

**® USING AMOS GRAPHICS TO ENHANCE THE UNDERSTANDING AND COMMUNICATION OF MULTIPLE REGRESSION**

CUNNINGHAM, Everarda G and WANG, Wei C  
Swinburne University of Technology  
Australia

*Support for the saying “a picture is worth a 1000 words’ has been consistently found within statistics education. Graphical images are effective in promoting understanding and communication of statistical concepts and results to a variety of audiences. The computer software package, AMOS, was developed for the analysis of structural equation models (SEM) and has a user-friendly graphical interface. However, courses in SEM are generally found only at the postgraduate level. This paper argues that the graphical interface of AMOS has the potential to enhance conceptual understanding and communication of results in undergraduate statistical courses. More specifically, approaches to the teaching and communication of results of multiple regression models when using SPSS and AMOS will be examined and compared.*

**INTRODUCTION**

Many undergraduate students, especially in the areas of business and the social sciences, come to the study of statistics with a wide range of academic backgrounds, learning styles and motivation (Biktimirov & Nelson, 2003). For these students, quantitative statistical methodologies and analyses are frequently considered as esoteric methods that are difficult to understand and use, yet integral to their studies or research proposals. This reality poses a challenge to lecturers to communicate statistical concepts and methodologies in a variety of ways that maximize student engagement and understanding. One strategy that is still underutilized as a communication tool for statistical understanding is visual learning. Visual learning modes potentially reach many more students, especially those whose preferred learning style is nonverbal (Biktimirov & Nelson, 2003). Following a brief discussion of the benefits of visual learning and features of the *AMOS* graphical interface, this paper then illustrates how visual learning through the utilization of *AMOS* graphics can be used to enhance the understanding of multiple regression, a topic commonly taught in later years of undergraduate statistics courses.

Research emanating from both education and cognitive psychology suggests that visual learning is among the very best methods for teaching students how to think, learn and communicate. Lynch (1985) viewed knowledge that was represented in the form of graphics and diagrams as of fundamental importance to learning and communication. Visual learning allows students to clarify concepts and to process, organize, and articulate information through graphical ways of working with presenting information (Biktimirov & Nilson, 2003). Visual diagrams readily reveal the patterns, interrelationships and interdependencies of models under investigation and such models, if thoughtfully constructed, become a means of communicating not only conceptual models but also the results from statistical analyses to broad audiences. In fact, Seeger (2004) believed that thinking of knowledge without thinking about how it is represented is very difficult, if not impossible, for most people.

Over the last two decades, computer programs that were specifically designed for the purpose of analyzing structural equation models (SEM) have emerged. A model in this sense is essentially “a statistical statement about the relations among variables” (Hoyle, 1995, p. 2). SEM is a multivariate analysis technique that encompasses standard methods such as multiple regression. However, SEM has the additional advantage of modeling relationships between latent variables after accounting for measurement error (Bollen, 1989). Indeed, much of the popularity of SEM lies in the ease with which more recent derivations of computer programs (e.g., *LISREL 8.5*, *AMOS 5*, *EQS 6*) allow users to estimate parameters and test hypotheses through graphical interfaces. Of these programs, the *AMOS 5* graphical interface is the most user-friendly because it

requires little more than the ability “to drag and drop boxes (manifest variables) and eggs (latent variables) on the screen” (Nachtigall, Kroehne, Funke, & Steyer, 2003, p.12).

Within *AMOS* 5 graphics, users draw a path diagram directly on the screen. The observed variables are represented with rectangles and linear equations are represented with arrows from the independent to the dependent variables. In this sense, the pre-cursor to the use of *AMOS*, excluding basic computer skills, involves knowledge of the theory of the variables under investigation and the translation of that theory into a set of linear regression equations that are simply represented by arrows connecting the variables (Nachtigall et al., 2003). Essentially, *AMOS* accepts a path diagram as the model specification and provides drag-and-drop drawing tools that allow rapid model specification in intuitive and user-friendly ways. The researcher then gathers data on the variables of interest and attaches the data set to the *a priori* specified model. After running the model, the parameter estimates are displayed on the path diagram. In other words, the *AMOS* graphic interface also presents the results from analyses in a visual framework that is very easy to understand. In fact the ease of use and wider availability of *AMOS* have enhanced the general view that SEM is a “natural mode of thought” (Steiger, 2001).

The *AMOS* program is frequently used at postgraduate level for the teaching of SEM. However, given the ease of use and the visual learning environment that the *AMOS* graphics interface provides, it is surprising that *AMOS* and similar SEM computer packages that utilize graphical interfaces are rarely considered as a teaching tool that might enhance the communication and understanding of statistical concepts at the undergraduate level. While the model specification starting point for using *AMOS* may not initially seem user-friendly, similar decisions are required with respect to independent and dependent variables in conducting multiple regression analyses within the more traditional framework when using programs such as *SPSS* or *Minitab* (Kühnel, 2003).

Presented first in the following section is an example from educational psychology that compares the approaches of conducting and reporting a standard multiple regression analysis within an *SPSS* environment to that of an *AMOS* graphical environment. The example is then extended to consideration of the teaching of hierarchical regression by introducing a potentially mediating variable. From a didactical perspective, it is expected that visual representations of *a priori* models that are drawn as path diagrams will assist not only in communicating statistical concepts to undergraduate students but will also enhance the communication of the findings from such analyses in ways that are more meaningful to broad audiences.

## ILLUSTRATIVE EXAMPLES

The goal in multiple regression is to predict scores on a dependent or criterion variable on the basis of scores on multiple independent or predictor variables. Researchers are usually also interested in determining the amount of variance in the dependent variable that is explained by the predictor variables (i.e.,  $R^2$ ) and in identifying the most efficient predictors in the multiple regression model (e.g., Aron & Aron, 1994; Francis, 2003). Building on previous knowledge of bivariate regression, a common starting point in the conventional teaching of multiple regression might be to represent such models as a regression prediction equation which, in standardized form for  $i$  variables, might be expressed as:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \dots \dots + \beta_i X_i$$

where  $\hat{Y}$  = the predicted value of the dependent variable  
and  $\beta_{1\dots i}$  = the standardized regression weights derived from the *SPSS* analyses.

A residual or error term is sometimes added to the equation because it is not expected that scores on the dependent variable will be perfectly predicted from scores on the independent variables. In other words, the representation of the regression model would include adding a residual term ( $\epsilon$ ) to the above equation to account for that portion of the variance that is not explained by the independent variables.

## EXAMPLE 1: STANDARD MULTIPLE REGRESSION

Consider a regression model of a type that is commonly presented in undergraduate statistics courses in which the aim is to predict academic achievement on the basis of the degree to which students are connected to their peers, teachers, and school. The regression model consists of three independent variables relating to connectedness to the school environment, namely school, peer, and teacher connectedness, and the single dependent variable of academic achievement. The variables are all measured on a continuous scale with higher scores indicating more of the construct. The regression analysis is then frequently conducted using programs such as *SPSS* in which the model is specified through the correct selection of the dependent and predictor variables. Selected output obtained from the *SPSS* analysis for a sample of 300 secondary students is presented in Tables 1 to 5.

Table 1

*Descriptive Statistics*

	<i>Mean</i>	<i>Std. Deviation</i>	<i>N</i>
Academic achievement	3.79	.75	300
C-peer	3.96	.79	300
C-teach	3.29	.82	300
C-school	2.75	.94	300

Table 2

*Correlations between the selected variables*

		<i>Academic achievement</i>	<i>C-peer</i>	<i>C-teach</i>	<i>C-school</i>
Pearson Correlation	Academic achievement	1.00			
	C-peer	.57	1.00		
	C-teach	.44	.35	1.00	
	C-school	.39	.31	.65	1.00

Table 3

*Model summary statistics*

<i>Model</i>	<i>R</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>Std. Error of the Estimate</i>
1	.63(a)	.40	.39	.59

Table 4

*Analysis of Variance (ANOVA) for regression analysis*

<i>Model</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
1	Regression	66.93	3	22.31	64.43	.00(a)
	Residual	102.49	296	.35		
	Total	169.42	299			

Table 5

*Standardized and unstandardized regression coefficients*

<i>Model</i>		<i>Unstandardized Coefficients</i>		<i>Standardized Coefficients</i>	<i>t</i>	<i>Sig.</i>
		<i>B</i>	<i>Std. Error</i>	<i>Beta</i>		
1	(Constant)	1.20	.19		6.22	.00
	Connectedness-peers	.44	.05	.46	9.44	.00
	Connectedness-teachers	.18	.06	.20	3.30	.00
	Connectedness-school	.10	.05	.12	2.05	.04

a Dependent Variable: Academic achievement

Based on the output presented in Tables 1 to 5, a summary report might then be produced. For example:

The regression analyses revealed that student perceptions of their peers, teachers, and school explained 40 percent of the variance in academic achievement scores. While all of the connectedness factors were statistically significant predictors of academic achievement, connectedness to peers ( $\beta = .46$ ) was the strongest predictor of academic achievement.

Now consider the same example when the regression analysis is conducted within the *AMOS* graphics environment. The initial step in this instance is the specification of the regression model as a path diagram (see Figure 1).

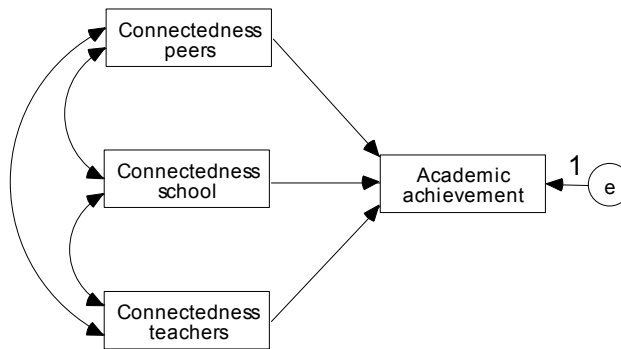


Figure 1. Multiple regression model specification using *AMOS*.

Importantly, to be correctly specified, the model must include a residual or error term (e) as an ellipse which indicates a latent or unmeasured variable. The path diagram specification of the model also highlights that the predictor variables are correlated in the model. Once the model is correctly specified (i.e., drawn using the drop and drag tools in *AMOS*), the same *SPSS* data file for 300 high school students is attached to the model prior to running the model. The standardized and unstandardized parameter estimates derived from the analyses may then be viewed in visual form on the graphic interface as shown in Figure 2.

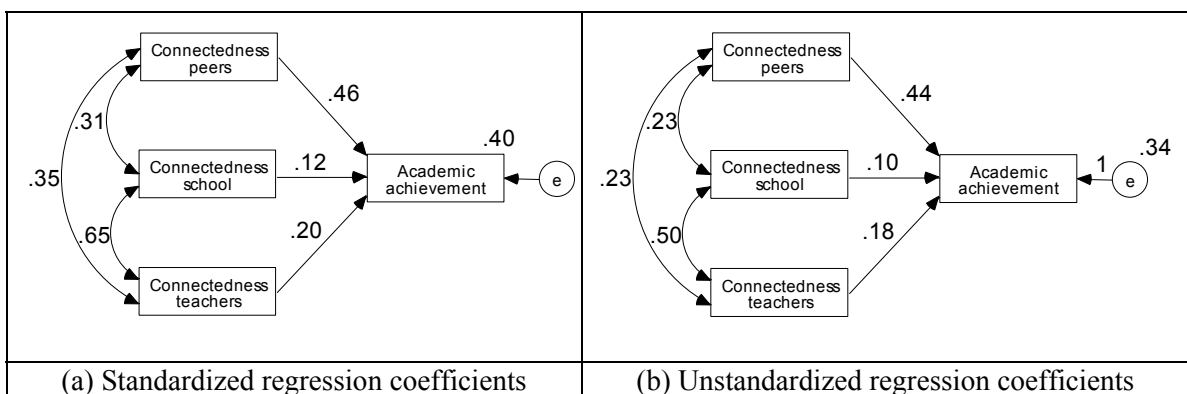


Figure 2. Standardized and unstandardized regression coefficients in predicting school achievement from perceptions of connectedness using *AMOS*.

The parameter estimates shown in Figure 2 may be directly compared to the output from the *SPSS* analysis shown in Tables 1 to 5. The standardized solution (see panel (a) in Figure 2) displays the identical correlations to those shown in Table 2 and the standardized regression coefficients are the same as the *Beta* values in Table 5. The  $R^2$  value of .40 displayed above the

rectangle for academic achievement in Figure 2 equates to the value found in Table 3. Similarly, the unstandardized regression coefficients displayed in panel (b) of Figure 2 are the same as those reported in the Table 5. Panel (b) of Figure 2 also displays the covariances (i.e., the products of pair-wise correlations and their respective standard deviations) between the predictor variables. More importantly, Figure 2 presents the derived parameter estimates of the multiple regression analysis in a visual form that is explicit about the model under consideration. The inclusion of panel (a) in Figure 2 would complement a text-only report and enhance the communication of the findings for both students and the community at large.

#### EXAMPLE 2: HIERARCHICAL REGRESSION

In hierarchical regression, the predictor variables are entered in sets of variables according to a pre-determined order that may infer some causal or potentially mediating relationships between the predictors and the dependent variable (Francis, 2003). Such situations are frequently of interest in the social sciences. The logic involved in hypothesizing mediating relationships is that “the independent variable influences the mediator which, in turn, influences the outcome” (Holmbeck, 1997, p. 600). However, an important pre-condition for examining mediated relationships is that the independent variable is significantly associated with the dependent variable prior to testing any model for mediating variables (Holmbeck, 1997). Of interest is the extent to which the introduction of the hypothesized mediating variable reduces the magnitude of any direct influence of the independent variable on the dependent variable.

To illustrate, consider an extension of the previous example in which self-efficacy, namely the belief that one has the skills and strategies to affect desired behaviours, is hypothesized to mediate or partially mediate the relationships between the school environment connectedness variables and academic achievement. To test these hypotheses, conventional hierarchical regression analyses would enter the set of connectedness variables as a block of variables in the first step of the regression and then enter self-efficacy in the second step of the regression procedure. Selected *SPSS* output from this analysis is shown in Tables 6 to 9.

Table 6

##### *Model summary statistics*

<i>Model</i>	<i>R</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>Std. Error of the Estimate</i>
1	.63(a)	.40	.39	.59
2	.67(b)	.45	.44	.56

Table 7

##### *Further model summary statistics*

<i>Model</i>	<i>Change Statistics</i>				
	<i>R Square Change</i>	<i>F Change</i>	<i>df1</i>	<i>df2</i>	<i>Sig. F Change</i>
1	.40	64.43	3	296	.00
2	.05	27.03	1	295	.00

Table 8

##### *Results from Analysis of Variance (ANOVA)*

<i>Model</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
1	Regression	66.93	3	22.31	64.43	.00(a)
	Residual	102.49	296	.35		
	Total	169.42	299			
2	Regression	75.53	4	18.88	59.33	.00(b)
	Residual	93.89	295	.32		
	Total	169.42	299			

Table 9  
Standardized and unstandardized parameter estimates

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations	
		B	Std. Error	Beta			Zero-order	Partial
1	(Constant)	1.20	.19		6.22	.00		
	Connectedness-peers	.44	.05	.46	9.44	.00	.57	.48
	Connectedness-teachers	.18	.06	.20	3.30	.00	.44	.19
	Connectedness-school	.10	.05	.12	2.05	.04	.39	.12
2	(Constant)	.83	.20		4.18	.00		
	Connectedness-peers	.40	.05	.42	8.98	.00	.57	.46
	Connectedness-teachers	.13	.05	.14	2.40	.02	.44	.14
	Connectedness-school	.03	.05	.04	.69	.49	.39	.04
	Self efficacy	.23	.04	.27	5.20	.00	.48	.29

a Dependent Variable: Academic achievement

Tables 6 to 9 are divided into two sections in which the first section corresponds to the standard multiple regression results that were reported in Tables 3 to 5. Of interest here is the effect of including self-efficacy in the second step of the regression analyses. When self-efficacy is entered, connectedness to school is no longer a statistically significant predictor of academic achievement, which suggests that the relationship between school connectedness and academic achievement is mediated by self-efficacy. In particular, an examination of the correlations in Table 9 reveals a substantial drop between the zero-order and partial correlation coefficients, thus further supporting the mediational effect of self-efficacy.

Conducting the analogous hierarchical regression within the *AMOS* graphics environment initially involves the same model specification path diagram as shown in Figure 1. The second step in the specification of the hypothesized mediated model is the inclusion of self-efficacy in the path diagram (see Figure 3).

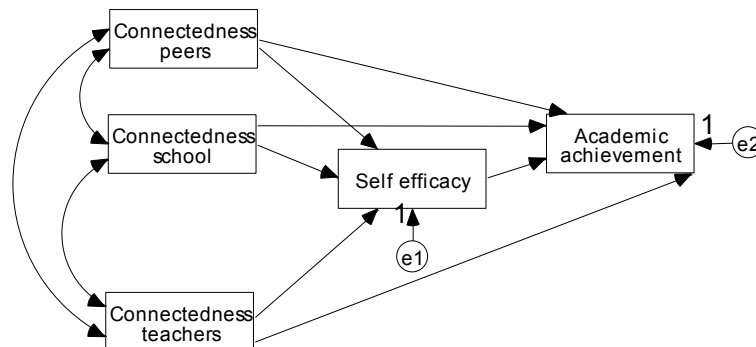


Figure 3. Hypothesized mediated model specification for *AMOS* graphics input.

Once the model has been drawn, the data attached, and the analyses conducted, the parameter estimates are then viewed within *AMOS* graphics. Figure 4 displays the standardized parameter estimates equivalent to the first and second steps of the hierarchical regression conducted in *SPSS*. The parameter estimates indicating direct effects in panel (a) and panel (b) of Figure 4 respectively correspond to the blocks of *SPSS*-derived standardized *Beta* coefficients reported under Model 1 and Model 2 in Table 9.

Of interest is the ease of interpretation of the visual representations depicted in Figure 4 when compared to Table 9. From Figure 4 it is apparent that the introduction of the mediating variable of self-efficacy is accompanied by a drop in the values of the regression coefficients representing the direct effects of the connectedness variables on academic achievement. The

greatest decrease in *Beta* weights occurs for the path from school connectedness on academic achievement (from .12 to .04). This suggests that self-efficacy mediates the relationship between these two variables. Equally, Figure 4 shows that the addition of self-efficacy to the model has smaller effects on the direct paths from peer and teacher connectedness to academic achievement and only contributes an additional 5 percent of the explained variance in academic achievement. While significance tests of the parameter estimates, which are also readily obtained in *AMOS*, would provide more rigorous testing of the mediating hypotheses, the visual representation of results suggest that the relationships between connectedness to the school environment and academic achievement are mediated (e.g., school connectedness) or partially mediated (e.g., peer and teacher connectedness) by self-efficacy. When compared to the sea of tables and numbers derived from the *SPSS* analyses, the simplicity of the visual representations can only assist in communicating a better understanding of more difficult statistical concepts and analyses.

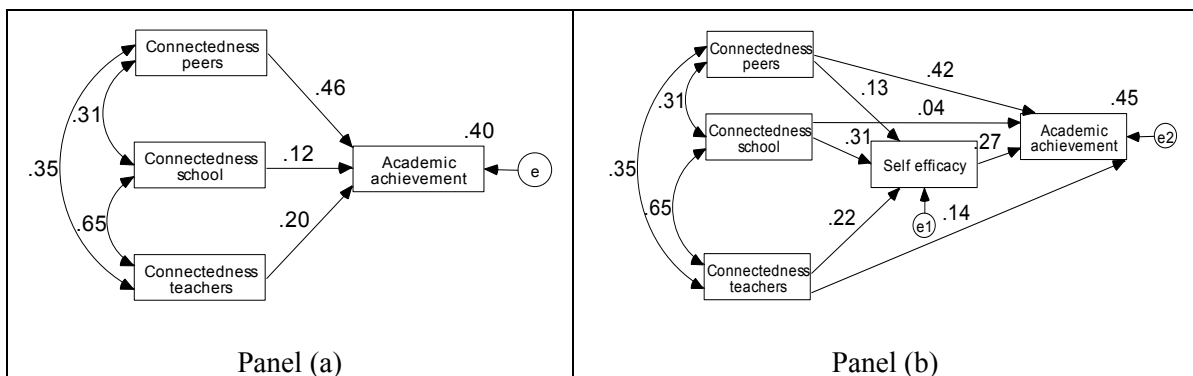


Figure 4. Standardized parameter estimates for standard and mediated models.

#### STUDENT FEEDBACK

Classroom observations and student evaluations suggest that when undergraduate students, with the aid of computer packages such as *SPSS* are asked to conduct, interpret and communicate the results from multiple regression analyses, they generally display a reasonable understanding of standard multiple regression. However, when hierarchical regression models involving potential mediating variables are introduced, many students display only very limited understanding of the concept of mediating variables and significant numbers have difficulty communicating the findings of their analyses. In contrast, direct feedback from postgraduate students whose first encounter with multiple and hierarchical regression was via the *AMOS* graphics interface do not seem to have these difficulties. Anecdotal student comments include:

‘Using *AMOS* is really helpful for understanding what’s going on here and the constructs and their interrelationships are very clearly displayed’

‘Seeing the figures (i.e., drawings) makes it is so much easier to explain and understand. I really like the way that multiple regression equations can be represented by a diagram’

‘Because of my limited statistical background, I was worried about doing the course but with the models so clearly displayed I am finding that I’m not having any real difficulty understanding things and I am sure that I could explain my findings to others’

‘Most of my previous statistical work was in qualitative data analysis. However, I really like the ease of using *AMOS* graphics and the fact that I can *see* the regression coefficients, explained variance and the like. Using *AMOS* graphics has elevated my interest in quantitative analysis.’

## CONCLUSION

This paper has shown how use of graphical interfaces might enhance the teaching of regression in undergraduate statistical courses. It seems that for many students who are visual learners, the graphical interface helps them “see” the model and the hypothesized relationships and interrelationships among the variables of interest. From a didactical perspective, having model specification as the starting point when working within the *AMOS* 5 environment has the added advantage of forcing one “to think about what assumptions about reality one has in the field of interest and how these assumptions are transformed to relationships between variables” (Kühnel, 2003, p. 82). In particular, classroom observations and anecdotal feedback from students suggest that such an approach generates increased student interest and understanding.

Further research that compares the outcomes from equivalent undergraduate classes who are taught regression either in the more traditional manner or via a graphical interface is now needed. The utilization of graphical interfaces could also readily be extended to other areas of statistical courses. For example, for many students, the conventional techniques employed in exploratory factor analyses (EFA) when using *SPSS* seem little more than waving a magic wand at a set of questionnaire items that somehow become transformed into a meaningful and valid solution. More importantly, some researchers have suggested that EFA now has no place in the social sciences (Bollen, 1995). Alternatively, confirmatory factor analysis (CFA) requires researchers to decide *a priori* the structure of the factor model (Kline, 1998). Such models are readily represented and analysed within the *AMOS* graphics environment. Perhaps it is time that the study of *EFA* in undergraduate courses be at least extended to include consideration of CFA.

In summary, the greater utilization of user-friendly graphical interfaces in teaching statistical courses is likely to engage many students who prefer a visual learning style and should assist also lecturers in communicating statistical concepts. If a picture is truly worth a thousand words, then students should benefit in understanding, absorbing, reviewing, communicating and retaining statistical concepts via exposure to the graphical representations and hands-on practice.

## REFERENCES

- Aron, A., & Aron, E. N. (1994). *Statistics for psychology*. Englewood Cliffs, NJ: Prentice Hall.
- Biktimirov, E. N., & Nilson, L. B. (2003). Mapping your course: designing a graphic syllabus for introductory finance. *Journal of Education for Business*, 78(6), 308-312.
- Bollen, K. (7 Aug 1995). *SEMNET Discussion List*. Retrieved 16 December, 2004, from <http://bama.ua.edu/cgi-bin/wa?A2=ind9508&L=semnet&P=R2541&I=1>
- Francis, G. (2003). *Multiple Regression*: Swinburne University Press.
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, 4, 599-610.
- Hoyle, R. H. (Ed.). (1995). *Structural Equation Modeling: Concepts, Issues and Applications*. Thousand Oaks, CA: SAGE Publications.
- Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- Kühnel, S. (2001). The didactical power of Structural Equation Modeling. In R. Cudeck, S. du Toit & D. Sörbom (Eds.), *Structural Equation Modeling: Present and Future, A Festschrift in honor of Karl Jöreskog*. Chicago: Scientific Software International, Inc.
- Lynch, M. (1985). *Art and artifact in laboratory science*. London: Routledge & Kegan Paul.
- Nachtigall, C., Kroechne, U., Funke, F., & Steyer, R. (2003). (Why) should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research Online*, 8(2), 1-22.
- Seeger, F. *Graphical representations and learning*. Retrieved December 16, 2004, from <http://lilt.ics.hawaii.edu/CSCL99/CollaborativeReps/Contributions/Seeger/Seeger.html>
- Steiger, J. H. (2001). Driving fast in reverse: The relationship between software development, theory, and education in structural equation modeling. *Journal of the American Statistical Association*, 96, 331-338.