Technical Report # 1306

Basic Concepts of Structural Equation Modeling

Daniel Anderson

Chalie Patarapichayatham

Joseph F. T. Nese

University of Oregon

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Note: Funds for this dataset were provided by the Oregon Department of Education Contract No. 8777 as part of Project OFAR (Oregon Formative Assessment Resources) Statewide Longitudinal Data System (SLDS), CFDA 84.372A Grant Program that is authorized by the Educational Technical Assistance Act of 2002.

Abstract

In this paper we introduce the basic concepts of structural equation modeling (SEM) for consumers of research. The purpose is to help provide readers a basis from which articles employing SEM can evaluated; but not necessarily to teach readers how to conduct an analysis. We assume little prior statistical knowledge, and thus begin by introducing the concepts of simple linear regression. We then expand to multiple regression, path analysis, confirmatory factor analysis, and finally, structural equation models.

Basic Concepts of Structural Equation Modeling

Structural equation modeling (SEM) is a powerful and flexible approach to statistically model relations among variables, or measured characteristics of interest (e.g., student achievement). Two characteristics of SEM differentiate it from other statistical techniques; the ability to model: (a) latent, unobserved or unmeasured theoretical variables by using a combination of two or more observed variables; and (b) complex structural relations rather than simple X affects Y relations.

In this paper, we present a basic overview of SEM and the intended audience is consumers of research using SEM practices. Thus, the purpose of the paper is to provide a conceptual understanding of SEM, but not necessarily to show the reader how to conduct an SEM analysis. We begin by first introducing common SEM symbolism, sometimes referred to as reticular action model (RAM) symbols (Kline, 2010). We then introduce simple linear regression, which serves as the basis for SEM, and provide an example using real data. These same data are then used to build increasingly complex models using multiple regression, then path analysis, confirmatory factor analysis, and finally a full SEM model. We show how the RAM symbols can be used to describe the modeled relationships regardless of the level of complexity. We conclude with a discussion of the flexibility of SEM.

RAM Symbols

Five basic RAM symbols can be used to describe essentially any relation among variables: directional arrows $\longrightarrow$, double-headed curved arrows $\curvearrowleft\!\curvearrowright$, circular curved arrows $\circlearrowright$, boxes $\square$, and circles $\bigcirc$. The boxes and circles represent the variables of interest, while all arrows display relations. The key distinction between variables depicted by boxes versus circles is whether the variable is observed or measured in some way (e.g., student achievement test) or if it represents some theoretical entity (e.g., motivation). Boxes represent observed variables and circles represent theoretical variables. The directional arrows imply one variable having a direct affect on another (i.e., one variable regressed on the other), the double-headed curved arrows imply a covariance between two variables (or in its standardized form, a correlation), and the circular curved arrows represent the variance of a variable. The circular curved arrow symbol is used because the variance of a variable is literally its covariance with itself. The key distinction between directional and double-headed curved arrows is whether the model is implying that one variable has an affect on the other (directional), or whether the variables are simply related (double-headed curved).

Generally, the choice of the parameter (directional or double-headed curved arrow) is based on the hypothesized relation among the variables (i.e., the theory behind the model), but an important consideration is also temporal ordering. For example, a variable cannot have a direct affect on another variable that has been measured earlier in time, but the variable measured first can have a direct affect on the variable measured second, or the variables can be correlated. When the symbols are used in conjunction to describe a set of relations, the resultant figure is known as a path diagram. Circular curved arrows should always be represented in the path diagram and always with the exogenous variables (the independent variables, or those most "upstream" in the path diagram).

Figure 1 – *Correlation and Regression Path Diagrams*

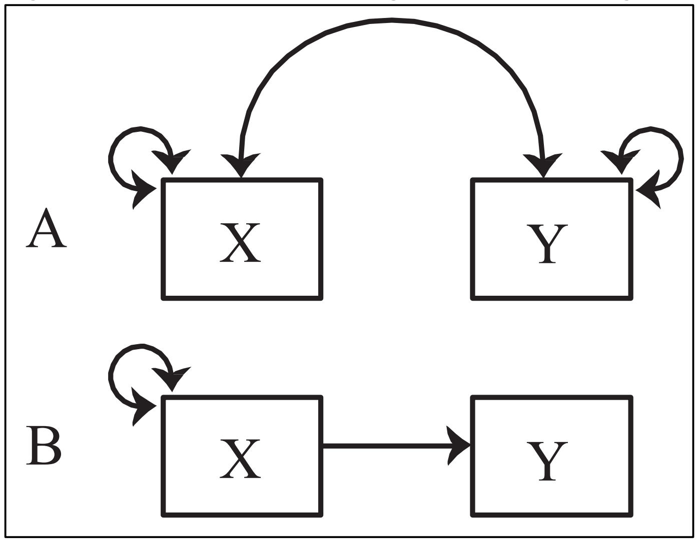


Figure 1 displays two path diagrams depicting simple relations with the RAM symbols. Note that only observed variables are shown to this point, but later in the paper we introduce latent, unobserved variables.

Path diagram A shows two variables that are related, with no directionality. The variables are allowed to correlate (as indicated by the double headed curved arrow) but one variable does not necessarily affect the other. By contrast, path diagram B implies a direct affect of X on Y. It is important to be cautious with language and not state that X *causes* Y, but simply that X affects Y. In path diagram A the variables are simply correlated, each with a variance component. The diagram implies that the variance in one variable "goes along" with variance in the other variable. Thus both variables have a variance component. For example, elementary students' shoe size may be correlated with their reading ability, because older students have received more schooling – leading to higher reading levels – while also being more mature physically. Thus there is a positive correlation between shoe size and reading ability, despite one variable having no affect on the other.

In contrast, path diagram B represents a regression equation for two variables. Note that only the X variable has a variance component. In path diagram B, the X variable is *accounting for* the variance in Y. The variance in Y is being explained by the variance in X (through the regression equation). It is also important to note that at this point there is no residual term represented, which would mean that X is *fully* accounting for the variance in Y *and* that the variables are measured without error. In application this would essentially never be the case and a residual term would need to be included for Y. Residual terms will be discussed in more depth later in the paper.

Simple Linear Regression

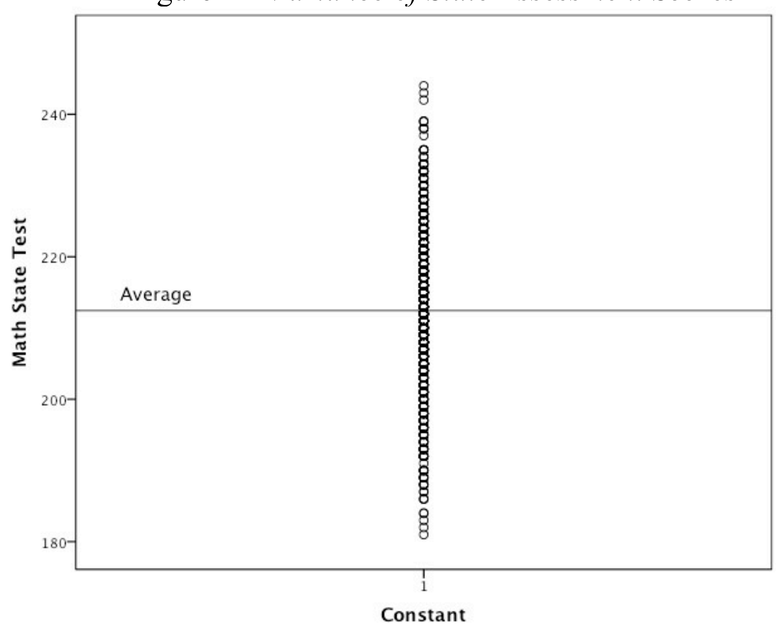
The goal of most statistical analyses is to explain an event (e.g. treatment or outcome). In public schools, the treatment is typically instructional and the outcome performance or progress on tests and measures of achievement.

Problem**:** *Why do some students perform well on the state test while others perform poorly?*

The first step in answering this question is to hypothesize some variables that may play a role in students' state test performance. For instance, do students perform differently because of the instruction they are receiving? Is it because of a general intelligence level? Is it because of the students' home life? All of these are possible sources of variance for performance on a state test. For instance, perhaps one student performs differently on the state test from other students, in part, because he or she has parents that require homework every night. Knowing this information can help us to understand, in part, *why* the student is performing differently on the state test. No single variable (i.e., intelligence, home life, etc.), however, can ever explain fully why a student performs differently from other students. In application, we are always trying to explain, or account, for as much of the variance in students' scores as possible.

In regression, the variance in students' scores can be divided into three components: the average of all scores, the coefficients in the model, and unaccounted variance or residual. If we are provided a group of scores, and we know nothing about any individual student, our best guess as to how any one student would perform on the test would be the average of all the other scores (Galton, 1886). In Figure 2, you see students' math state test scores plotted along the y-axis, and a constant along the x-axis. This figure shows the variance in the state test scores. Again, if we know nothing about these students, our best estimate of an individual student's performance would be the average of all students. For some students this guess would be fairly close (those whose scores are near the average). For most students, however, this guess would not summarize their performance very well.

Figure 2 – *Variance of State Assessment Scores*



If we add a predictor variable to the x-axis, the scatterplot changes and we're able to get a regression line. Figure 3 shows the same data, but this time plotted against students' scores on a formative math assessment. Now the predictor variable – the formative assessment – is regressed on the outcome and provides more information to estimate any individual student's performance.

The goal of conducting a regression analysis is to fit a line that summarizes the data best by minimizing the distance between the line and the data points.

Figure 3 – *State Assessment with Predictor Variable and Regression Line*

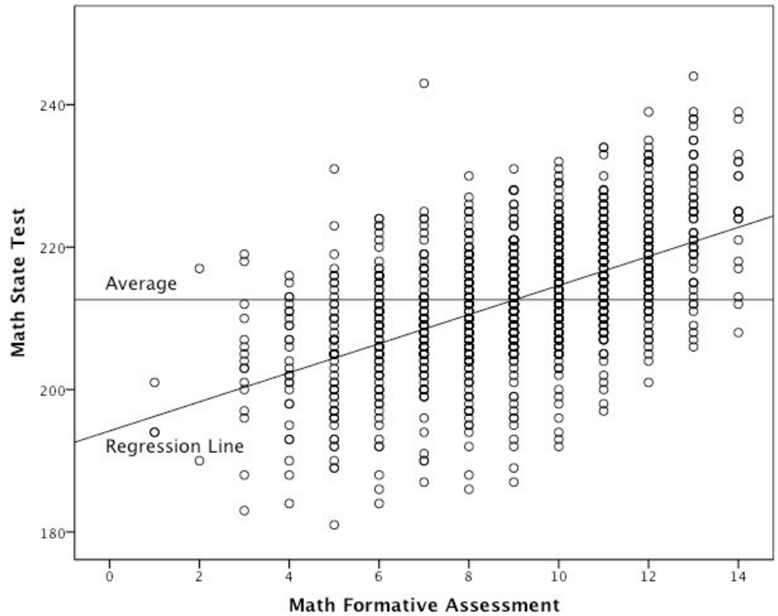
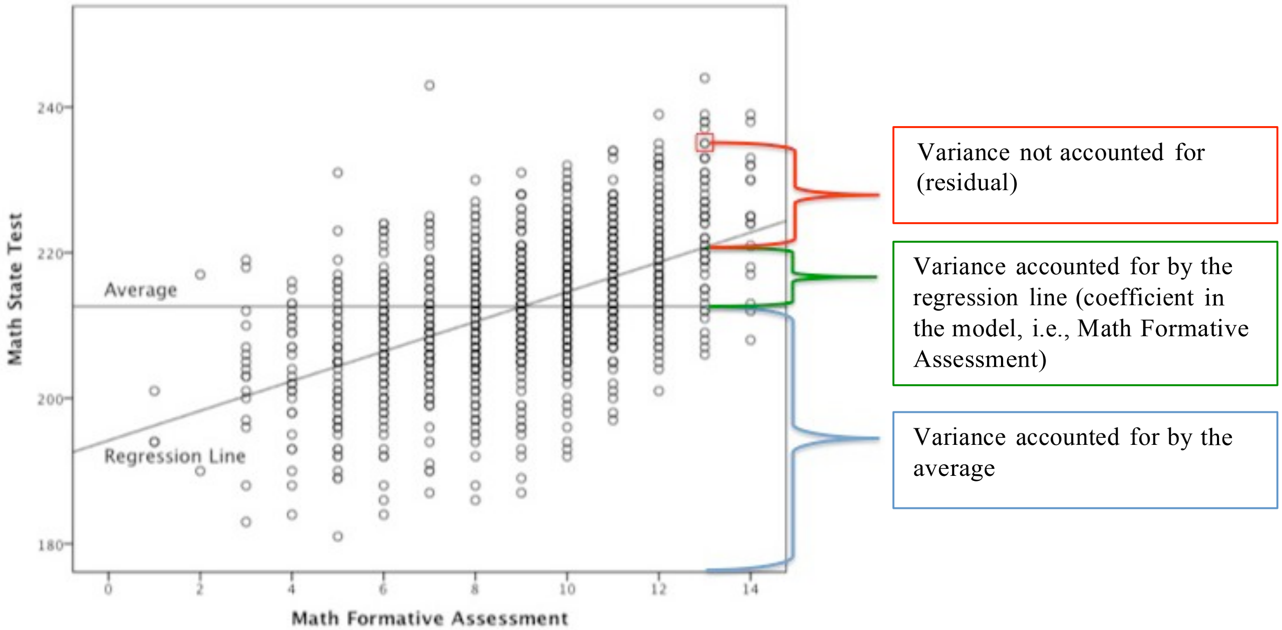


Figure 4 below again shows the same data, but with the variance decomposed into its three sections for a single student. Notice that for this individual student (outlined with a red box), the variance associated with the average score is largest. However, using just that average score does not summarize this student's performance well. Adding the regression line gets us much closer to the student's score, but still does not fully account for *why* the student scored where he or she did. The red portion in the figure is the variance not accounted for and is generally called the residual. Residual refers to what is "left over" after the regression line. Although it includes measurement error (e.g., unreliability in the measurement tool), the residual is a result of multiple different factors including unreliability in the measure *and* omitted variables that might also explain the outcome. Thus we use the term "residual" throughout this paper in reference to unaccounted variance.

Figure 4 – *Regression Line with the Variance Decomposed*

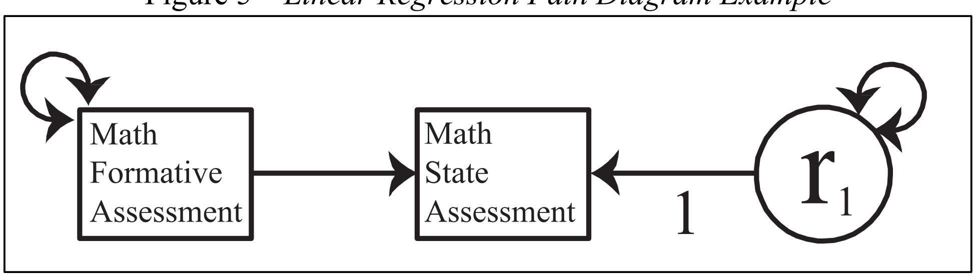


When conducting a regression analysis, we attempt to maximize the variance accounted for by the regression equation (regression line) and minimize the variance not accounted for (red line – distance each "dot" falls away from the regression line). If 100% of the variance were accounted for, all the dots on the scatterplot would fall on the regression line. However, as previously stated, we are essentially never able to completely explain all the variance and there will nearly always be some amount of residual variance. The scatterplots displayed in Figures 3 and 4 account for 29% of the variance in state test scores. The actual statistic behind the variance accounted for is simply the squared correlation coefficient. In other words, the correlation between the math formative assessment and math state test is .54, which when squared equals .29.

The regression line can be summarized by the regression coefficient, beta, which represents the change in the outcome with every one-unit change in the independent variable. It is often reported in its standardized form, , which represents the change in the outcome in standard deviation units for every one standard deviation change in the independent variable.

We can also draw the same relation depicted in the scatterplot with a path diagram (see Figure 5 below). Notice that this path diagram differs from the one in Figure 1B in one important way – there is now a residual term. From inspecting the scatterplot in Figures 3 and 4 it is quite apparent that our predictor variable, Formative Assessment, does not account for all the variance in our dependent variable, State Assessment. The variance that is not accounted for (the red part in Figure 4) ends up in the residual term in the path diagram. The relation can be conceptualized through the path diagram as two competing entities vying for the variance in State Assessment. Our predictor variable, Formative Assessment, gets to go first and “eat up” all the variance that is common between the two variables. The residual term goes next, and takes everything left, leaving State Assessment without any variance. The variance of State Assessment is thus parsed out into a section accounted for by the predictor variable and a section not accounted for, the residual. Notice that the residual term has its own variance, which is simply the left over State Assessment variance not associated with the predictor variable.

Figure 5 – *Linear Regression Path Diagram Example*



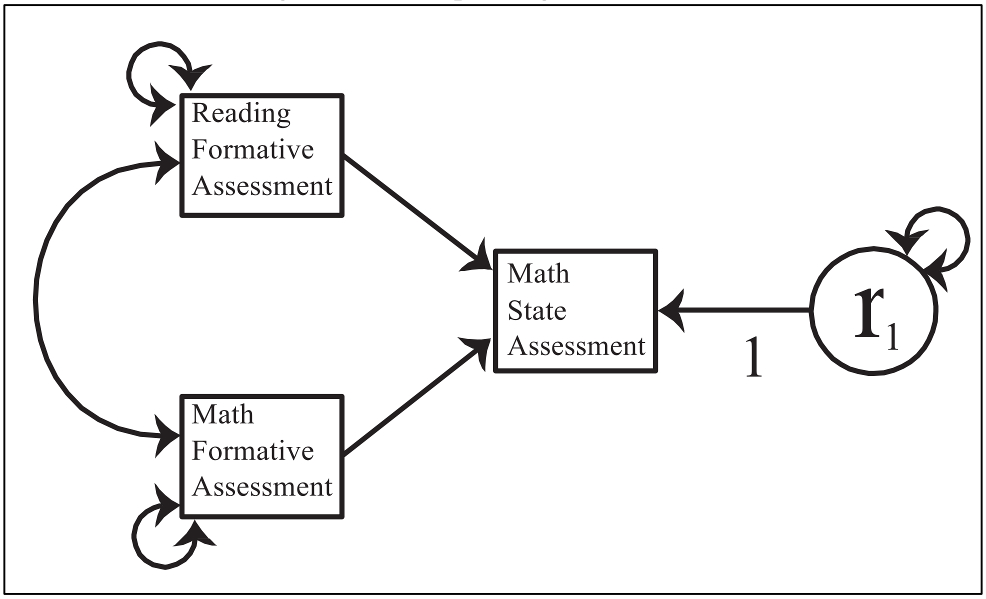
Notice that the residual term is a circle, not a box, implying an *unobserved* variable. By definition, the residual term is unaccounted for variance, which means it also is unmeasured and thus unobserved. Notice that the path from the residual to the dependent variable has a 1 by it, which implies that the path has been “fixed” to that value. All latent variables must have at least one parameter (i.e., path or variance) fixed at a specified value to provide a scale. Without at least one path being fixed, the latent variable cannot be identified. Fixing a parameter at a specified value means that it cannot be tested for statistical significance. However, the model can still estimate a *standardized* value for fixed paths, which can be tested for statistical significance. Generally residual terms have the paths fixed, and not the variance, because we are typically interested in the variance in the residual and not the path from the residual to the dependent variable. Again, the variance within the residual term includes unreliability in the measured variables and/or variance associated with additional variables that are not included in the model.

Multiple Regression

A basic extension of simple linear regression, and a fundamental concept of SEM, is to apply multiple predictor variables to a single dependent variable. These multiple predictors may more fully account for the variance in the dependent variable and thus help us explain outcomes. Figure 6 below displays a multiple regression extension of the example outlined in the linear regression section. Notice that there are now two predictor variables, Reading Formative Assessment and Math Formative Assessment. These two variables are still predicting a single outcome – state test performance in Math.

As with all research, it is important to have strong theoretical justifications for all statistical models applied. The model displayed in Figure 6 implies that students' reading and math skills both play a role in the score they receive on a Math State Assessment. Theoretically, we could imagine both students' reading and math skills playing a role in the Math State Assessment score because many items on the test require students to read. Notice also that the independent variables are allowed to correlate, given the double-headed curved arrow connecting them. That is, we expect students' reading and math formative scores to be related, so we account for that in the model.

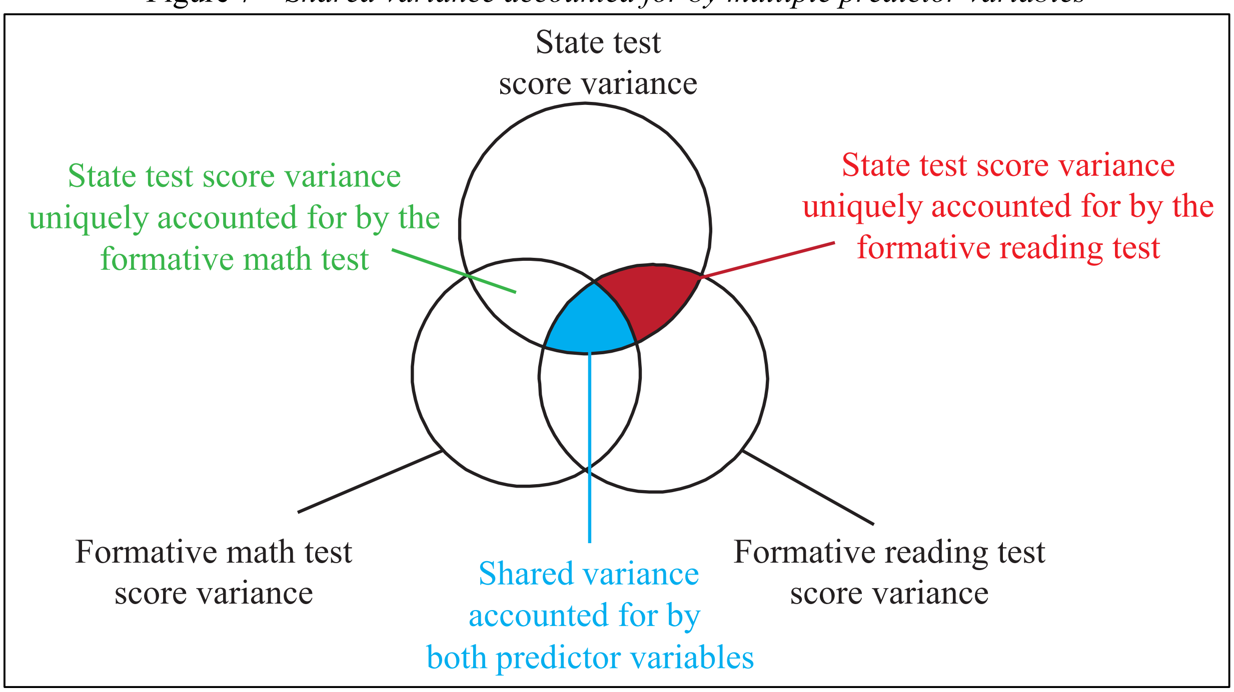
Figure 6 – *Multiple Regression Model*



With a multiple regression model we can obtain a squared multiple correlation coefficient, otherwise known as the R^2 value. The R^2 value indicates the percent of variance accounted for by *all* the predictor variables. When there is just a single predictor variable, as in simple linear regression, the r^2 value is the correlation coefficient squared. However, in multiple regression the R^2 value represents the combined effects of all the predictor variables. The variance accounted for by simple linear regression and multiple linear regression are often differentiated by capitalizing the *r*. In other words, r^2 represents the variance accounted for by a single predictor while R^2 represents the variance accounted for by multiple predictors.

In the simple linear regression model used before we had only used the Math Formative Assessment, which correlated with the state assessment at .54 and thus accounted for 29% of the total variance, $r^2 = .29$. In the multiple regression example shown here, the formative math assessment still correlates at .54, but we now have the Reading Formative Assessment that correlates at .55. The R^2 value takes the combined effect of these two variables to calculate the total variance accounted for in the dependent variable, which in this case is .42. We would thus conclude that 42% of the variance in students' state math assessment performance is accounted for by the combined effects of both formative assessments. Note that there is overlapping variance among the predictor variables. In other words, the addition of a new variable does not bring entirely new information. Figure 7 depicts this relation in a Venn diagram. The top circle represents the total variance in the dependent variable – state test scores. The two bottom circles represent the variance associated with each independent variable. Taken individually, the Math Formative Assessment would account for the green and blue variance within state test scores, while Reading Formative Assessment would account for the red and blue variance. Together, each variable accounts for a unique portion of the overall variance (green and red for math and reading respectively), but also have shared variance (blue). The R^2 value in multiple regression represents, conceptually, the green, blue and red portions of the total variance.

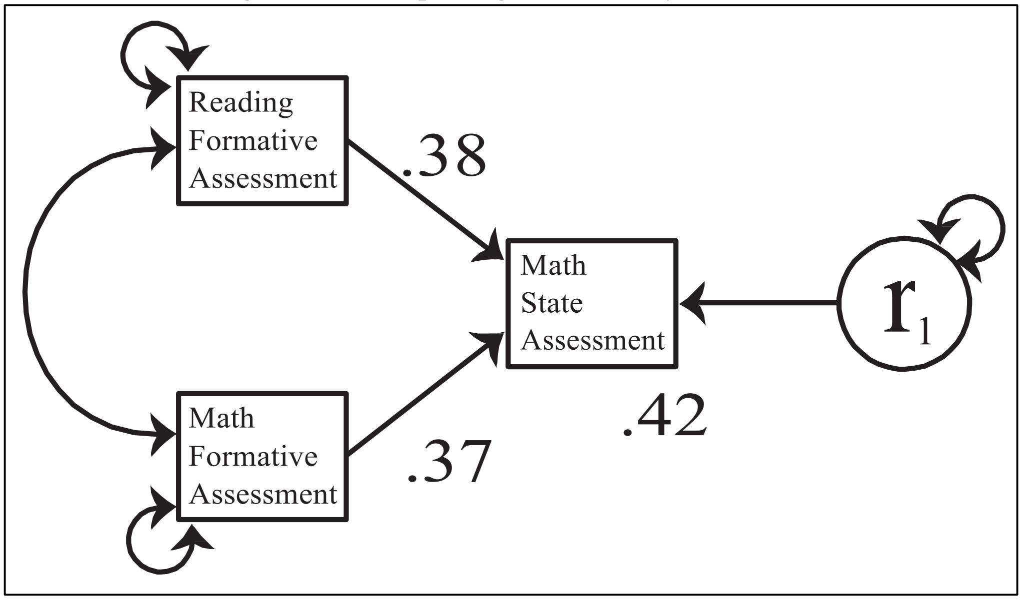
Figure 7 – *Shared variance accounted for by multiple predictor variables*



Path diagrams also can be used to display results of models. Figure 8 displays the results of the multiple regression model above. Generally, path diagrams are displayed with the standardized results (as in this case). When researchers apply more complex models they will often present the results with the path diagram, but for multiple regression analyses this generally is not the case. However, it may help to conceptualize the statistics by viewing them through the path diagram. The path coefficients are analogous to the standardized beta weights (i.e., the standardized regression coefficients) of regression, .

R^2 values are important statistics in any model, but are not always displayed on the path diagram in more complex models. Because the multiple regression model is fairly simple, the single R^2 value is displayed just below the dependent variable (.42). Thus, the path diagram displays the correlation between the predictors, the standardized regression weights, and the variance in the dependent variable accounted for by the predictors. We can also conclude that 58% of the variance in the dependent variable is represented in the residual term (or unaccounted for by our model), given that 42% of the variance is accounted for by the predictor variables (1.00 - .42 = .58).

Figure 8 – *Multiple regression analysis results*



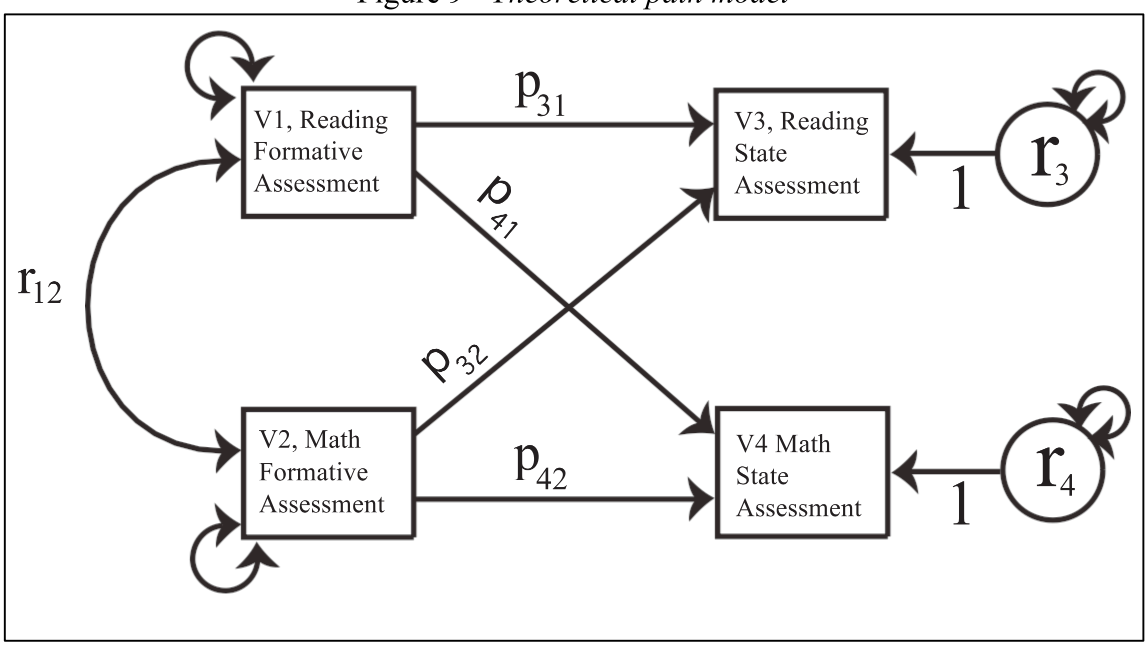
Path Analysis

Path analysis considers the correlation between variables based on the theory expressed in the path diagram. In theory, one can use path analysis to hand-compute the standardized regression weights (i.e., the standardized values for the directional arrows) of a path model by using the observed correlations between variables. In order to do so, one would need to follow a set of rules for path analysis.

Every SEM model consists of two types of variables: exogenous and endogenous. Exogenous variables are those that do not have straight path arrows pointing at them, and endogenous variables are those that do. In other words, exogenous variables are independent variables that are not influenced by other variables in the model, and can be either observed or latent (unobserved) variables. For example, in Figure 8 above, the Math Formative Assessment and Reading Formative Assessment variables are exogenous, and the Math State Assessment variable is endogenous. An easy way to remember the distinction is to think back to functions in your algebra courses, in which inputs of the function are often referred to as “x” and outputs are referred to as “y;” thus, exogenous variables are akin to “x” variables, and endogenous variables are then akin to “y” variables. We can take this example a step further to define the equation that describes an endogenous, or outcome variable. Figure 9 below shows a path diagram with two exogenous variables (V1: Reading Formative Assessment, and V2: Math Formative Assessment) and two endogenous variables (V3: Reading State Assessment, and V4: Math State Assessment).

This diagram represents the relation between the variables based on theory. This is not the only way in which the path diagram can be expressed, rather the diagram expresses a researcher’s theory about how the variables interrelate. In this case, the path diagram represents a theoretical model in which students’ performance on a Reading Formative Assessment (exogenous variable) has a direct effect on their state test performance in both reading and math (both endogenous). Similarly, how students perform on the Math Formative Assessment (exogenous) has a direct effect on their state test performance in reading and math (both endogenous). Note that a plausible alternative model may remove the paths crossing subject areas (i.e., math only effects math and reading only effects reading).

Figure 9 - *Theoretical path model*



Note that all the directional arrows are labeled as paths, *p*, and denote where the path begins and where it ends. For example, p_{31} indicates the path that began at V1 ended at V3. Also note that the double-headed curved arrow labeled r_{12} indicates the *correlation* between V1 and V2. We can use this notation to describe the regression-type equation that expresses each endogenous variable as a function of all elements having a direct affect on it (where a direct affect is represented by a directional arrow).

$$V3 = p_{31}V1 + p_{32}V2 + r_3 \qquad \text{and} \qquad V4 = p_{41}V1 + p_{42}V2 + r_4$$

One can derive these correlations by following Sewall Wright's Standardized Path Tracing Rules (Wright, 1934), which state that a valid path tracing:

1) Does not loop through the same variable twice.
2) Allows going through only one double-headed curved arrow per trace.
3) Can go forward (with) or backward (against) a directional arrow, but once forward tracing has begun, it cannot then go backward.

A path diagram implies correlations between two variables. Based on these model-implied correlations, the observed correlations between the variables, and the three tracing rules, we can use path analysis to hand-compute the standardized regression weights of a path model. The table below displays the observed correlations (right portion of table) along with the model-implied correlations (left portion of table). Note that at this point the model-implied correlations are the path tracing rules, which we use to compute actual values.

Table 1 – *Model-implied and observed correlations*

Model-Implied Correlations					Observed Correlations				
	V1	V2	V3	V4		V1	V2	V3	V4
V1	1				V1	1			
V2	r_{12}	1			V2	.48	1		
V3	$p_{31}+ r_{12}p_{32}$	$p_{32}+ r_{12}p_{31}$	1		V3	.67	.47	1	
V4	$p_{41}+ r_{12}p_{42}$	$p_{42}+ r_{12}p_{41}$	0	1	V4	.55	.54	.67	1

To illustrate an example, let's decompose the paths from V1 to V3; there are two ways we can do this. We can go directly from V1 to V3 (p_{31}), and we can go from V1 through V2 to V3, in which case we multiply the two separate paths ($r_{12}p_{32}$). To get the entire path for V1 to V3, we add the paths of the two tracing methods together to get $p_{31} + r_{12}p_{32}$. We can now use a series of algebraic equations to solve for each of the path, p, coefficients. As an example, we use just two of the equations to derive p_{41} and p_{42}. We begin by setting the results of our path tracing rules (i.e., model-implied) equal to the observed correlations.

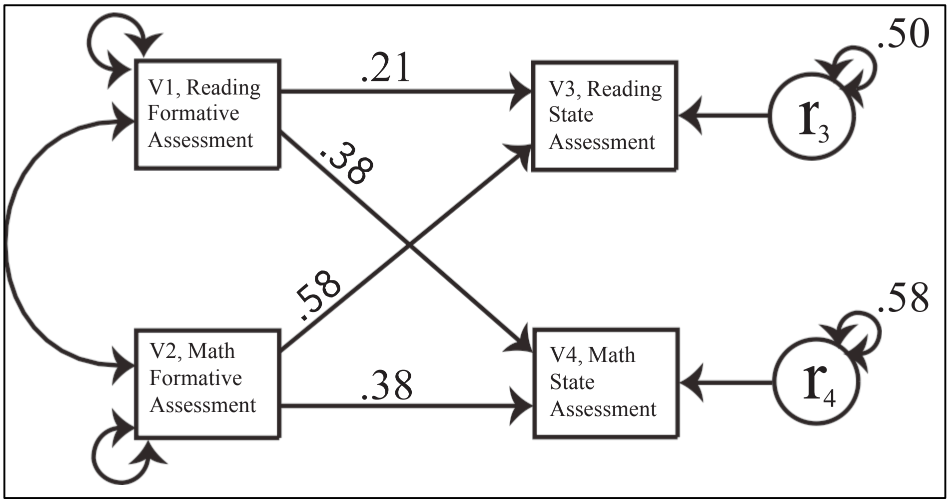
Given: (a) $p_{41}+ r_{21}p_{42} = .55$ and
(b) $p_{42}+ r_{21}p_{41} = .54$

1) $p_{41} + .48p_{42} = .55$	given by the observed correlations
2) $p_{41} = .55 - .48p_{42}$	rearrange the terms
3) $p_{42} +.48(.55 - .48p_{42}) = .54$	substitute equation (a) into equation (b)
4) $p_{42} + .26 - .23p_{42} = .54$	solve for p_{42}
5) $.77p_{42} = .28$	
6) $\boldsymbol{p_{42} = .36}$	
7) $.36+.48p_{41} = .54$	substitute solution for p_{42} into equation (b)
8) $.48p_{41} = .18$	solve for p_{41}
9) $\boldsymbol{p_{41} = .38}$	

Thus, the standardized regression weight for p_{41} = .38 and the standardized regression weight for p_{42} = .36. Assuming the model is correct, we can state that a one standard deviation increase in a student's Reading Formative Assessment (V1) leads to, on average, a .38 standard deviation increase in his or her state math assessment (V4), controlling for formative math scores. Similarly, a one standard deviation increase in the student's Math Formative Assessment (V2) leads to, on average, a .36 standard deviation increase in his or her state math assessment (V4), controlling for Reading Formative Assessment. Thus, both the Reading Formative Assessment and the Math Formative Assessment are almost equal in their capacity to explain performance on the State Math Assessment (.38 versus .36, respectively).

We could repeat this same procedure for both p_{31} and p_{32} to estimate the effect of the formative reading and math assessments on the state reading assessment. Instead, let's compare our hand-calculated results to those offered by statistical software, in this case, M*plus* version 5.21 (Muthén & Muthén, 1998-2010), shown in Figure 10 below.

Figure 10 – *Path Analysis Results*



Note that our hand-calculated estimates for p_{41} and p_{42} are slightly different than the results in Figure 10. These small differences are due to rounding error (i.e., as we rounded to the hundredths' place, the software carries out computations to many more points past the decimal).

There are several ways in which to interpret the results. The first is the same way in which we interpreted our hand-calculated results, as we interpret all regression results. For example, a one standard deviation increase in formative reading score (V1) leads to, on average, a .21 standard deviation increase in the state reading assessment (V3), holding all else constant. We can do this for all path coefficients, and because these results are standardized, we can compare the coefficients to one another. For example, it is surprising that the affect of Math Formative Assessment on the Reading State Assessment is greater than that of the effect of Reading Formative Assessment. Similarly, both formative assessments have about the same effect on the Math State Assessment. It is left to the researcher to theorize why this might be.

In Figure 10, you will also note the residual variances of our endogenous variables: .50 for state reading and .58 for state math. These are the standardized estimates for each outcome variable of what the model did not account for, that is, the residuals. Conversely, their inverses are the estimates of what the model *did* account for, so that $1 - r_y$ is the proportion of variance accounted for in that endogenous "y" variable. You might recognize these as R^2 estimates. Thus, 1 – .50, or .50, is the proportion of variance (50%) accounted for in state reading by the model, and 1 – .58, or .42, is the proportion of variance (42%) accounted for in state math test scores by the model. One of the potential purposes of a path model, or regression model, is to account for variance in the outcome variable(s). This is one indicator of how well the model "fits" the data, and there are other fit indices as well. One thing to note in Table 1 is that the model implies that there is no correlation between V3 and V4 – that the correlation equals zero. Had we theorized a correlation between these two endogenous variables, we would have drawn a double-headed curved arrow between the two. Also note that the observed correlation between V3 and V4 was .67, which is a fairly strong correlation. There are likely implications with fixing this correlation to zero, rather than specifying it in the model, and one of those implications involves model fit.

Goodness of fit. In general, the goodness of fit (GOF) of a model indicates how well the theorized model aligns with the observed data. There are several GOF indicators, each taking a different approach to estimating model fit. The goal of testing fit is to find little difference between the theorized model and the empirical (found) model. This effort is opposite of typical hypothesis testing in which a null hypothesis is used to test the probability of finding an outcome listed in the alternative hypothesis. Here, when we find no difference, the theory is supported.

One class of fit indices compares the degree of correspondence between the observed and model-implied variance-covariance matrices (remember, the unstandardized version of Table 1), and includes the model χ^2 statistic (or chi-squared) and the standardized root mean squared residual (SRMR). The model χ^2 statistic, if significantly different than zero ($p < .05$), indicates that there is a statistically significant difference between the observed and model-implied variance-covariance matrices, but this statistic is affected by the sample size so that the larger the sample the more likely it will be that significant differences will be observed, indicating poor fit. The χ^2 statistic, though routinely reported, is not the most reliable GOF indicator because it runs counter to good research practice. Generally we want very large sample sizes so our observed data approximates normal distributions and we can more adequately estimate the standard errors of our estimates. With the χ^2 statistic, however, large sample sizes decrease the chances of obtaining adequate model fit. Given these limitations, other fit indices have been developed. The SRMR, for example, expresses the average difference between the observed and model-implied correlations; generally values of .08 or less indicate adequate fit (Hu & Bentler, 1999).

Another class of fit indices evaluates the overall discrepancy between the observed and model-implied variance-covariance matrices while also taking into account the model's simplicity (i.e., adjusting for the number of parameters estimated). These indices represent improvements as more parameters with useful contributions are added to the model. Among them are Akaike's information criterion (AIC), Bayesian information criterion (BIC), and the root mean squared error of approximation (RMSEA). For the AIC and BIC, smaller values indicate better fit, but the magnitude is not directly interpretable. These criteria are often used to compare one model to another and therefore are frequently referred to as incremental fit indices. For the RMSEA, a value of .06 or less generally indicates adequate fit (Hu & Bentler, 1999).

A third class of fit indices evaluates the model fit relative to a model that specifies no relations between the variables; these include the comparative fit index (CFI) and the Tucker-Lewis Index (TLI). A value greater than or equal to .95 indicates good fit for both the CFI and TLI. We can apply these criteria to our model presented in Figure 10 to determine how well the model fits the data.

Table 2 – *Fit Indices*

χ^2	SRMR	AIC	BIC	RMSEA	CFI	TLI
281.639	.07	35243.64	35310.77	.47	.84	.18
$p < .05$				90% CI .42 - .51		

We can evaluate these GOF indices based on the guidelines suggested above. The χ^2 statistic is significant, indicating poor model fit, but again, this is dependent on the sample size and may not be our best indicator; in our example of formative assessments, $n = 1{,}292$ students. Because we are only looking at one model the SRMR value is less than .07, which indicates good fit. The RMSEA is much higher than .06, which indicates poor fit. We have no model by which to compare the AIC and BIC values, so we can ignore them here. Finally, neither the CFI nor the TLI are greater than .95, indicating poor model fit. Thus, as sometimes happens, we have conflicting GOF indicators. It is generally considered best practice to report one or more indicators from each class to determine model fit. Here, because most of our indices suggest our model fits the data poorly, the correct interpretation would be that this model does not fit the data well.

To this point in the paper, we have only dealt with variables we can directly observe or measure. Often, however, we would like to estimate a value for an unobserved or latent variable. We can do so by treating multiple observed variables as “indicators” of the latent trait of interest in a measurement model. We can then test the adequacy of the measurement model through confirmatory factor analysis.

Confirmatory Factor Analysis (CFA)

Confirmatory factor analysis (CFA), a special form of factor analysis, is used to test whether measures of a construct are consistent with a researcher's understanding of the nature of that construct (or factor). CFA is a confirmatory technique to investigate the relations between the observed and latent, unobserved variables, here referred to as factors. Guided by theory, researchers can specify the number of factors (unobserved variables), the number of observed variables that "load" on the factors (i.e., contribute towards), and the associated residuals. Figure 11 below displays the theoretical framework for a one-factor CFA model, Math Ability, which is theorized to influence or predict students' scores on the observed variables *Number & Operations*, *Geometry*, and *Algebra*. This figure describes the latent variable as measured by three observed variables, and each of the observed variables has a corresponding residual, r_n. For the model to be identified, either the variance of the latent variable or one of the paths (or factor loadings) from the factor to an observed variable need to be fixed. In Figure 11, the variance has been fixed at 1.0 so that each path can be freely estimated. The choice of fixing a path over a variance is largely arbitrary and has to do in part with the research questions. If the researcher is interested in the unstandardized estimates of all paths then he or she may fix the variance of the factor. Similarly, if the researcher is interested in the unstandardized estimates of the variances he or she may choose to fix a path.

Just like in other forms of regression, the results from CFA can be presented in standardized or unstandardized form. In many cases, standardized results are easier to interpret because all the estimates have been placed on the same scale, and thus are directly comparable, so it is easier to compare the effect of each variable. In CFA, the standardized loadings are equivalent to the correlations between the observed variables and latent factor. Theoretical CFA models are shown in Figure 11 for both reading and math. For reading, the model displayed has negative degrees of freedom.

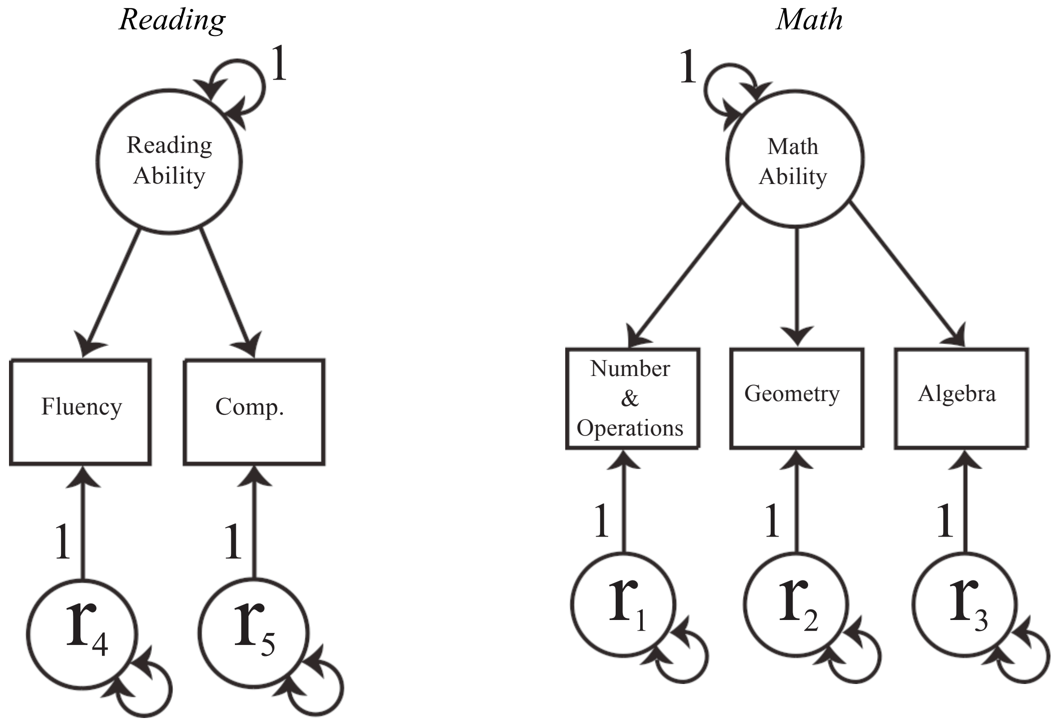
For models to be estimated they must have greater than or equal to 0 degrees of freedom. The degrees of freedom of any SEM model are equal to the number elements in the variance-covariance matrix used to estimate the model minus the number of parameters being estimated –

regression weights, variances, and covariances. The formula for calculating the number of elements in the variance-covariance matrix is $m(m + 1) / 2$, where m = the number of observed variables (i.e., the number of boxes in the path diagram). In the theoretical reading model displayed in Figure 11 below there are two observed variables and thus $2(2 + 1) / 2 = 3$ elements in the variance covariance matrix. There are 7 total parameters – 3 variances and 4 path coefficients – but 3 of these are fixed at 1.0. There are thus 4 parameters to be estimated and the model, as displayed, has 3 – 4 = -1 degrees of freedom.

To make the model estimable, we must place further constraints on the model to obtain more degrees of freedom. One way to do this is to constrain the residual terms to be equal, thus necessitating the estimation of one residual parameter instead of two. The approach is less than ideal and when employed in the current model led to results that suggest the model does not adequately measure students' reading ability. Generally one should aim to have at least three observed variables for any CFA model, which allows for six degrees of freedom and greater flexibility in specification.

Note as well that model fit indices are only produced when the model is “over identified”, which means the model has at least 1 degree of freedom. Models with zero degrees of freedom are generally referred to as “just identified”. Just identified models have only one possible solution, and thus always fit the data perfectly because there are no other possible solutions. Over identified models have multiple possible solutions, and thus must be evaluated for how well the converged solution fit the observed data. Model fit is a fundamental concern of SEM and CFA models, and when the model is just identified the fit of the model cannot be evaluated.

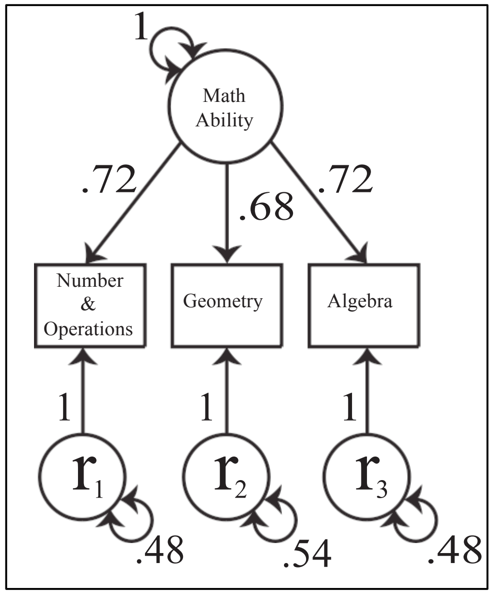
Figure 11 Theoretical *One-Factor Confirmatory Factor Analysis, Reading*



Given the problems in estimation with the reading model, we present here only the results for math, which had 6 – 6 = 0 degrees of freedom. The results for the math model are displayed in Figure 12 below. The correlations (loading) between the latent *Math Ability* factor and the *Number and Operations*, *Geometry*, and *Algebra* factor indicators are relatively large at .72, .68, and .72, respectively. The Math Ability latent factor is thus quite strongly associated with the three observed variables. In other words, the observed variables are adequately measuring our targeted latent construct. The residual variances (unexplained variances) of the *Number and Operations, Geometry,* and *Algebra* factor indicators were .48, .49, and .52, respectively. We can determine the amount of variance "explained" or "accounted for" in each latent factor by computing the R^2 value, which is equal to 1- residual (or, the unexplained variance). Thus, in our model Math Ability explained 52%, 51%, and 48% of the variance in students' responses to the

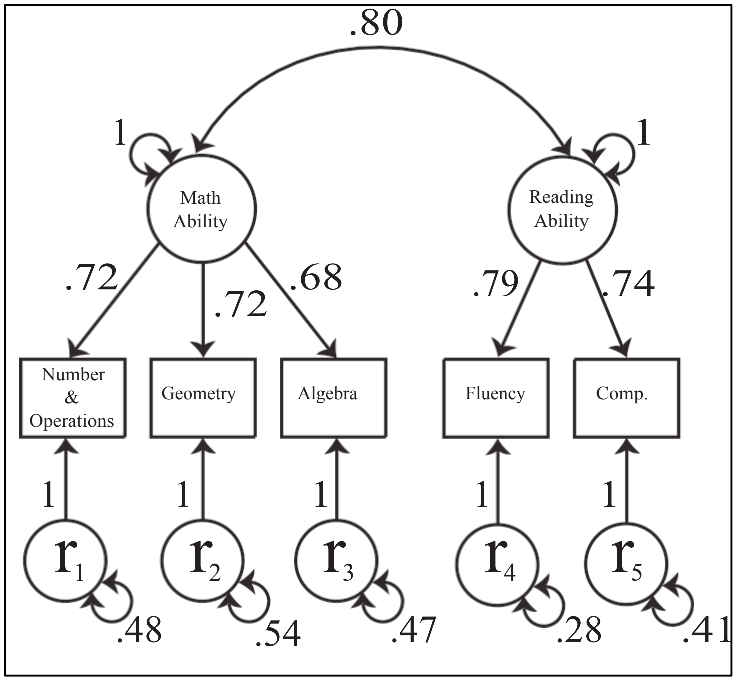
Number and Operations, Geometry and *Algebra*, tasks, respectively. Fit statistics could not be produced due to the model being just identified.

Figure 12 *One-Factor Confirmatory Factor Analysis Results, Math*



Researchers applying CFA are also able to test multiple latent factors simultaneously. For example, if we have two latent factors, *Math Ability* and *Reading Ability*, we can test both factors under one CFA model. Figure 13 below shows the results of a two-factor CFA model, where the correlation between the latent variables is estimated, represented by the double-headed curved arrow between the latent factors. The two-factor CFA model is interpreted just like a one-factor CFA model. In this example, the correlation between Math Ability and Reading Ability is 0.80, which indicates a strong correlation between latent factors. The model has 15 – 11 = 4 degrees of freedom and thus was over-identified and fit statistics could be calculated. Results suggest the model fits the data quite well, with a CFI of .996, TLI of .990, RMSEA of .038, and an SRMR of .012 – all meeting the Hu and Bentler (1999) fit criteria described in the previous section. Again, keep in mind that models should be based on theory, and specify the number of factors, the number of observed variables within factors, the correlation(s) between latent factors, and any residual *before* the parameters are estimated.

Figure 13 –*Two-Factor Confirmatory Factor Analysis Results*



CFA generally requires large sample sizes to adequately estimate parameters. When running CFA, many different fit statistics are used to help determine whether the model provides adequate fit for the data. Many of the rules of interpretation of model fit and model modification in either path analysis or SEM apply equally to CFA (e.g., RMSEA, SRMR, CFI, TLI, AIC, BIC, or ABIC; note that all fit statistics and interpretations are discussed in the *Path Analysis* section). If the fit statistics are acceptable, the parameter estimates can be examined. It is important to note that the parameter estimates should not be interpreted if the model does not fit the data well, as the individual path coefficients cannot be trusted if the model as a whole does not fit the data. Additionally, correlations between the factors should be checked to see how the factors relate to each other. If correlations are sufficiently high, consolidating the corresponding factors into a single factor should be considered, because the factors overlap to such an extent that they may be redundant.

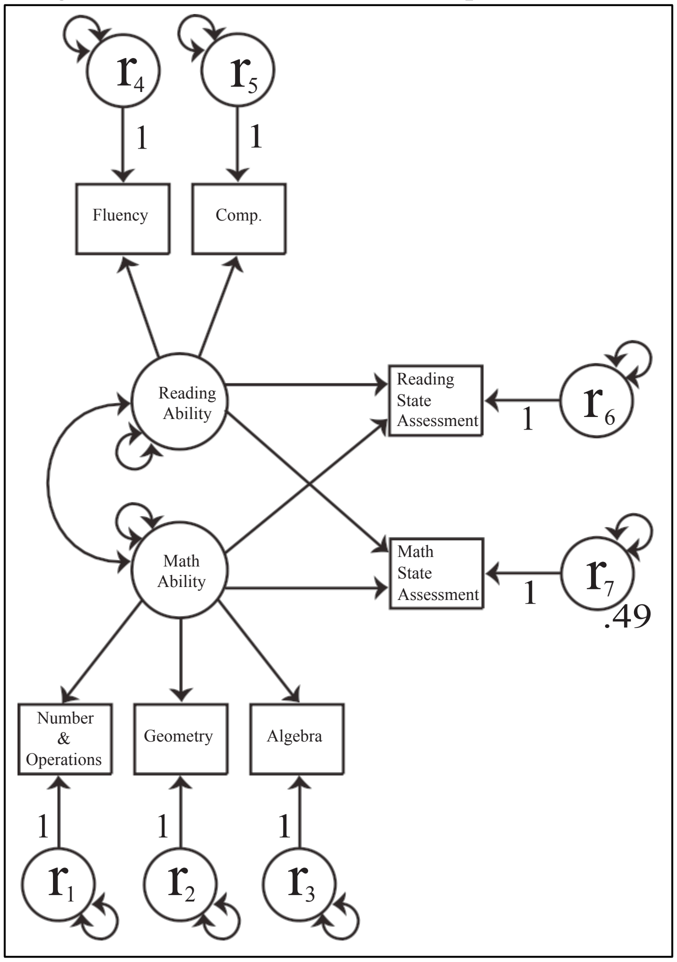
Structural Equation Modeling (SEM)

CFA is frequently used as a first step to assess the proposed measurement model in a full structural equation model (SEM). In the context of SEM, the CFA portion of the model is referred to as "the measurement model," while the relations between variables (with directional arrows) is called "the structural" portion of the model. One cannot estimate directional relations under the CFA model; in general, SEM is used when researchers would like to test the relations between several observed and unobserved variables. Although researchers can apply regression, path analysis, or CFA to many research problems in education and psychology, SEM is by far the most flexible. In many cases, regression analysis, factor analysis, and path analysis represent special cases of SEM; however, SEM is more general than regression, because, in particular, a variable can act as both an independent and dependent variable in the same model. CFA is distinguished from SEM by the fact that the factors are only related (double-headed curved arrows) and not theorized to directly affect each other (directional arrows). Additionally, researchers using SEM can model mediating variables, structural residuals, and multiple dependent variables (see Baron & Kenny, 1986; Kline, 2010).

The structure of full SEM has two main parts, both of which we have been discussed in isolation: a measurement model (CFA) and a structural model (path analysis). The measurement model is a multivariate regression model (multivariate generally implies multiple variables being analyzed simultaneously) describing the relation between a set of observed dependent variables (observed variables) and a set of continuous latent variables (unobserved, theoretical constructs). The observed dependent variables are often referred to as factor indicators and the continuous latent variables are referred to as factors. In SEM, both exogenous (independent, or “upstream”) and endogenous (dependent, or “downstream”) variables can be observed or unobserved, depending on the model being tested. Within the context of structural modeling, exogenous variables represent constructs that theoretically influence other constructs under study and are not influenced by other factors in the model. Endogenous variables are theoretically influenced by exogenous variables, and, depending on the model specified, other endogenous variables in the model.

As an example of a full SEM model, we can easily replace the observed math and reading variables in Figure 9, with the measurement models displayed in Figures 11 and 12. This model is shown in Figure 14 below. This figure displays the theoretical framework with two correlated unobserved (latent) variables. Analogous to the model displayed in Figure 9, the path diagram indicates that students’ Reading Ability (exogenous variable) has a direct effect on their state test performance in both reading and math (both endogenous). Similarly, students’ Math Ability (exogenous variable) has a direct effect on their state test performance in both reading and math (both endogenous). Perhaps the primary advantage of the model displayed in Figure 14 over the model displayed in Figure 9 is that, because each indicator of the latent factors already includes a residual term, the latent factor is theoretically free of measurement error (i.e., the measurement error has already been accounted for). Thus, the latent factors may more adequately represent students’ math or reading ability, which is being used to explain variance in the state assessment scores, over the raw observed score. In sum, full SEM can be viewed as a combination of path analysis and CFA.

Figure 14 *Theoretical Structure Equation Model*

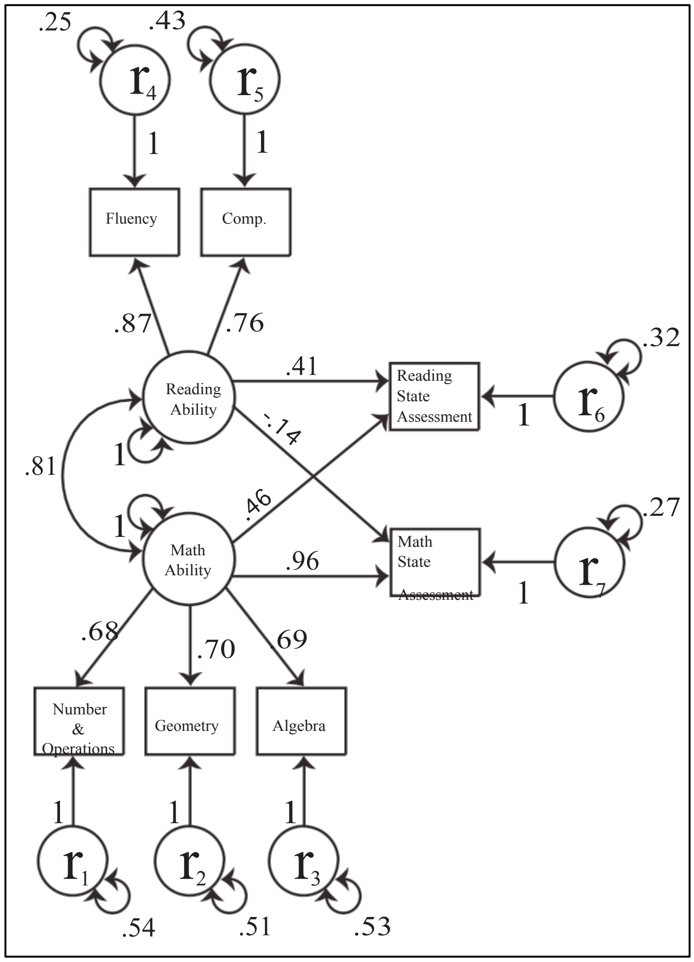


The results of the SEM model showed that all fit indices were relatively good. The CFI was .972, the TLI was .951, the RMSEA was .082, and the SRMR was .023. The model thus fits the data adequately, but is not great. The standardized results are presented in Figure 15 below. The correlation between the two latent factors was large at .81, indicating that students Reading and Math Abilities were strongly related. The standardized factor loading (correlations) between the Math Ability latent factor and the *Number and Operations, Geometry,* and *Algebra* indicators were quite large at .68, .70, and .69, respectively. Thus, the Math Ability latent factor and the three observed variables were strongly associated. The correlations between the Reading Ability latent factor and *Fluency* and *Comprehension* are also large at .87 and .76, respectively, again indicating a strong association. The residual variances of the *Number and Operations, Geometry, Algebra, Fluency,* and *Comprehension* were .54, .51, .53, .25, and .43, respectively.

The interpretation of standardized SEM results is similar to the interpretation of standardized path analysis results; each one standard deviation increase in Math Ability produced, on average, a .96 and .45 standard deviation increases in *State Math Assessment* and *State Reading Assessment* scores, respectively, holding all else constant. Because these results are standardized, the coefficients can be directly compared. These results indicate that Math Ability has a much stronger effect on *State Math Assessment* than on the *State Reading Assessment* (.96 vs. .45). This is reasonable, as we expect students with higher mathematical competency to perform higher on the state math assessment, but we might not predict the same about the effects on *State Reading*, given that the skills are not closely related.

For reading, a one standard deviation increase in Reading Ability produce, on average, a -.14 and .41 standard deviation increase in *State Math Assessment* and *State Reading Assessment* scores respectively, holding all else constant. The effect of Reading Ability on the *State Reading Assessment* is perhaps not as strong as we expected (.41), given that each variable represented similar constructs. On the other hand, there is a negative effect of Reading Ability on the *State Math Assessment* (-.14), which is surprising. Interestingly both Math Ability and Reading Ability have a similar effect on *State Reading Assessment* (.45 vs. .41). Also, the residual variances of *State Math Assessment* and *State Reading Assessment* is .27 and .32, respectively, indicating that our model explains a good proportion of variance in these outcomes measures (.73 and .68, respectively).

Figure 15 – *Full Structure Equation Modeling Results*



Conclusion

Deeper inspection of the measures used, the sample investigated, or other mediating or moderating variables could help to explain some of the unexpected results. However, the primary purpose of this paper is to illustrate the technique of SEM and its interpretation through an example. With that in mind, it is perhaps most helpful to view the differences in the model results between Figure 10 and Figure 15. Although essentially the same data are used, and the theoretical framework is identical, different statistical models are specified and the resulting inferences based on each model are substantively different. The differing results highlight the importance of correct model specification. For instance, while the relation between the formative reading assessment and the state reading assessment is modest in the path analysis (.38), the relation is actually slightly negative in the full SEM model (-.14). Further, while the negative relation is difficult to interpret, the model fit statistics suggest that the full SEM model more adequately represents the observed relations among the variables. It should also be noted, however, that only one measure of reading competency is included in the path analysis model, while two are included in the full SEM model, making comparisons between the two models somewhat tenuous. Further, the results of our reading-only CFA is not great and perhaps suggest that we may need to reconceptualize our theoretical reading model before applying it to a more complex SEM model. That is, perhaps some of the unexpected results are due to poor precision in our measurement of the reading ability latent construct. The math measures remain the same between models, with the formative assessment simply representing the sum of the three factor indicators.

The flexibility of SEM offers great potential for researchers to model complex relations. To keep this report concise, we concluded with a relatively simple full SEM model. However, a wide range of theoretical models of considerably greater complexity can readily be specified. Perhaps the greatest strength of SEM is that it allows researchers to not only theorize about the ways in which variables interrelate, but to explicitly test the relations. In the end, SEM affords the researcher a rich basis from which to theorize and test the adequacy of complex models representing the real world. In turn, this flexibility helps us better understand and provide solutions to pressing problems.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropoligical Institute of Great Britain and Ireland, 15*, 246-263.

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

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (Third ed.). New York: Guilford Press.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (Sixth ed.). Los Angeles, CA: Muthén & Muthén.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statisics, 5*, 161-215. doi: 10.1214/aoms/1177732676