the possibility of unmeasured common causes more completely?

2. Figure 16.13 shows a model to test the effects of participation in Head Start on children's cognitive ability. This example is a classic reanalysis of a controversial quasi-experiment; I have seen variations of it presented in Kenny (1979) and Bentler and Woodward (1978), among others. The measured background variables in the model include measures of mother's and father's educational attainment, father's occupational status, and family income. Head Start was hoped to improve participants' cognitive skills, and the latent Cognitive Ability outcome was indexed by scores on two tests: the Illinois Test of Psycholinguistic Abilities (ITPA) and the Metropolitan Readiness Test (MRT). The Head Start variable is a dummy variable coded 1 for those who participated in Head Start and 0 for children in the control group. The data are shown in Table 16.4. These are data from 303 white children from an early Head Start evaluation, 148 who attended Head Start in the summer and 155 who did not. To understand why the example is so controversial, note the correlation between Head Start and the two cognitive outcomes: both are negative (–.10, –.09), suggesting that Head Start may have negative effects on Ability! The model is one of several possible models designed to determine what the outcomes of Head Start are after taking the family's background characteristics into account. The correlations and *SD*s are also included here and in the Excel file "head start.xls" (the SDs are not included in most presentations of these data. I estimated these and the means from data presented in Magidson & Sörbom, 1982). All continuous variables are standardized.
   a. Draw (set up) and estimate the model. Is the structural portion of the model just-identified or overidentified? Evaluate the fit of the model and, if adequate, focus on parameter estimates. Interpret the model. According to these results, does Head Start have a positive effect on cognitive ability, a negative effect, or no effect at all? Interpret the other aspects of the model.

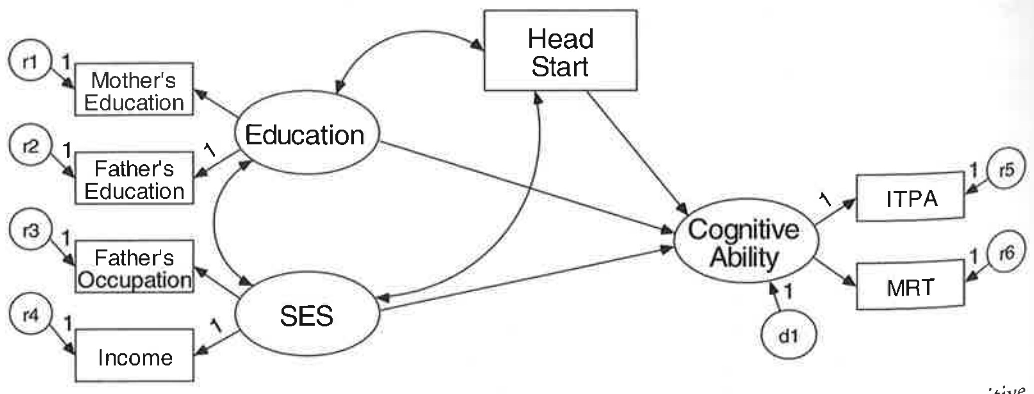


Figure 16.13 Model testing the potential effects of Head Start participation on children's cognitive ability

b. Fix the path from Head Start to Cognitive Ability to zero; compare the fit of this model to the initial model. Do you still come to the same conclusion as before?
c. Are there other alternative models that you are interested in testing? Are they equivalent to the initial model? Test these models; be sure to evaluate the relative fit of the model and to interpret your findings.
d. Are there any common causes that the research may have neglected? How could you investigate the possibility of unmeasured common causes more completely?

3. Kimmo Sorjonen and colleagues used SEM to estimate the relative effects of intelligence, family of origin SES, and emotional capacity (at the time of their conscription into the military) on Swedish men's occupational status at ages approximately 35-40 (Sorjonen, Hemmingsson, Lundin, Falkstedt, & Melin, 2012). The authors were interested in the relative effects of these variables as well as the extent to which their effects were mediated by educational attainment. Figure 16.14 shows the authors' model (minus one correlated error). A dataset of 1000 cases, simulated to give similar findings to the article, are on the website in the file "Sorjonen et al simulated 7.sav" (the actual research had an *N* of over 48,000). Note that while the simulated data are designed to mimic the means and variances of the original data, I have not been strict in the scaling; thus there are items that have (impossible) negative values. A brief explanation of the variables in the analysis are shown in Table 16.4.

Estimate the model shown. Create a table of direct, indirect, and total effects on the final outcome (Attained Occupation). Which variables are the most important influences on these men's eventual occupations? Which variables are less important? Interpret your findings. Is there anything unusual about this model?

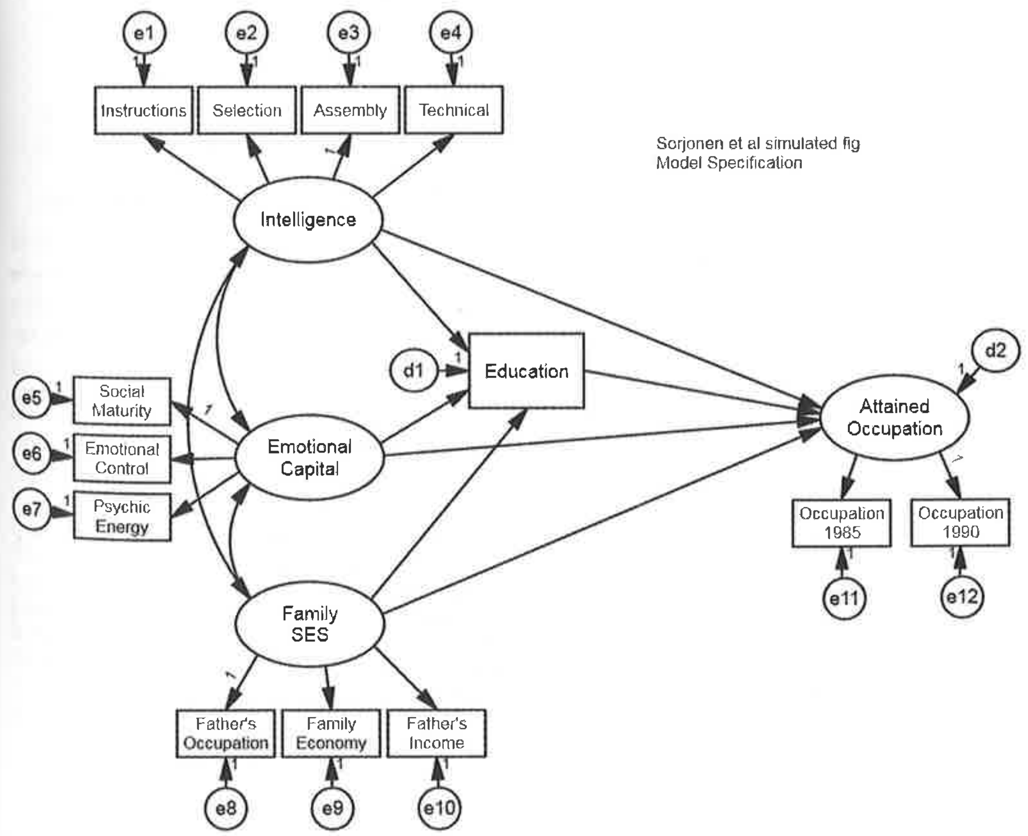


Figure 16.14 Model for the Sorjonen and colleagues (2012) exercise.

Table 16.4 Variables in the Sorjonen et al. (2012) example.

Variable Name	*Label in Figure*	*Description*
Instructions		Short measure of verbal intelligence & inductive reasoning
Selection		Short measure of verbal intelligence & inductive reasoning
Assembly		Short measure of visual-spatial reasoning
Technical		Short measure of "mechanical ability" and "technical understanding" (p. 270)
Pop Occ	Father's Occupation	Occupation status on a 5-point scale from census
Fam Economy	Family Economy	Participant's ratings of their family's economic standing from very poor (1) to very good (5), rated in 1969/70 at time of conscription
Pop Income	Father's Income	Natural log of participant's father's income, from census data, for 1970
Maturity	Social Maturity	Psychologist's ratings in 1969/70 irresponsibility and maladjustment versus "responsibility . . . independence, . . . and extraversion" (p. 271)
Control	Emotional Control	Psychologists' ratings of nervousness and anxiousness versus calmness
Energy	Psychic Energy	Psychologists' ratings of a lack on initiative versus initiative and ideas
Occ 85	Occupation 1985	Occupational status from 1985 census
Occ 90	Occupation 1990	Occupational status from 1990 census
Education		Level of education (7 point scale) from 1990 Census data

Note

1 Here is the calculation of *df* for the measurement and structural models: with eight measured variables, there are 36 elements in the variance/covariance matrix: $\frac{p\times(p+1)}{2}=\frac{8\times 9}{3}=36$. For the measurement model we estimate 22 parameters: 6 correlations/covariances among the factors, 4 factor loadings (recall that for each factor one factor loading is set to one to set the scale), 4 factor variances, and 8 unique/error variances (r1 through r8). 36 – 22 = 14 *df* for the measurement model. For the full latent variable SEM we are estimating 21 parameters: the 8 unique/error variances, 4 factor loadings, 1 factor variance (for the exogenous variable, Rejection) and 3 variances of disturbances (d1 through d3), and 5 paths. For this model, 36 – 21 = 15 *df*. Another way of thinking about *df* is to apportion them to the measurement versus structural models (Figures 16.3 versus 16.4). As already calculated, the measurement model accounted for 14 *df*. In the structural model, the 6 factor correlations are replaced by 5 paths, resulting in 1 additional *df*.

17

Latent Variable Models

More Advanced Topics

In the previous chapter we introduced and explored latent variable structural equation models. This chapter will review and consolidate that learning by reviewing another example. We will continue our exploration with several more advanced topics and an assessment of where we stand in our efforts to conduct meaningful nonexperimental research. The chapter will begin with a model that incorporates two complexities that we have touched on previously: single-indicator variables and correlated errors.

SINGLE INDICATORS AND CORRELATED ERRORS

A Latent Variable Homework Model

Figure 17.1 shows a latent variable version of our earlier Homework model from Chapter 13. The primary variables in the model are Homework, indexed by student reports of average time spent on homework in 8th (Homework 8th) and 10th (Homework 10th) grades, and students' overall Grades in high school, a latent variable estimated by students' high school GPAs in English, Math, Science, and History–Social Studies