DOCUMENT RESUME

ED 454 245 TM 032 848

AUTHOR	Witta, E. Lea
TITLE	Does Method of Handling Missing Data Affect Results of a Structural Equation Model?
PUB DATE	2001-02-00
NOTE	30p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; *High School Students; High Schools; Regression (Statistics); *Research Methodology; *Statistical Analysis; *Structural Equation Models
IDENTIFIERS	*Missing Data; National Education Longitudinal Study 1988

ABSTRACT

The influence of method of handling missing data on estimates produced by a structural equation model of the effects of part-time work on high-school student achievement was investigated. Missing data methods studied were listwise deletion, pairwise deletion, the expectation maximization (EM) algorithm, regression, and response pattern. The 26 variables selected from the National Educational Longitudinal Survey of 1988 database were those previously used by K. Singh and M. Ozturk (1999) in an analysis of part-time work. Results indicate the data was not missing completely at random, and although the covariance matrices, measurement models, and structural models using the five missing data methods were not significantly different statistically, the individual best fitting structural model for each missing data method differed substantively. Results are discussed. (Contains 4 figures, 3 tables, and 20 references.) (Author/SLD)

Running Head: Missing Data - Structural Models

Does method of handling missing data affect results of a structural equation model?

E. Lea Witta

University of Central Florida

Department of Educational Foundations

Abstract

The influence of method of handling missing data on estimates produced by a structural equation model of the effects of part-time work on high-school student achievement was investigated. Missing data methods investigated were listwise deletion, pairwise deletion, the EM algorithm, regression, and response pattern. The 26 variables selected from National Educational Longitudinal Survey of 1988 database were those previously used by Singh and Ozturk (1999) in an analysis of part-time work. Results indicate the data was not missing completely at random, and although the covariance matrices, measurement models, and structural models using the five missing data methods were not significantly different statistically, the individual best fitting structural model for each missing data method differed substantively. Results are discussed.

Does method of handling missing data affect results of a structural equation model?

Different methods of handling missing values may produce different results. When Jackson (1968) entered data on all the available variables in a discriminant analysis, the significance of the regression coefficients of individual variables, as well as the interpretation of the importance of these variables, changed with the missing value method used. Witta and Kaiser (1991) also reported that the regression coefficients and total variance accounted for by the variables changed depending on the method used to handle missing values. After re-analyzing three studies of private/public school achievement, Ward and Clark III (1991) concluded that the method used to handle missing data influenced the outcome of these studies. Thus, it would seem that the method chosen to handle missing values affects the substantive results of that study. If, however, the initial model covariance matrices are equivalent, is there a difference in substantive interpretation of the final models based on missing data handling method used?

There are many methods used to investigate effectiveness of missing data methods. Some researchers compare covariance matrices or variable means for equality. Some researchers compare other non-missing variables for the incomplete cases to those of the complete cases. In using the National Educational Longitudinal Study of 1988 database to investigate the effects of part-time work on school outcomes Singh and Ozturk (1999, p. 10) stated "The initial sample for this study was N=4600 but the final analyses (structural equations models) are based on 1582 cases after listwise deletion of all incomplete data." They further add that the incomplete cases were similar to the complete cases. In addition to questions concerning representativeness of the population, the removal of 66% of the cases leads to the question, what changes in interpretation of the structural model if another missing data handling method were used?

The purpose of the current study was to determine what changes in interpretation of the structural model would occur if different methods of handling missing data were used. The incomplete cases for the 26 variables in the Singh and Ozturk (1999) study were treated using the listwise deletion, pairwise deletion, regression imputation, expectation maximization algorithm, and the response pattern missing data methods. The equality of the covariance matrices, measurement and structural models for data produced by each missing data method were compared. Then each model (data produced by use of missing data method) was analyzed individually to determine if there were substantive changes in interpretation of the model.

Until recently, the only methods available with popular statistical computer software focused on handling the missing data problem by deleting subjects with incomplete information, deleting the variables with missing values, or replacing the missing value with some reasonable estimate. Now, however, new subroutines are available to provide more assistance in handling missing data and providing analysis choices using iterative regression or expectation maximization (EM) procedures. These relatively new methods (in current software) also provide the possibility of specifying the model to be used (i.e., multivariate normality, adding a randomly selected error). In addition, the PRELIS 2 preprocessor for the LISREL 8 computer program provides a response pattern method of handling missing data.

Methods Studied

Listwise Deletion

Listwise deletion is probably the most frequently used method of handling missing data and is available as a default option in several statistical software programs including. This method discards cases with a missing value on any variable and thus is very wasteful of data. Listwise deletion, however, has been shown to be effective with low average intercorrelation, less than four variables and a small proportion of missing values (Chan, et.al., 1976; Haitovsky, 1968; Timm, 1970). The assumption of missing completely at random is crucial to the use of this method. It is more likely, however, to find the complete sample different in important ways from the incomplete sample (Little & Rubin, 1987). Problems for a researcher using this method include a reduction in power and an increase in standard error due to reduced sample size and the possible elimination of sub-populations.

Pairwise Deletion

When using pairwise deletion, covariances are computed between all pairs of variables having both observations, eliminating those that have a missing value for one of the two variables (Glasser, 1964). Means and variances are computed on all available observations. The assumption made is that the use of the maximum number of pairs and all the individual observations yield more valid estimates of the relationship between the variables. It is assumed that when two variables are correlated, information on one improves the estimates of the other variable. It is also assumed that the pairs are a random subset of the sample pairs. If these assumptions are true, pairwise deletion produces unbiased estimates of the variable means and variances (Hertel, 1976). When missing data are not missing completely at random, however, the correlation matrix produced by pairwise deletion may not be Gramian (Norusis, 1988).

Marsh (1998) investigated the estimates produced when using pairwise deletion for randomly missing data. From this study, which included five levels of missing data and three sample sizes, Marsh concluded parameter variability was explained, parameter estimates were unbiased, and only one covariance matrix was nonpositive definite.

Regression

Regression as an imputation method has many variations. The regression methods rely on information contained in non-missing values of other variables to provide estimates of missing values. As the average intercorrelation and the number of variables from which these methods can obtain information increases, the regression methods, theoretically, perform better. Too many variables, however, can cause problems with over prediction (Kaiser & Tracy, 1988) and too high an average intercorrelation can result in a singular matrix. In these cases, regression does not perform well.

Variations in the regression methods include differences in methods of developing the initial correlation matrix (listwise deletion, pairwise deletion, and mean substitution) and the presence or absence of iteration procedures. Differences in regression methods also include the use of randomly selected residuals for iterations and assumptions of a normal distribution. Theoretically, the more variables considered that provide additional information, the better the estimate. Mundfrom and Whitcomb (1998) investigated the effects of using mean substitution, hot-deck imputation, and regression imputation on classification of cardiac patients. Mean substitution and hot-deck imputation correctly classified patients more frequently than regression imputation.

Expectation Maximization

Dempster, Laird, and Rubin (1977) recommended the use of the EM algorithm which imputes estimates simultaneously in an iterative procedure. The E step of this algorithm finds the conditional expectation of the missing values. The M step performs maximum likelihood estimation as if there were no missing data. The primary difference between this procedure and the regression procedure is that the values for the missing data are not imputed and then iterated. The missing values are functions based on the conditional expectation (Little & Rubin, 1987). This method of handling missing data represents a fundamental shift in the way of thinking about missing data (Schafer & Olsen, 1998).

Response Pattern

The response pattern method of handling missing values is available in the PRELIS 2 preprocessor for LISREL 8. Using this method, the "value to be substituted for the missing value for a case is obtained from another case that has a similar response pattern over a set of matching variables" (Jöreskog & Sörbom, 1996b, p.78). This method provides intuitive appeal in that it provides imputation only if there is a similar response pattern.

Pattern of Missing Values

All of the missing data handling procedures discussed except response pattern require data missing at random (MAR) or missing completely at random (MCAR). Yet Cohen and Cohen (1983) suggested that in survey research the absence of data on one variable may be related to another variable (MCAR) and may be due to the value of the variable itself (MAR). When investigating simultaneously missing values, Witta (1996/97) found concurrently missing values (p<.001) in three of four samples using data from a national database.

Schafer and Olsen (1998), however, argue convincingly that "every missing-data method must make some largely untestable statistical assumptions about the manner in which the missing values were lost" (p551). Consequently, they (Schafer & Olsen, 1998) suggest when analyzing real data, researchers typically assume missing at random.

Procedure

All high school seniors who had reported working during their tenth grade and senior year of high school and for whom base-year and first follow-up data were available were included in this study. The initial sample contained the 26 variables used in the Singh and Ozturk (1999) study for 4337 subjects. The four grades variables were eliminated and twelve composite variables were created in a method similar to Singh and Ozturk (see Table A-1). This resulted in a sample containing approximately 28% incomplete cases (3128 complete cases and 1209 incomplete).

Because the initial sample contained 28% incomplete cases and Singh and Ozturk (1999) had indicated that in their final model 60% of the cases were removed by listwise deletion, an additional model was also analyzed. All incomplete cases (1209) were retained. Eight hundred fifty nine cases were randomly selected from the 3128 complete cases. Merging these files resulted in a second sample (n=2068) for analysis with 58% of the cases containing one or more missing values.

Analysis

The composite indicators were treated by each missing data handling method in the missing data subroutine in SPSS 10.0. Correlation matrices, means, and standard deviations for the missing data handling methods were produced by this subroutine. The test for missing completely at random and pattern of missing data was also produced by this subroutine. In addition, the response pattern method available in PRELIS 2 was used to treated the missing data. The correlation matrix for this method was produced by PRELIS 2. Because the response pattern method converted variables with less than 14 distinct values to z scores, the means and standard deviations for the three variables affected were converted to the means and standard deviations of all possible values.

After treatment by each missing data handling method, multi-sample analysis in LISREL 8.3 (Jöreskog & Sörbom, 1996a, chap. 9) was used to test the equality of the covariance matrices, and the measurement and structural models produced by each missing data handling method. Then, the data produced by each missing data handling method were analyzed independently. Paths that were statistically nonsignificant ($p>.05$) were deleted from each model. The resulting models were compared logically across missing data methods. Although the actual sample size varied across missing data methods, in order to provide estimates that were not distorted by sample size, all correlation matrices, means, and standard deviations were entered into LISREL using a sample size of 1500.

Results

Randomness of Missing Values

When data were tested for randomness of missing values, results suggested the missing data may not be missing at random and was not missing completely at random as measured by Little's chi square ($\chi^2 = 646$, df=350, p<.01). The frequency of missing data (simultaneously and independently) is depicted in Figure 1. The category of 'Tests' consists of four simultaneously missing standardized test variables (History, Math, Reading, and Science). The standardized test variables were also missing in conjunction with missing values for homework 10, homework 12, and a motivation variable. If a variable did not contain a missing value for 10% of the sample cases (either alone or concurrently with other variables), it was included in the 'Other' category. The majority of the cases containing missing values consisted of concurrently missing values for standardized tests, the dependent variable in this analysis.

Insert Figure 1 About Here

28% Incomplete Cases

The initial test of equality of covariance matrices produced by use of each missing data method when 28% of the cases were incomplete was not statistically significant (χ^2=78.37, df=312, p>.05). The initial model, which is similar to the model used by Singh and Ozturk (1999), is depicted in Figure 2. Although the initial fit for the model used for each missing data method did not fit when measured by chi-square (χ^2, $\underline{p} \leq .05$), the standardized residuals for each missing data method model were approximately 0.02 and the goodness of fit index was at 0.98. When analyzed simultaneously for the same pattern, the omnibus χ^2 for all models was 879.03 (df=220, $\underline{p} < .01$) with a root mean square error of approximation (RMSEA) of 0.04. When the measurement portion of the model for each missing data method was constrained to equivalence, chi-square increased by a non-significant 3.58 with 28 degrees of freedom. The analysis was further constrained by forcing the structural portion of each model to equivalence for each missing data method. Chi-square increased to 893.34 (df=288), a chi-square increase of 10.73 (df=40) - again a nonsignificant increase. The individual results are depicted in Table 1.

Insert Table 1 and Figure 2 About Here

Each model was then analyzed separately to determine the best fitting model if non-significant (statistically) paths were removed. Criteria used was, the final model could not have a statistically significant chi-square increase for the change in degrees of freedom. This resulted in removal of one path in the listwise deletion and response pattern models, two paths in the regression model, and four paths in the EM algorithm and pairwise deletion models.

The path from part-time work to homework was removed from all models except the one produced by the response pattern method. In addition, the path from attendance to motivation was removed from all models except listwise deletion. The paths from part-time work to motivation and from attendance to tests were removed in the pairwise deletion and EM algorithm models (see Figure 3). These changes also affected the influence of variables on the dependent standardized test scores. The total effect of part-time work on tests ranged from -5.68 (standardized = -.36) in the listwise deletion model to -8.05 (standardized= -.41) in the EM model. The variance in standardized test accounted for by other variables in the model ranged from 25% (listwise deletion) to 30% (pairwise deletion and the EM algorithym). These results are displayed in Table 2.

Insert Table 2 and Figure 3 About Here

58% Incomplete Cases

The test of equality of covariance matrices produced by use of each missing data method when 58% of the cases were incomplete was not statistically significant (χ^2=353.01, df=312, p>.05). Again, the initial fit for the model used for each missing data method (see Figure 2) did not fit when measured by chi-square (χ^2, $p \leq .05$), but the standardized residuals for each missing data method model did not exceed 0.04 for any model and the goodness of fit index was never below 0.97. When analyzed simultaneously for the same pattern, the initial χ^2 for all models was 1095.42 (df=220, p <.01) with a root mean square error of approximation (RMSEA) of 0.05. Chi-square increased by a non-significant (statistically) 8.04 with 28 degrees of freedom when the measurement portion of the model for each missing data method was constrained to equivalence. When the analysis was further constrained by forcing the structural portion each model to equivalence for each missing data method, chi-square increased to 1144.60 (df=288), a χ^2 increase of 41.14 (df=40) - again a nonsignificant increase. The individual results are depicted in Table 1.

When 58% of the cases in a sample were incomplete, paths from the motivation variable to tests and from attendance to motivation were not statistically significant and were removed from all models (see Figure 4). When listwise deletion was used, however, all paths leading to tests (both direct and indirect) from attendance were removed. If, on the other hand, regression or the response pattern methods were used, attendance is not only a statistically significant contributor to test score, but has a larger total effect (standardized = -0.11, -0.13 respectively) on test scores than motivation (standardized 0.09. 0.10 respectively). In addition, when using the pairwise deletion model, part-time work has a total effect on test score of -9.74 (standardized -.48). When using the response pattern method, part-time work has a total effect on test score of -3.09 (standardized = -.38). These results are depicted in Table 3 and Figure 4.

Insert Table 3 and Figure 4 About Here

Discussion and Conclusions

Although the proportion of incomplete cases was small (28%) in the initial sample and there were no statistically significant differences in the covariance matrices, measurement models, or structural models based upon missing data method used, there were differences in interpreting an individual model. The listwise deletion model is the only model indicating a direct effect of attendance on motivation. In addition, the path coefficients varied from one model to another. For example, the path coefficient between part-time work and attendance is 0.23 in the listwise model, 0.31 in the pairwise and EM models, 0.25 in the regression model, and 0.24 in the response pattern model. The path between attendance and tests is -0.08 in the listwise and regression models, -0.07 in the response pattern model, and does not exist in the pairwise and EM models. Thus, interpretation of the meaning of each model changes based upon which missing data method was chosen to handle the incomplete cases. And, as the proportion of incomplete cases increases, this situation becomes more pronounced.

When 58% of the cases were incomplete, the covariance matrices produced by the missing data handling methods for the initial model did not differ significantly. In addition, the measurement and structural models did not differ. When examining the structure of the final model for each missing data method, however, the latent variable of attendance had no effect on tests in the listwise deletion model. Under these circumstances in the regression and response pattern models, attendance not only had a direct effect on tests, but also an indirect effect through homework. In the pairwise and EM models there was only an indirect effect of attendance through homework. On the other hand, in the pairwise and EM models, the motivation variable became an exogenous variable. Again, as in the models produced when 28% of the cases were incomplete, the covariance matrices, the measurement model, and the structural model did not differ, but the interpretation of the individual model produced changed.

This study was limited to one sample size and proportion of incomplete cases. Consequently, results may be specific to these samples. This study did not evaluate the effectiveness of the missing data methods used. Therefore, no conclusions concerning which is the better method can be made. The findings from this study, however, imply that the missing data method chosen for a study will influence the substantive interpretation of the final model.

In addition, the results from the current study imply that use of equality of covariance matrices to test effectiveness of missing data methods may be questionable. Consequently, researchers should provide a logical reason for the method of handling missing data chosen for their study. Because decisions made concerning models and removal of paths was based solely on statistical significance in the current study, a further caution is added concerning this use of a single criteria for decision making. Further research providing evidence of the effectiveness of methods of handling missing data and into criteria for judging effectiveness is needed .

References

Chan, L.S., Gilman, J.A., & Dunn, O.J. (1976). Alternative approaches to missing values in discriminant analysis. *Journal of the American Statistical Association, 71*, 842-844.

Dempster, A.P., Laird, N.W., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B, 39*, 1-38.

Glasser, M. (1964). Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association, 59*, 834-844.

Haitovsky, Y. (1968). Missing data in regression analysis. *Journalof the Royal Statistical Society, B, 30*, 67-82.

Hertel, B.R. (1976). Minimizing error variance introduced by missing data in survey analysis. *Sociological Methods & Research, 4*, 459-474.

Hill, M.A. (1997). *SPSS Missing Value Analysis 7.5* [Computer program manual]. Chicago: SPSS Inc.

Jackson, E.C. (1968). Missing values in linear multiple discriminant analysis. *Biometrics, 24*, 835-844.

Jöreskog, K.G. & Sörbom, D. (1996a). *Lisrel 8: User's reference guide*. Chicago: Scientific Software International, Inc.

Jöreskog, K.G. & Sörbom, D. (1996b). *PRELIS 2: User's reference guide*. Chicago: Scientific Software International, Inc.

Kaiser, J. & Tracy, D.B. (1988). Estimation of missing values by predicted score. *Proceedings of the Section on Survey Research, American Statistical Association 1988*. 631-635.

Little, R.J.A., & Rubin, D.R. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Marsh, H.W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling, 5* (1), p 22-36.

Mundfrom, D.J. & Whitcomb, A. (1998). *Imputing missing values: The effect on the accuracy of classification*. Paper presented at the annual meeting of the American Educational Research Association, San Diego. ED419817.

Norusis, M.J. (1988). *SPSS-X Introductory Statistics Guide: Release 3* [Computer program manual]. (pp 107-108). Chicago: SPSS Inc.

Schafer, J.L. & Olsen, M.K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavior Research, 33* (4), p 545-571.

Singh K., & Ozturk, M. (1999). *Part-time work and school-related outcomes for high school seniors: An analysis of NELS:88*. Paper presented at the 1999 Annual Conference of the American Educational Research Association, Montreal, Canada.

Timm, N.H. (1970). The estimation of variance-covariance and correlation matrices from incomplete data. *Psychometrika, 35*, 417-437.

Ward, Jr., T.J. & Clark III, H.T. (1991). A reexamination of public-versus private-school achievement: the case for missing data. *Journal of Educational Research, 84*, 153-163.

Witta, E.L. (1996/97). Randomness of missing values in survey data. *Louisiana Education Research Journal, XXII* (2), p 73-86.

Witta L. & Kaiser, J. (1991, November). *Four methods of handling missing data with GSS-84*. Paper presented at the meeting of the Mid-South Educational Research Association, Lexington, KY

Appendix A

Table A-1

Study Questions and their suggested Construct

Construct	Variable Code	Question
Part-time Work (Grade 10)	F1S85	HOW MANY HRS DOES R USUALLY WORK A WEEK
Part-time Work (Grade 12)	F2S88	CURRENT JOB, # HRS WORKED DURING SCHL YR
Attendance (Grade 10)	F1S10A	HOW MANY TIMES WAS R LATE FOR SCHOOL
	F1S10B	HOW MANY TIMES DID R CUT/SKIP CLASSES
	F1S13	HOW MANY DAYS WAS R ABSENT FROM SCHOOL
Attendance (Grade 12)	F2S9A	HOW MANY TIMES WAS R LATE FOR SCHOOL
	F2S9B	HOW MANY TIMES DID R CUT/SKIP CLASSES
	F2S9C	HOW MANY TIMES DID R MISS SCHOOL
Participation (Grade 10)	F1S40A	OFTEN GO TO CLASS WITHOUT PENCIL/PAPER
	F1S40B	OFTEN GO TO CLASS WITHOUT BOOKS
	F1S40C	OFTEN GO TO CLASS WITHOUT HOMEWORK DONE
Participation (Grade 12)	F2S24A	GO TO CLASS WITHOUT PENCIL/PAPER
	F2S24B	GO TO CLASS WITHOUT BOOKS
	F2S24C	GO TO CLASS WITHOUT HOMEWORK DONE
Homework (Grade 10)	F1S36A1	TIME SPENT ON HOMEWORK IN SCHOOL
	F1S36A2	TIME SPENT ON HOMEWORK OUT OF SCHOOL
Homework (Grade 12)	F2S25F1	TOTAL TIME SPENT ON HMWRK IN SCHOOL
	F2S25F2	TOTAL TIME SPENT ON HMWRK OUT SCHL
Standardized Tests (Grade 12)	F22XHSTD	HISTORY/CIT/GEOG STANDARDIZED SCORE
	F22XMSTD	MATHEMATICS STANDARDIZED SCORE
	F22XRSTD	READING STANDARDIZED SCORE
	F22XSSTD	SCIENCE STANDARDIZED SCORE

Table 1

Comparison of Model Fit across Missing Data Treatments and Proportion of Incomplete Cases

Condition		28% Incomplete Cases						58% Incomplete Cases					
		χ^2	standardized	GFI	RMSEA	$\Delta\chi^2$	Δdf	χ^2	standardized	GFI	RMSEA	$\Delta\chi^2$	Δdf
Pattern	Listwise	182.28	0.023	0.98				268.05	0.033	0.97			
Same	Pairwise	172.01	0.021	0.98				195.32	0.023	0.98			
	EM	171.38	0.021	0.98				189.18	0.023	0.98			
	Regression	176.95	0.022	0.98				231.00	0.033	0.98			
	Response Pat	176.4	0.022	0.98				211.87	0.026	0.98			
	Total (df=220)	879.03			0.044			1095.42			0.052		
Measure	Listwise	182.43	0.023	0.98				270.48	0.033	0.97			
ment	Pairwise	172.15	0.022	0.98				196.32	0.023	0.98			
Model	EM	171.52	0.021	0.98				191.21	0.024	0.98			
Invariant	Regression	177.06	0.022	0.98				232.32	0.028	0.98			
	Response Pat	178.73	0.022	0.98				213.14	0.026	0.98			
	Total (df=248)	882.61			0.041	3.58	28	1103.46			0.049	8.04	28
Adds	Listwise	184.52	0.024	0.98				292.60	0.041	0.97			
Structural	Pairwise	173.14	0.022	0.98				203.93	0.027	0.98			
Model	EM	172.66	0.02	0.98				196.94	0.027	0.98			
Invariant	Regression	177.59	0.023	0.98				233.64	0.028	0.97			
	Response Pat	183.35	0.024	0.98				217.49	0.028	0.98			
	Total (df=288)	893.34			0.037	10.73	40	1144.60			0.045	41.14	40

Table 2

Effects after Removal of Non-significant Paths when 28% of the Cases were Incomplete

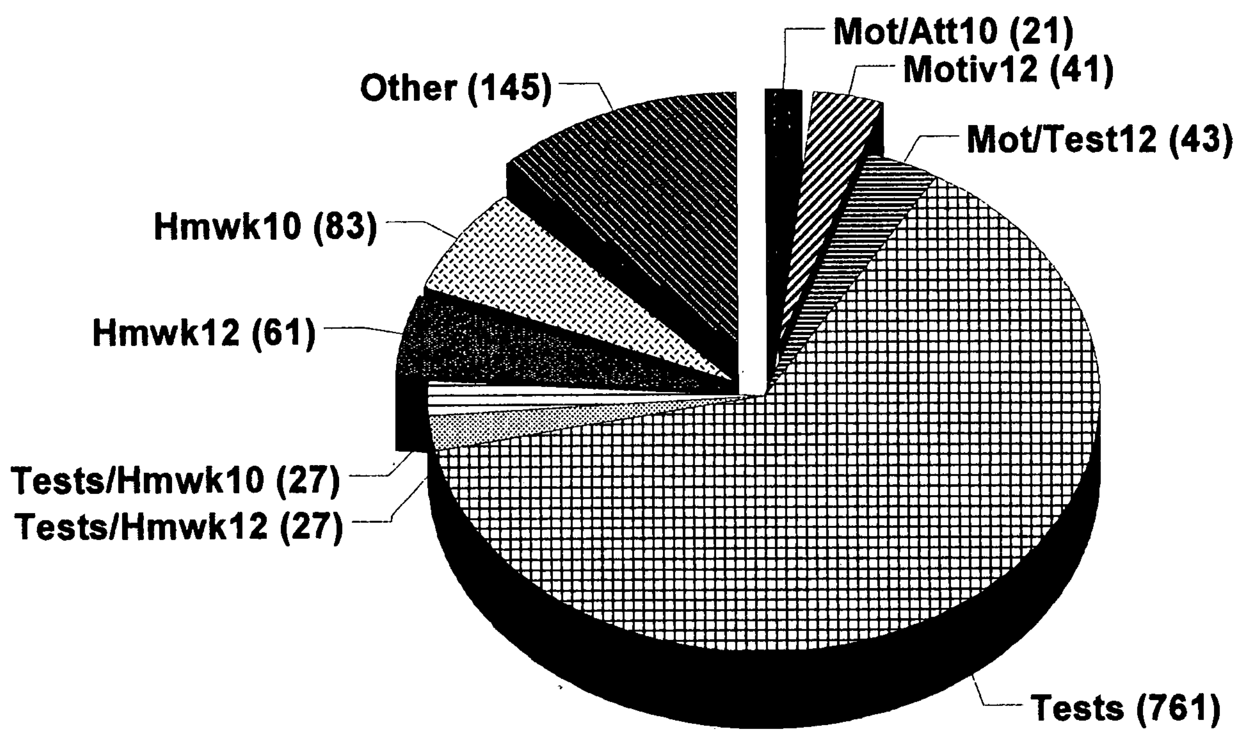
	Initial χ^2	Initial df	Paths Removed	Final χ^2	Final df	Variable	Effects Total	Effects Indirect	Standardized Effects Total	Standardized Effects Indirect
Listwise										
	179.1	44	Part-time to Homework	182.23	45	Part-time	**-6.78***	**-0.60***	**-0.36***	**-0.03***
						Attendance	**-1.41***	**-0.51***	**-0.13***	**-0.05***
						Motivation	0.07	**2.00***	0.00	**0.15***
						Homework	**1.68***	0.00	**0.37***	0.00
Pairwise										
	170.69	44	Part-time to Homework	184.32	48	Part-time	**-7.75***	**-0.49***	**-0.40***	**-0.03***
			Part-time to Motivation			Attendance	**-0.84***	**-0.84***	**-0.08***	**-0.07***
			Attendance to Motivation			Motivation	0.67	**2.22***	0.05	**0.16***
			Attendance to Tests			Homework	**2.01***	0.00	**0.40***	0.00
EM										
	169.73	44	Part-time to Homework	182.86	48	Part-time	**-8.05***	**-0.49***	**-0.41***	**-0.03***
			Part-time to Motivation			Attendance	**-0.85***	**-0.85***	**-0.08***	**-0.08***
			Attendance to Motivation			Motivation	0.54	**2.19***	0.04	**0.16***
			Attendance to Tests			Homework	**1.94***	0.00	**0.38***	0.00
Regression										
	174.12	44	Part-time to Homework	179.17	46	Part-time	**-6.86***	**-0.67***	**-0.37***	**-0.04***
			Attendance to Motivation			Attendance	**-1.43***	**-0.63***	**-0.14***	**-0.06***
						Motivation	0.15	**2.04***	0.01	**0.15***
						Homework	**1.75***	0.00	**0.37***	0.00
Response Pattern										
	173.78	44	Attendance to Motivation	175.65	45	Parttime	**-5.68***	**-1.37***	**-.36***	-0.08
						Attendance	**-1.37***	**-0.69**	**-.13**	-0.07
						Motivation	0.05	**1.94***	0.00	**0.15***
						Homework	**1.65***	0.00	**0.40***	0.00

Table 3

Effects after Removal of Non-significant Paths when 58% of the Cases were Incomplete

	Initial			Final			Effects		Standardized Effects	
	χ^2	df	Paths Removed	χ^2	df	Variable	Total	Indirect	Total	Indirect
Listwise										
	266.97*	44	Motivation to Tests	**270.67***	48	Part-time	**-7.13***	**-1.36***	**-0.41***	**-0.08***
			Attendance to Motivation			Attendance				
			Attendance to Tests			Motivation	**0.96***	**0.96***	**0.07***	**0.07***
			Attendance to Homework			Homework	**0.93***	0.00	**0.20***	0.00
Pairwise										
	197.87*	44	Part-time to Motivation	**204.27***	48	Part-time	**-9.74***	**-1.39***	**-0.48***	**-0.07***
			Motivation to Tests			Attendance	**-0.45***	**-0.45***	**-0.05***	**-0.05***
			Attendance to Tests			Motivation	**1.55***	**1.55***	**0.11***	**0.11***
			Attendance to Motivation			Homework	**1.61***	0.00	**0.31***	0.00
EM										
	191.98*	44	Part-time to Motivation	**197.96***	48	Part-time	**-9.32***	**-1.36***	**-0.42***	**-0.07***
			Attendance to Motivation			Attendance	**-0.45***	**-0.45***	**-0.05***	**-0.05***
			Attendance to Tests			Motivation	**1.43***	**1.43***	**0.11***	**0.11***
			Motivation to Tests			Homework	**1.51***	0.00	**0.3***	0.00
Regression										
	229.77*	44	Motivation to Tests	**230.95***	46	Part-time	**-3.23***	**-0.93***	**-0.4***	-0.10
			Attendance to Motivation			Attendance	**-1.59***	**-0.42***	**-0.11***	**-0.03***
						Motivation	**1.19***	**1.19***	**0.09***	**0.09***
						Homework	**1.16***	0.00	**0.29***	0.00
Response Pattern										
	212.50*	44	Motivation to Tests	**216.04***	46	Parttime	**-3.09**	**-0.98**	**-0.38**	**-0.12**
			Attendance to Motivation			Attendance	**-1.68**	**-0.68***	**-0.13***	**-0.05**
						Motivation	**1.19***	**1.19***	**0.10***	**0.10***
						Homework	**1.40***	0.00	**0.32***	0.00

Figure 1
Incomplete Cases Grouped by Variable



MCAR $\chi^2 = 646.710$, df = 350, P< .01

Figure 2

Initial Model for Testing

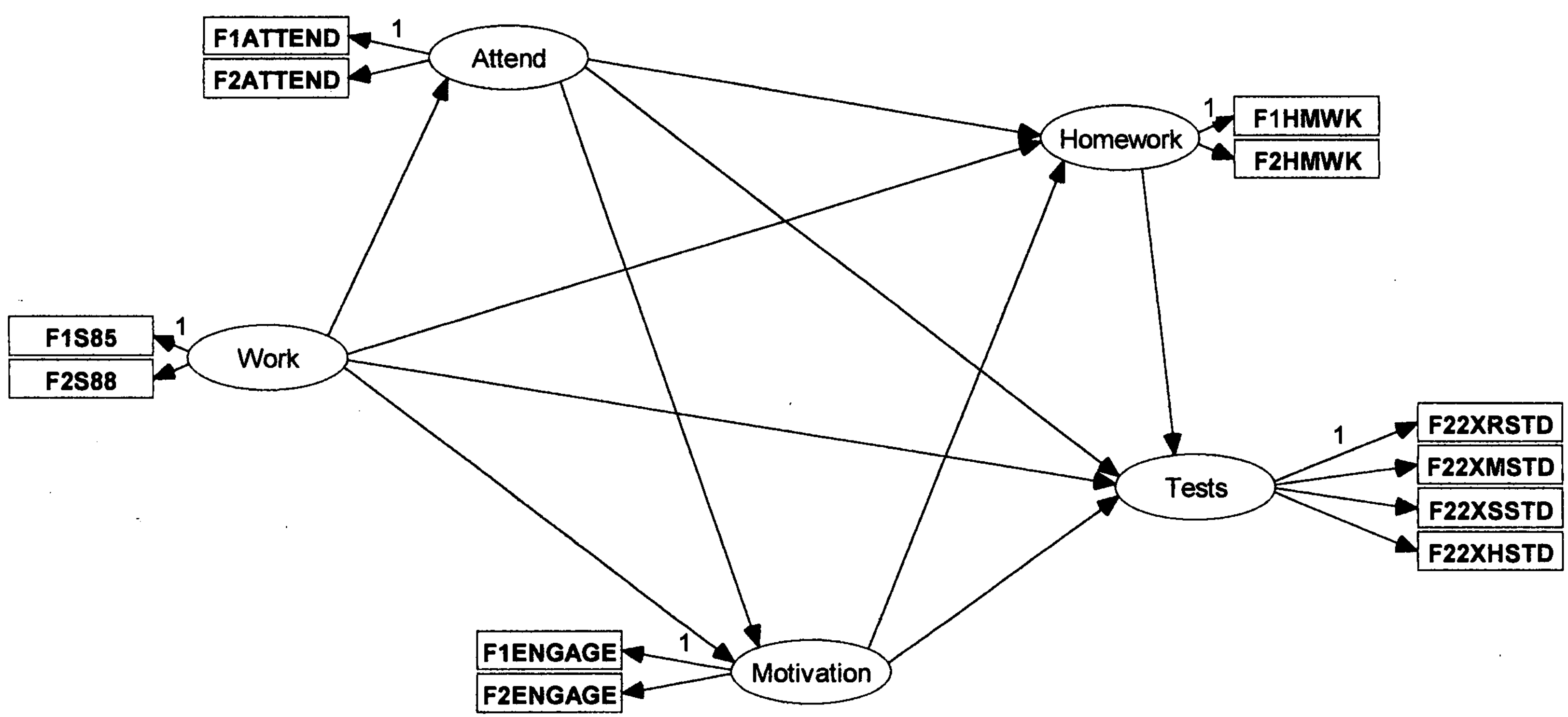


Figure 3

Models Produced when 28% of the Cases were Incomplete

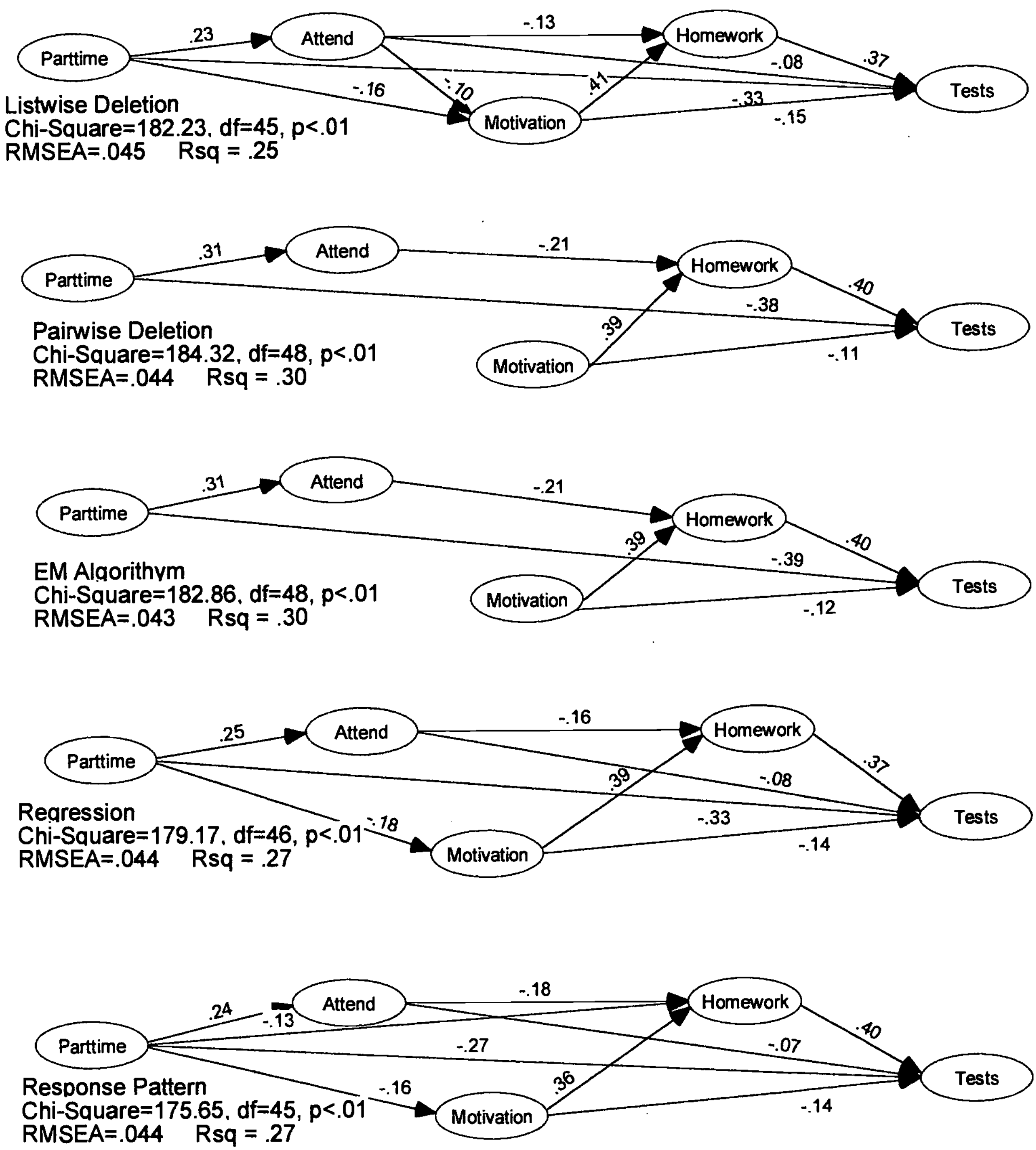
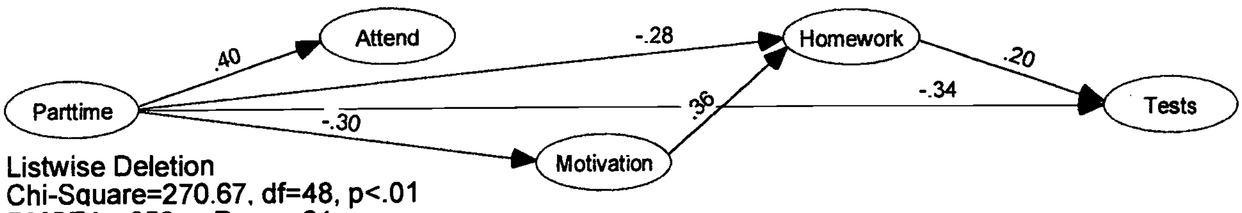
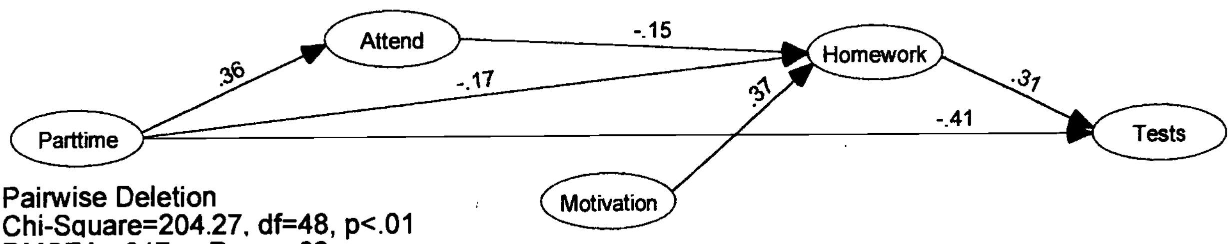


Figure 4

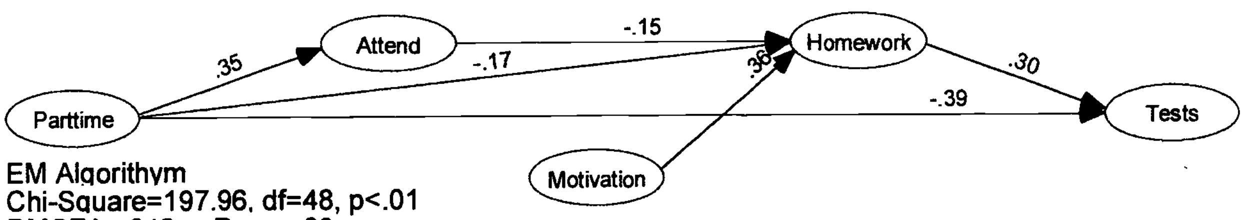
Models produced when 58% of the Cases were Incomplete



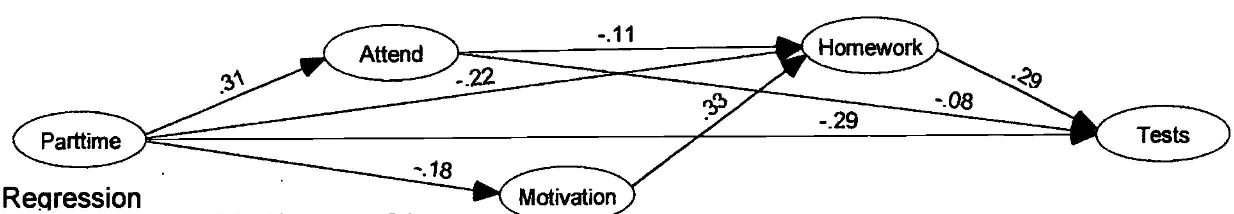
Listwise Deletion
Chi-Square=270.67, df=48, p<.01
RMSEA=.056 Rsq = .21



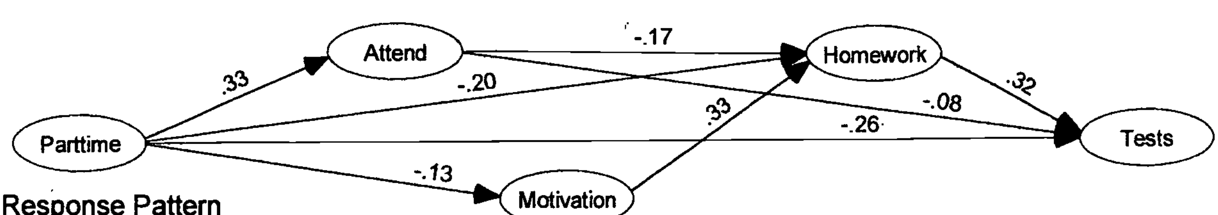
Pairwise Deletion
Chi-Square=204.27, df=48, p<.01
RMSEA=.047 Rsq = .32



EM Algorithym
Chi-Square=197.96, df=48, p<.01
RMSEA=.046 Rsq. = .30



Regression
Chi-Square=230.95, df=46, p<.01
RMSEA=.052 Rsq = .25



Response Pattern
Chi-Square=216.04, df=46, p<.01
RMSEA=.050 Rsq. = .25

Paper presented at SERA

TM032848

ERIC

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Does method of handling missing data affect results of a structural Equation Model	
Author(s): E. Lea Witta	
Corporate Source: Southwest Educational Research Association Conference	Publication Date: February 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: E Lea Witta	Printed Name/Position/Title: E. Lea Witta	
Organization/Address: University of Central Florida, PO Box 161250, Orlando, FL 32816-1250	Telephone: 407-823-3220	Fax: 407-823-5141
	E-mail Address: lwitta@mail.ucf.edu	Date: April 2001

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706
Telephone: 301-552-4200
Toll Free: 800-799-3742
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)