STRUCTURAL EQUATION MODELING, 2(1), 79–91

SOFTWARE REVIEW

AMOS, EQS, and LISREL for Windows: A Comparative Review

AMOS 3.1. Originally distributed by James Arbuckle, Department of Psychology, Temple University, Philadelphia, PA 19122. $50.[1] Requirements for AMOS 3.1 for Windows: IBM PC–compatible computer, MS-DOS 3.1 or later, 512K free memory, 386 or 486, Microsoft Windows 3.0 or later.
EQS/Windows 4.0. BMDP Statistical Software, Inc., Suite 316, 1440 Sepulveda Boulevard, Los Angeles, CA 90025, (800) 238-BMDP, (310) 479-7799. (In Europe, EQS/Windows 4.0 is available from ProGamma, P.O.B. 841, 9700 AV Groningen, The Netherlands, +31-50-636900.) Prices vary. Upgrades and academic pricing available. Requirements: 386 or better, 2.5MG hard disk space, 4MB RAM, math coprocessor, Microsoft Windows 3.1.
LISREL for Windows 8.01. Scientific Software International, Suite 906, 1525 East 53rd Street, Chicago, IL 60615–4530. (In Europe, LISREL for Windows 8.01 is available from ProGamma, P.O.B. 841, 9700 AV Groningen, The Netherlands, +31-50-636900.) DOS/DOS Extender versions, $495; all three versions, $575; upgrades from LISREL 7, $195; all three versions, $245. Requirements: 2MB RAM (4MB recommended), 3.6MB hard disk space, Microsoft Windows 3.1.

Reviewed by Joop J. Hox
University of Amsterdam

Requests for reprints should be sent to Joop J. Hox, Department of Education, IJsbaanpad 9, University of Amsterdam, 1076 CV Amsterdam, The Netherlands. e-mail: A716HOX@HASARA11.BITNET.

[1]AMOS 3.1 was recently upgraded to AMOS 3.5 and will be reviewed in an upcoming issue of SEM. AMOS 3.5 is distributed by SmallWaters Corp., 1507 East 53rd Street, # 452, Chicago, IL 60615, (312) 667-8635, Internet: smallwaters@acm.org. $465 (retail), $325 (educational), $170 (live or competitive upgrade). Requirements: Intel-compatible PC, 4+ MB RAM, MS-Windows 3.1, NT, or IBM-OS/2 2.1.

To compare the three programs on a common data set, I generated a data set based on the solution of Holzinger and Swineford's (1939) Nine Psychological Variables data in the LISREL manual (Jöreskog & Sörbom, 1989, p. 102, second solution). This is a confirmatory factor analysis with nine variables and three factors. I generated a multinormal data set of 250 cases and a categorical data set with five-category variables having equally spaced thresholds, using the same random seed for both simulations. This exercise proved an interesting reminder that random samples are not necessarily representative; two factor loadings were consistently estimated too high by all analysis methods, with the population value actually falling outside the 95% confidence interval. Because the main object was to compare the estimates produced by the three programs, this data set was nevertheless used.[1]

With identical estimation methods, the three programs produced equal parameter estimates and standard errors; almost all parameter estimates were identical within two decimals, and any discrepancies were not larger than .01. There were some larger discrepancies in the goodness-of-fit indices; these are discussed subsequently. EQS and LISREL also produced a different solution when polychoric correlations and weighted least squares (WLS) were used for the categorical data set (AMOS does not have this feature). This is also discussed subsequently.

MODEL FIT

A common problem with the usual chi-square model test for structural equation models is that its power depends on the sample size. In large samples, models may be rejected because of trivial misspecifications, whereas in small samples, models with large misspecifications may be accepted. To address this problem, we need indices that evaluate the degree of fit of a model. Many fit indices have been proposed for this purpose. Table 1 lists the fit indices produced by AMOS, EQS, and LISREL (with maximum likelihood [ML] estimation: An explanation of the abbreviations in Table 1 is given in the Appendix). For a description and critical discussion of these and other indices, I refer to Bollen (1989) and Gerbing and Anderson (1992).

Before using Table 1 as a shopping list, one should realize that it is not clear which indices are superior. Some of these indices have been criticized because their value depends on the sample size or because they appear to be sensitive only to large misspecifications. Furthermore, all these indices are some function of the chi-square, number of parameters, and sample size. All

[1]It is possible to generate a data set in which the population model holds exactly. I decided to use simulated sample data instead, because in a recent review by Waller (1993), all programs and all estimation methods found the population values. Apparently, comparing procedures on a data set that has an exact solution is not very informative.

TABLE 1
Goodness-of-Fit Indices in AMOS, EQS, and LISREL

Index	*AMOS*	*EQS*	*LISREL*
Chi-square–based			
Model chi-square	Y	Y	Y
Independence chi-square	Y	Y	Y
Robust chi-square		Y	
Fit function value	Y	Y	Y
Noncentrality parameter			Y
General			
GFI	Y	Y	Y
PDF_0			Y
RMSR	Y	Y	Y
Standardized RMSR	Y	Y	Y
Hoelter's *N*	Y		Y
Incremental			
NNFI	Y	Y	Y
NFI	Y	Y	Y
IFI	Y	Y	Y
CFI		Y	Y
RFI	Y		Y
Based on information theory			
AIC	Y	Y	Y
CAIC	Y	Y	Y
BCC	Y		
Schwarz	Y		
ECVI			Y
Centrality Index			
McDonald's MFI		Y	
Parsimony Index			
Chi-square/df	Y		
AGFI	Y	Y	Y
RMSEA			Y
PGFI			Y
PNFI			Y

programs produce sufficient information to allow researchers to compute those indices that are not offered automatically.

Some minor discrepancies emerged in the fit indices. AMOS and LISREL produce different values for Hoelter's *N*. Hoelter's critical *N* is the largest sample size for which the model is not rejected. As it turns out, AMOS uses a .05 significance level for this test, and LISREL a .01 level. Hoelter (1983) used .05 and suggested that values of 200 and higher indicate a satisfactory fit. Using the .05 level leads to a lower critical *N* than using .01; this means that LISREL would indicate a good model more often than AMOS would.

EQS produces a chi-square of 18.00 using unweighted least squares (ULS) estimation, whereas LISREL calculates this chi-square as 6.28. With ULS, AMOS does not calculate a chi-square, but the formula χ^2 = Sample Size ×

Likelihood Function Value yields a chi-square of 6.23. Strange enough, the same formula on the Likelihood Function value reported by EQS produces a chi-square of 6.30. Two other program bugs show up in some LISREL fit indices with generalized least squares (GLS) and WLS estimation. In both cases, LISREL does not calculate the chi-square for the independence model correctly, and all fit indices that involve this chi-square—such as the non-normed fit index (NNFI) and normed fit index (NFI)—are also incorrect. This can be shown easily by specifying the independence model; obviously in this case the chi-square for the model and for the independence model should be equal, and the NFI should be zero. This is the case with AMOS and EQS, but not with LISREL (GLS/WLS method).

A minor annoyance is that the program manuals do not always use the same abbreviations for various fit indices. AMOS uses Bollen's (1989) nomenclature: delta-1 for NFI, delta-2 for incremental fit index, rho-1 for relative fit index, and rho-2/Tucker–Lewis Index for NNFI. EQS refers to McDonald's Centrality Index as MFI. (All these indices are discussed in Gerbing and Anderson [1992]). All three program manuals include the formulas used to calculate the fit indices.

CATEGORICAL VARIABLES

AMOS has no provision for categorical (ordinal) variables. LISREL handles categorical data through PRELIS. PRELIS uses the marginal univariate distribution of the observed categorical variables to estimate thresholds for the underlying latent normal variable. These thresholds are used in the estimation of the polychoric correlations and the associated asymptotic weight matrix (Jöreskog & Sörbom, 1988). The previous version of PRELIS, PRELIS 1, estimated the asymptotic covariance matrix under the assumption that two different polychoric correlations are asymptotically uncorrelated. PRELIS 2 estimates the asymptotic covariance matrix without this assumption. In fact, the estimates produced by PRELIS 2 are now identical to the estimates produced by Muthén's LISCOMP (Muthén, 1987). The polychoric correlation matrix and the associated weight matrix are input into LISREL for further analysis by WLS.

EQS uses different methods. The thresholds and polychoric correlations are estimated simultaneously for each pair of categorical observed variables, and the corresponding asymptotic covariance matrix is estimated from the information matrix. The correlation matrix is analyzed by GLS, using a modified model that ensures that the entries on the diagonal of the estimated covariance matrix are equal to 1 (Bentler & Wu, 1993, p. 165). Lee, Poon, and Bentler (1992) argue for this approach because it uses a consistent model to estimate the thresholds, polychoric correlations, and weights.

If PRELIS/LISREL and EQS are used to estimate the parameters on the categorical data set for the Holzinger–Swineford example, there are small differences between the estimates. The PRELIS thresholds and the average EQS thresholds typically differ by about ±.03. The estimates for the polychoric correlations typically differ by about the same amount. For PRELIS/LISREL, the chi-square is 26.5 (df = 23, p = .28), and for EQS the chi-square is 19.5 (p = .67). The parameter estimates again differ by about .03. Because we know the population values, we can compare the bias (mean deviation) of both methods. In our example, the PRELIS/LISREL estimates had a mean bias of .05, and the EQS estimates had a bias of .03.

Although the EQS estimates appear to be better, the differences are not dramatic. A difference that is dramatic is the difference in processing time. With EQS, processing time is clearly some exponential function of the number of categorical variables. Table 2 shows the processing times for the Holzinger and Swineford categorical data for one, two, ... , nine variables declared as categorical.

As Table 2 indicates, the EQS estimation method becomes impractical when the number of categorical variables is greater than 9. (Larger numbers of variables also need more memory than their default share, but I was able to run all models within 8MB RAM. Most processing time is spent estimating the polychoric correlations (both PRELIS/LISREL and LISCOMP used less than 1 min with nine categorical variables).

To compare EQS and PRELIS/LISREL more systematically, I performed a small simulation study, using the simulation facilities in both programs. The population model was a model with one latent factor and four observed variables. All factor loadings were set equal to 0.707, and the error variances were set equal to 0.5. With the variance of the latent factor standardized at 1, this produced a population covariance matrix with 1s on the diagonal and 0.5s outside the diagonal. I used EQS to generate 200 samples of 200 cases from this population matrix with the observed variables categorized at the thresholds −1, 0, +1, which creates observed variables with four categories. EQS was further used to compute

TABLE 2
Processing Time for Different Numbers of Categorical Variables in EQS

Categorical Variables	*Time*
1	0.6 min
2	0.8 min
3	1.4 min
4	4.4 min
5	13.5 min
6	40.8 min
7	2.0 hr
8	6.0 hr
9	17.5 hr

Note. Computations on a 486 DX-50 PC.

ML and EQS Polychoric/WLS parameter estimates. The categorical data produced by EQS were input into PRELIS, which computed the 200 polychoric correlation matrices with their associated asymptotic covariance matrices. These were passed to LISREL to compute the 200 LISREL Polychoric/GLS solutions. The whole process required three program runs: one each in EQS, PRELIS, and LISREL. It took about 1 hr to set up the command files and 45 min to run them.

Table 3 compares the mean deviation and root mean square error of the parameter estimates, the number of chi-squares that were significant at the 0.05 level, and the *p* value of a test of the simulated chi-squares against the theoretical chi-square distribution (Kolmogorov–Smirnov test).

It is clear from Table 3 that the parameter estimates are all very close to their population values. The chi-square appears to be biased upward with ML estimation and downward with EQS/Polychoric estimation. This contradicts the result reported by Lee et al. (1992) that the chi-square estimated by their approach is unbiased. However, the statistical methods used in EQS/Windows are not exactly identical to the ones used by Lee et al. (cf. Bentler & Wu, 1993, p. 165). If the parameter estimates for the same simulated sample are compared, they typically differ by about 0.03. This confirms the conclusion from the analyses of the categorical Holzinger–Swineford data: The differences are real, in the sense that they are larger than the differences between the programs with other estimation methods, but from a substantive point of view, they are negligible. Of course, the model in this simulation is very simple, and only one sample size is used. How much better the EQS estimates are across different models and different sample sizes remains an interesting question, but extensive simulation studies (and practical applications) on a large data set will need a much faster algorithm.

TABLE 3
Simulation Results Categorical Variables in EQS and LISREL

BIAS (RMSD)	*ML*	*EQS/Polychoric*	*LISREL/Polychoric*
Loading 1	.00 (.06)	.00 (.06)	.00 (.06)
Loading 2	.01 (.06)	.01 (.06)	.01 (.06)
Loading 3	.00 (.05)	.00 (.05)	.01 (.05)
Loading 4	.00 (.06)	.00 (.06)	.01 (.06)
Error 1	.00 (.08)	.00 (.08)	−.01 (.08)
Error 2	−.01 (.08)	−.01 (.08)	−.02 (.08)
Error 3	−.01 (.08)	.00 (.08)	−.01 (.08)
Error 4	−.01 (.08)	.00 (.08)	−.01 (.08)
Proportion significant χ^2 (at .05)	.12	.02	.05
p value test against theoretical χ^2	.00	.00	.11

Note. BIAS = mean deviation, RMSD = root mean square error, ML = maximum likelihood.

TIMINGS

Windows extracts a price in memory and processing overhead. Subjectively, with the exception of the polychoric correlations in EQS, all programs were very fast. To gain more precise information on the speed of the three programs in Windows and DOS, I compared the processing time needed to produce ML estimates for a large (25 × 25) Multi Trait–Multi Method (MTMM) matrix for five traits and five methods, using the general MTMM model with five Trait and five Method factors, with covariances between Trait and Method factors constrained to zero.[2] The computing times are in Table 4.

All programs are fast enough to permit analysts to run a series of MTMM models (cf. Widaman, 1985) in one session. The time differences can become significant in simulation or bootstrapping. The differences between the programs reflect not only differences in the efficiency of the algorithms but also differences in the ability of the various compilers to produce fast code. The large difference between the DOS and Windows version of AMOS especially suggests that the Windows compiler used is just not very fast. It follows that other versions of these programs could be much faster just by using a different compiler. The conclusion is that if speed is important, as in large Monte Carlo studies, it is worthwhile to perform some preliminary test runs to determine which of the available programs is the fastest.

WHY WINDOWS?

Three reasons are often quoted why Windows should be used instead of DOS:

1. Windows allows access to all of the computer's memory.
2. Windows is multitasking and thus allows working with more than one program simultaneously.

[2]In earlier analyses with LISREL 4, I found that MTMM models for this matrix need good starting values and still use many iterations to converge. The MTMM data are available from Joop J. Hox, Department of Education, IJsbaanpad 9, University of Amsterdam, 1076 CV Amsterdam, The Netherlands. e-mail: A716HOX@HASRA11.BITNET.

TABLE 4
Processing Times in AMOS, EQS, and LISREL for MTMM Model

Program Running Under	*AMOS*	*EQS*	*LISREL*
Windows	90	67	58
DOS	33	67	53

Note. Time in seconds, computations on a 486 DX-50 PC.

3. Windows uses an intuitive graphical user interface (GUI) that is superior to using the DOS command file approach.

The first reason, that Windows allows access to all the computer's memory, has become less compelling since the appearance of reliable DOS-extenders, which allow access to all available memory under DOS. Because Windows involves its own overhead, using Windows only as a DOS-extender is not very efficient.

The second reason, that Windows allows multitasking, seems to be important to some. I almost never use this feature.

The third reason, the availability of a GUI, is more pertinent. Under Windows, AMOSDraw's approach of using a path diagram to specify a model is perfectly natural. The command generator built into EQS is also very easy to use. In fact, while working in AMOS, I have missed a command generator, not only for specifying analysis options (such as estimation

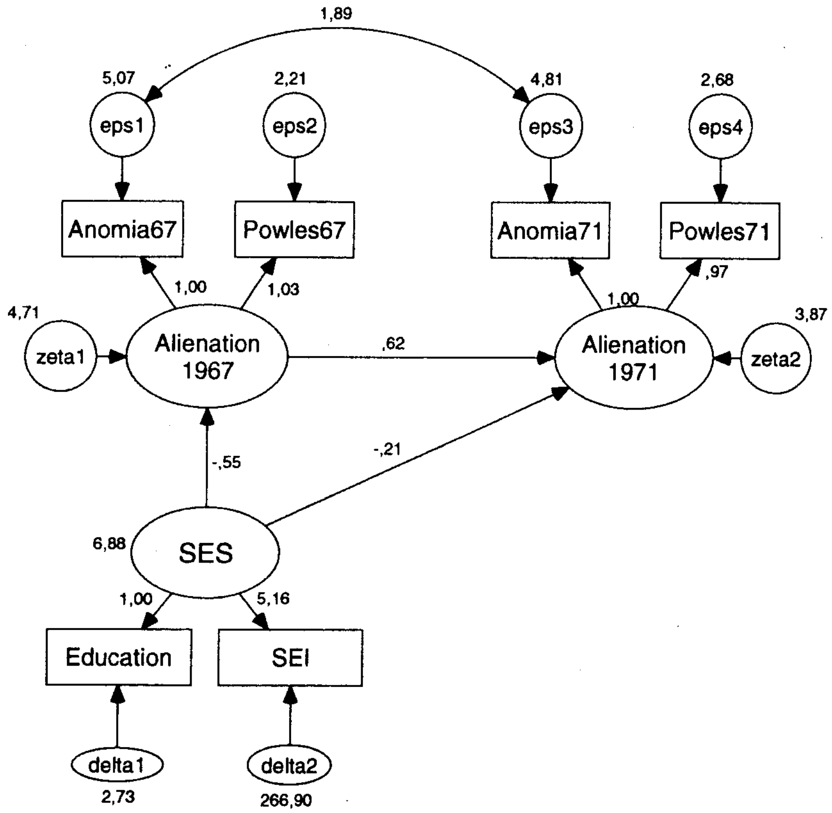

FIGURE 1 Amos path diagram for stability of alienation. $\chi^2 = 6.33$, $df = 5$.

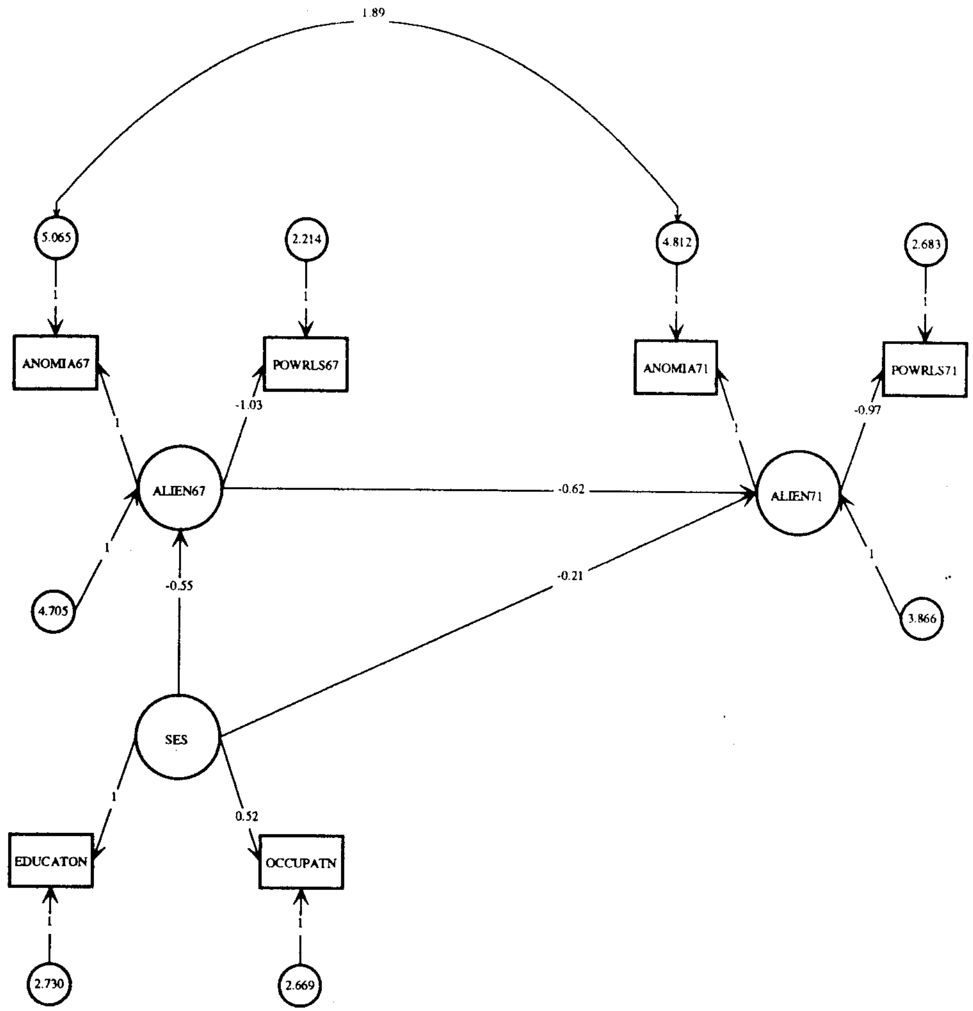

FIGURE 2 EQS path diagram for stability of alienation. $\chi^2 = 6.311$, $df = 5$, $p = 0.277$, BB NFI = .997, BB NNFI = .998, CFI = .999.

method) but also as a replacement for drawing a path diagram for highly structured complex models, such as the MTMM model. With such models, clicking on cells in a model matrix is much faster than drawing all the appropriate arrows. On the other hand, while working in EQS, I would have liked to have a drawing tool like AMOSDraw. With LISREL, I would have liked to have either, because LISREL gives no special support of any kind in setting up the model and analysis options beyond offering the Windows Notebook as a command file editor.

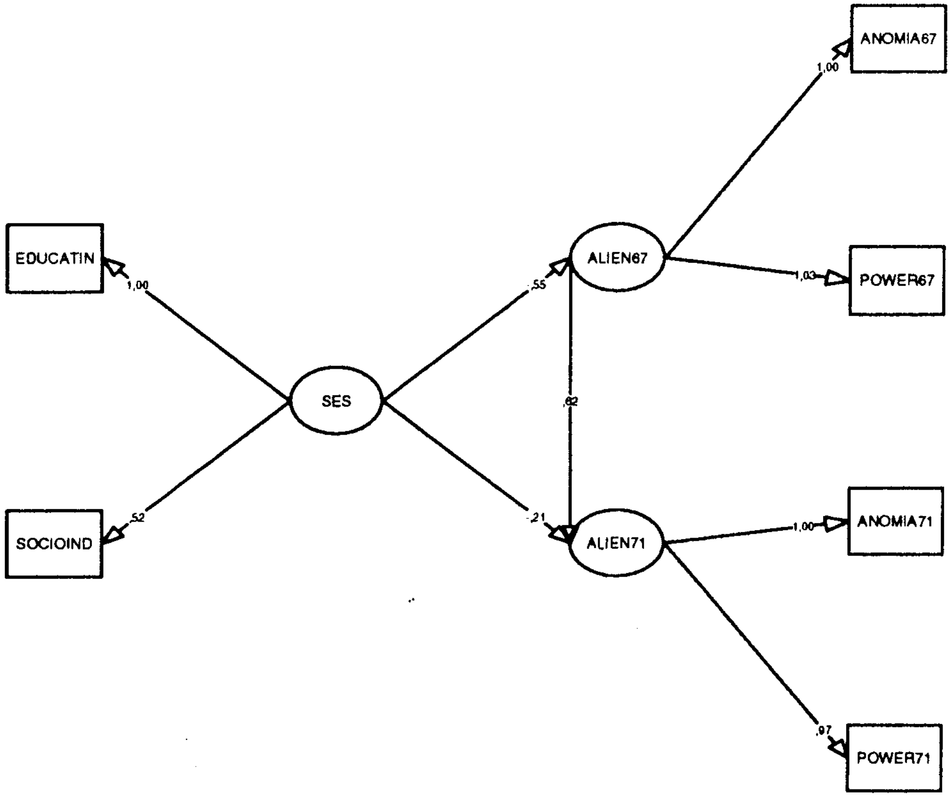

FIGURE 3 LISREL path diagram for stability of alienation.

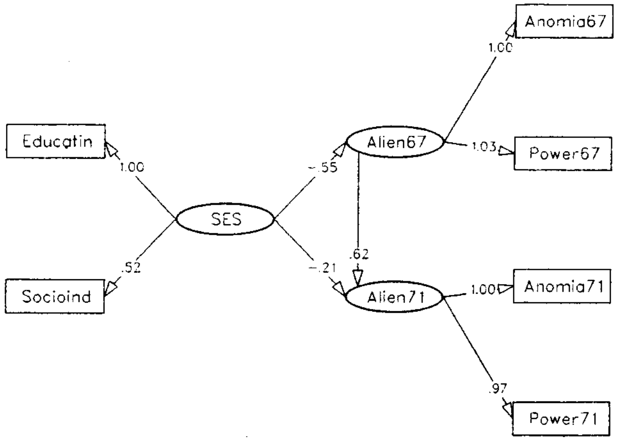

FIGURE 4 LISREL (DOS/Halo version) path diagram for stability of alienation.

Given the graphical capacities of Windows, it is a natural request that a Windows SEM program be able to produce publication-quality path diagrams as output. In this regard, AMOS is a clear winner. AMOS lets the analyst draw the path diagram precisely as wanted and produces the same path diagram as output with parameter estimates and other information added. There is also the possibility to change the layout of the output to some degree (AMOSDraw forbids eliminating paths from the output diagram). EQS lets the analyst specify the global layout of the path diagram in the command file and then produces a postscript file. LISREL allows no specification of the layout. A peculiarity with LISREL is that it draws several path diagrams for each analysis. The most complete diagram is the basic diagram, which shows all structural and measurement relations excluding the error terms. There are other diagrams that show only the structural model, the measurement model for *X*, or for *Y*, modification indices, and so on. If one needs a figure that shows all relations including the error terms, LISREL cannot produce that. To give an impression of the aesthetic value of the path diagrams, Figures 1, 2, and 3 show the path diagrams produced by AMOS, EQS, and LISREL for Windows (output on Postscript Laserprinter). For comparison, Figure 4 shows the path diagram produced by the HALO routines in the DOS version of LISREL (output on DeskJet printer).

Finally, in AMOS and EQS, I miss the capacity to write the path diagram to a graphics file. It is becoming increasingly common that conference proceedings and other book publications request that tables and figures be inserted in the text in a specified format. LISREL can output the path diagram as a graphics file, which can be read by a drawing program or a word processor. The EQS postscript file can be read by some programs (e.g., WordPerfect) but not edited. AMOS can communicate with other programs only through the Windows clipboard, which I could not get to work with some other programs (e.g., Windows Write works fine; Windows Paint and WordPerfect do not).

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107* 238–246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M., & Wu, E. J. C. (1993). *EQS Windows user's guide.* Los Angeles: BMDP Statistical Software.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC). *Psychometrika, 52,* 345–370.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavior Research, 24,* 445–455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258.

Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research, 21,* 132–160.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research, 11,* 325–344.

Holzinger, K., & Swineford, F. (1939). A study in factor analysis: The stability of a bifactor solution. *Supplementary Educational Monograph 48* (p. 91). Chicago: University of Chicago Press.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS. A program for multivariate data screening and data summarization.* Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7—A guide to the program and applications.* Chicago: Scientific Software.

Lee, S.-Y., Poon W.-Y., & Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika, 57,* 89–362.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97–103.

Mulaik, S., James, L., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. (1989). An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin, 105,* 430–445.

Muthén, B. O. (1987). *LISCOMP. Analysis of linear structural equations with a comprehensive measurement model.* Mooresville, IN: Scientific Software.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Waller, N. G. (1993). Software review. Seven confirmatory factor analysis programs: EQS, EzPath, LINCS, LISCOMP, LISREL 7, SIMPLIS, and CALIS. *Applied Psychological Measurement, 17,* 73–100.

Widaman, K. F. (1985). Hierarchically tested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement, 9,* 1–26.

APPENDIX: GOODNESS-OF-FIT INDICES: ABBREVIATIONS AND EXPLANATION

Abbreviation	*Explanation*
General	
GFI	Goodness-of-Fit Index (Jöreskog & Sörbom, 1989)
PDF_0	Population Discrepancy Function (Browne & Cudeck, 1992)
RMSR	Root Mean Squared Residual (Bollen, 1989)
Standardized RMSR	Standardized RMSR (standardized by estimated standard error
Hoelter's *N*	Critical *N* (Hoelter, 1983)
Incremental	
NNFI	Non-Normed Fit Index (Bentler & Bonett, 1980). Also: Tucker–Lewis index (Tucker & Lewis, 1973); rho_2 (Bollen, 1989)
NFI	Normed Fit Index (Bentler & Bonett, 1980). Also: $delta_1$ (Bollen, 1989)
IFI	Incremental Fit Index (Bollen, 1989). Also: $delta_2$ (Bollen, 1989)
CFI	Comparative Fit Index (Bentler, 1990)
RFI	Relative Fit Index. Also: rho_1 (Bollen, 1989)
Based on Information Theory	
AIC	Akaike's Information Criterium (Akaike, 1987)
CAIC	Consistent AIC (Bozdogan, 1987)
BCC	Browne & Cudeck Criterium (Browne & Cudeck, 1989)
Schwarz	Schwarz Criterium (Schwarz, 1978)
ECVI	Expected Cross Validation Index (Browne & Cudeck, 1989)
Noncentrality Index	
McDonald's NCI	Non-Centrality Index (McDonald, 1989). Also: MFI.
Parsimony Index	
AGFI	Adjusted Goodness-of-Fit Index (Jöreskog & Sörbom, 1989)
RMSEA	Root Mean Squared Error of Approximation (Browne & Cudeck, 1992)
PGFI	Parsimony Adjusted GFI (Mulaik et al., 1989)
PNFI	Parsimony Adjusted NFI (Mulaik et al., 1989)