Linear Regression Viewpoints
lume 12, Number 2 Fall 1983

A GENERAL APPROACH FOR TESTING FOR CORRELATED ERRORS IN LONGITUDINAL DATA

Joan K. Gallini
University of South Carolina

Abstract

The present study utilized structural equation methods (LISREL) to estimate models of the pre-posttest paradigm. The data set comprised a group of 6 - 8th grade students involved in a gifted and talented program.

Two types of analyses were conducted. The first analysis was applied to test the validity of Bloom's taxonomy underlying performance on the achievement measure used in the program, the Ross Test of Higher Cognitive Thinking Skills. For the most part, the results demonstrated the existence of the structure, such that analysis skills were preordered with respect to synthesis and evaluation skills.

The second LISREL analysis was applied to assess the model of "best fit" among a set of alternative models that varied in the correlations specified among the measurement errors. There was a significant improvement in model fit when measurement errors were allowed to correlate, as compared to the zero correlation specification on the errors in the null structure. Generally speaking, a nonzero covariation specification between errors associated with all similar measures across the two occasions resulted in the most efficient estimate of ability change. The study pointed to the efficacy of LISREL-type analyses in longitudinal data.

Presented at AERA 1983, MLR Special Interest Group
Not refereed by editorial staff

1. Introduction

The recent emergence of gifted and talented programs across the nation has given considerable attention to the appropriate identificatio of individuals for such programs. The movement has simultaneously evoked concern over the evaluation of student progress, once students are selected into these differentiated programs. Various questionnaires, norm-referenced standardized tests, and criterion-referenced tests have been developed and/or suggested for identification and evaluative purposes (e.g., Ganopole, 1982; Perrone and Chen, 1982; Renzulli, 1975; Vermilyea, 1981).

A number of different techniques should be employed in the assessment of abilities and skills for gifted programs. A major criteria in the choice of the instruments is that thy reflect the objectives and characteristics of the individual program (McFarland, 1980). For example, a student self-reporting questionnaire might be used to ascertain students' perceptions of activities implemented in the program. A criterion-referenced test, on the other hand, will be more suitable to assess performance on those activities, and probably more appropriate than a global norm-referenced measure of achievement.

Gifted and talented programs in the schools typically focus on the academically or intellectually-gifted dimension of "creativity." (For studies on factors of creativity, see Davis, 1980; Guilford, 1975; Klein, 1982). Thus program objectives will oftentimes emphasize the development of higher cognitive thinking skills. These include such aptitudes as problem awareness (formulating a problem), fluency (generating many and new ideas), flexibility (producing a variety of ideas), problem solving (analyzing situations), and divergent thinking (exploring alternative solutions). Valid standardized tests for evaluating "creative" thinking may be useful to tap such dimensions of a gifted and talented program. However, it is extremely important that the selected instrument(s) represent cognitive thinking levels that are indeed represented in the program's activities.

The Ross Test of Higher Cognitive Processes (Ross and Ross, 1976) is a standardized test used in gifted and talented programs. The Ross is designed to serve a general population and gifted students. The test can validly be used as a screening instrument to identify the academically gifted. It can also be employed in the assessment of special programs emphasizing critical thinking inquiry methods, problem solving, and logical thinking, or the development of more complex thinking skills. Another major use of the Ross test lies in the assessment of individual student performance. Such usage may involve a pre-posttest design to determine growth in a student's higher level thinking skills over a period of time.

While its usage may vary, the overall intent of the test is to measure ability in levels of higher cognitive thinking. The specific levels, named according to Bloom's taxonomy (Bloom, 1964), are analysis, synthesis, and evaluation.

Analysis relates to an individual's ability to break down information into its constituent parts. Synthesis refers to the ability to form a new whole. And evaluation is concerned with the ability to judge the value of material for a given purpose. The learning outcomes in these three levels basically represent higher intellectual levels. Synthesis generally stresses creative behaviors, with emphasis on the formulation of new patterns or structures. The behaviors in the taxonomy were attempted to be categorized by complexity, with analysis representing a lower level than synthesis and evaluation. Evaluation represents the highest level in the cognitive hierarchy.

The present study empirically examined the hierarchical structure of the Ross test (as modified for the particular gifted and talented program under study). Structural equation methods were employed to assess the validity of Bloom's taxonomy of higher skills as a reasonable representation of the test's underlying structure for a given data set. The model represents a confirmatory factor analysis model. That is, the structure hypothesized to underlie performance on the Ross is specified prior to the analysis. Maximum likelihood estimation procedures (Joreskog, 1969; 1970 ; Joreskog and Lawley, 1968); of the LISREL computer program (Joreskog and Sorbom, 1978; 1981) are then applied directly to the model specifications.

A second objective of the study was to assess student change in higher cognitive skills ability over the duration of the program. A series of "nested" structural models described by the pre-posttest paradigm were formulated. The models primarily varied in restrictions on correlations between the errors associated with the measurements across the two occasions. The LISREL analyses were employed to identify the model of "best fit" in describing changes in ability (Sorbom, 1979).

2. The Instrument

The Ross test consists of 105 items designed to assess the higher level thinking skills, analysis, synthesis, and evaluation. A description of the subtests and the number of items are shown below by cognitive level.

<u>Analysis.</u> Three subtests relate to the analysis level:
(1) analogies - ability to perceive analogous relationships between pairs of words (14)
(2) missing premises - ability to identify the missing premises needed to complete a logical syllogism (8)
(3) analysis of relevant and irrelevant information - ability to analyze data (14).

<u>Synthesis.</u> Three subtests relate to the synthesis level:
(1) abstract relations - ability to study data and synthesize a logically consistent scheme for organizing them to form a conceptual framework (14)
(2) sequential synthesis - ability to organize sentences in proper sequence (10)
(3) analysis of attributes - ability to formulate and appropriately modify hypotheses (10).

In the present study sequential synthesis was excluded. An "evaluation of arguments" subtest of five items from the Watson-Glaser Critical Thinking Test (Watson-Glaser, 1964) was substituted, since the items related more appropriately to the objectives of the particular gifted program in the present study. The evaluation of arguments items sample ability to distinguish between strong/relevant and weak/irrelevant arguments to a particular question at issue.

Evaluation. Two subtests relate to the evaluation level: (1) deductive reasoning - ability to analyze statements in logic (18) (2) questioning strategies - ability to evaluate methods of obtaining data (17).

3. The Models

A structural equation model was formulated to represent the structure of Bloom's taxonomy said to underlie performance on the Ross test. See Figure 1. The model shows three constructs, analysis, synthesis, and evaluation, so ordered by complexity. The model relates three indicators for analysis and synthesis, and two for evaluation. Descriptions of the indicators has been provided in the previous section. Posttest scores on the Ross, representing student achievement after nine months in the program, constituted the data base for this analysis.

The change in ability models are shown in Figures 2 and 3. The subtests of the three skills domains - analysis, synthesis, and evaluation - were aggregated. Thus, the basic structure reflects pre and posttest as latent variables, with the same three domains as indicators for both occasions.

Model A specified zero covariations between errors (measurement errors or residuals existing after variance due to true scores is extracted). Model B allowed errors to correlate between measures on the same occasions. The assumption is that the tests measure common traits not included in the model. Models, C, D, and E additionally allowed for covariations between the same measures across the occasions. Such covariation is likely to exist in longitudinal data, due to the correlation between the abilities at the two occasions as well as extraneous influences due to the similarity of testing at both times (e.g., recall of test items; practice effect; nature of the tests themselves).

4. Techniques

The study's sample constituted 337 6-8th grade students enrolled in a gifted and talented program in a southeastern school district. The program is oriented toward academically gifted children. Among other selection criteria, students in the program scored at the 95th or higher percentile on the Comprehensive Test of Basic Skills (CTBS) (McGraw-Hill, 1974).

Pre and posttest administrations of the Ross test took place in the Fall 1981 and Spring 1982. The maximum likelihood estimation procedures of LISREL[1] were applied to the posttest data to evaluate the hierarchical structure said to underlie test performance.

1 The LISREL IV computer program was used in the present analysis since LISREL V was unavailable at the time of the analyses.

Additional LISREL analyses were conducted to assess pre-posttest changes across the two occasions.

5. Results

Table 1 shows the means and standard deviations for the subtests of the analysis, synthesis, and evaluation skills domains. The covariance matrix for the variables appears in Table 2.

The LISREL estimates (unstandardized) have been included in Figure 1, along with their associated standard errors. Some of the error variance estimates are questionable, due to the relatively large standard errors. The γ_{11} path (analysis to synthesis) loses importance, as well, with the large standard arror.

Overall, however, the model was accepted as a plausibla structure for the data, based on the obtained chi-square value of 18.116, df=17, (p=.38). Nota that the path diagram portrays a nonzero correlation across the equations. The specification was necessary to obtain a tenable structure. Conceptually, the specification was reasonable since the data were longitudinal. A zero specification resulted in a poor model fit, with a chi-square of 31.773, df=18 (p=.02).

The second set of analyses was conducted to determine the model that best represented ability change among the studants during their participation in the program. Table 3 shows the means and standard

Table 1
Means and Standard Deviations
Higher Cognitive Thinking Skills Variables

Variable	Mean	Standard Deviation
Analogies	10.715	1.753
Missing Premises	4.59	1.817
Irrelevant Information	6.598	2.539
Abstract Relations	11.569	2.291
Attribute Analysis	10.263	2.576
Evaluation of Arguments	3.161	1.373
Deductive Reasoning	13.445	2.882
Questioning Strategies	8.751	1.834

Table 2
Covariance Matrix for the Subtests of
the Higher Cognitive Thinking Skills Instrument

	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4	Y_5
X_1	3.073							
X_2	1.098	3.301						
X_3	1.627	1.698	6.448					
Y_1	.796	.767	1.399	5.247				
Y_2	.693	.790	2.797	1.232	6.636			
Y_3	.215	.072	.433	.481	.693	1.886		
Y_4	1.495	1.915	1.820	1.201	1.801	.481	8.308	
Y_5	1.157	1.032	1.804	1.025	.867	-.099	1.611	3.364

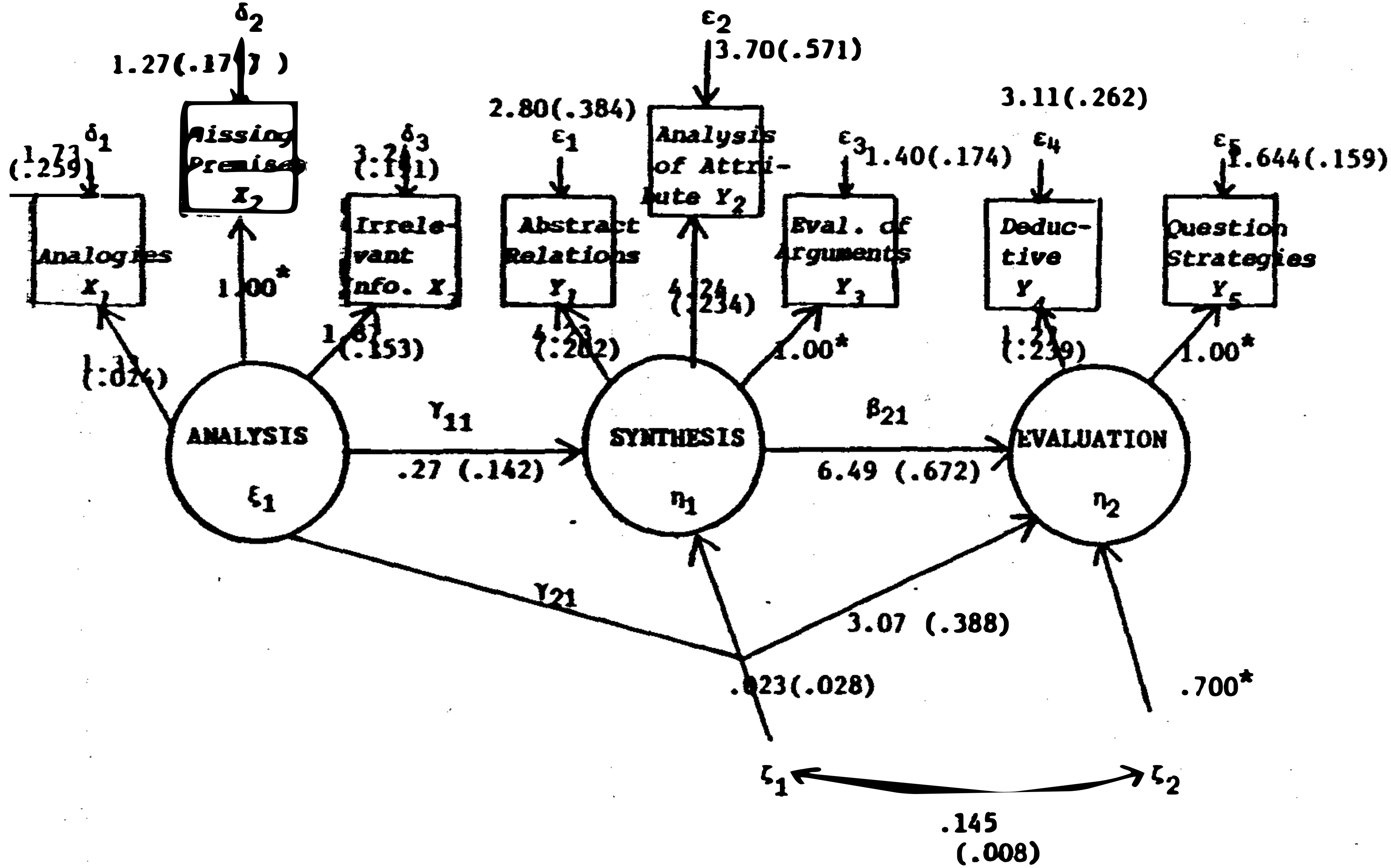


Figure 1. Structural Equation Model for High Cognitive Thinking Skills

*fixed paths

Table 3
Means and Standard Deviations
of the Three Skills
Domains for Pre and Post Administrations

Domain	Mean	Standard Deviation
Pre-Analysis	21.912	4.65
Pre-Synthesis	21.832	3.79
Pre-Evaluation	22.197	3.86
Post-Analysis	25.474	4.15
Post-Synthesis	23.912	3.18
Post-Evaluation	24.289	3.039

Table 4
Covariance Matrix for the Cognitive
Skills Domains of Both Occasions

	Pre			Post		
	Analysis 1	Synthesis 2	Evaluation 3	Analysis 4	Synthesis 5	Evaluation 6
1.	21.669					
2.	7.242	14.347				
3.	9.223	4.894	14.894			
4.	13.894	5.832	7.662	17.266		
5.	5.928	4.711	3.586	6.410	11.007	
6.	7.560	2.830	6.010	'7.796	4.271	9.237

deviations for the pre-and post-analysis, synthesis, and evaluation skills. Table 4 provides the covariance matrix for the variables.

Each pre-posttest model structure was evaluated for overall fit to the data. The individual chi-square values obtained from LISREL are reported in Table 5.

The zero covariation specification in the variable errors for pre and posttest administrations led to a rejection of both models A and B ($p < .01$). See Figure 2. This is not surprising, given the nature of the data. In addition, the first-order derivatives suggested most of the off-diagonal terms in the error covariance matrix were nonzero.

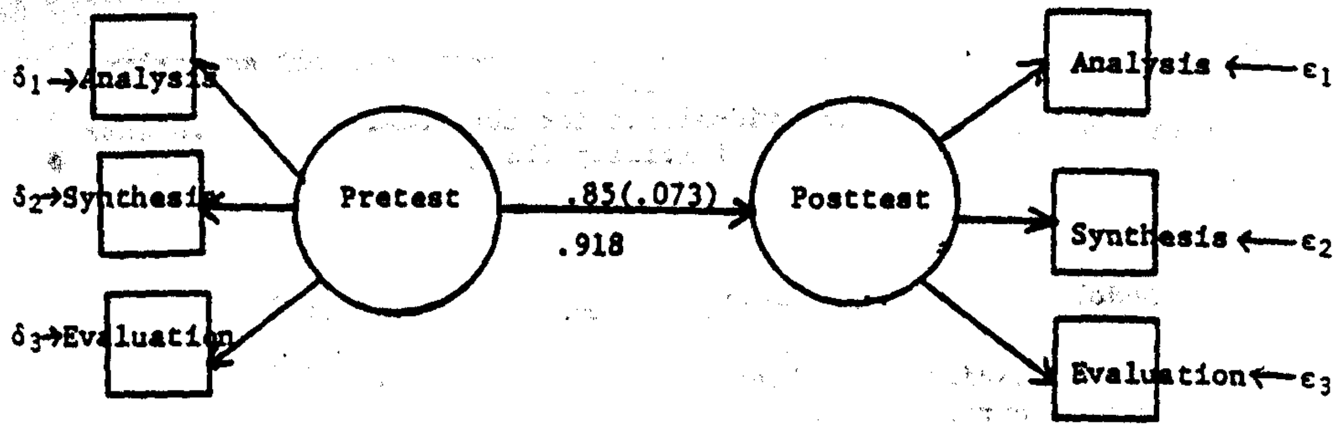
Model C included a correlation between the errors in the analysis variable at both times, but the model was rejected (marginally at $p < .01$). Therefore, it became necessary to add a covariation between the synthesis errors, as shown in model D. (See Figure 3. However, model E yielded the best fit in allowing for covariations between all three similar measures across time.

All four models (B-E) that allowed for some degree of correlation between the variable errors offered significant improvement over model A, with model E showing the greatest degree of improvement for the associated degrees of freedom (chi-square difference of 30.467, df=5). Table 6 shows the chi-square difference statistics

Table 5
Individual Fit for the Models
of Ability Change

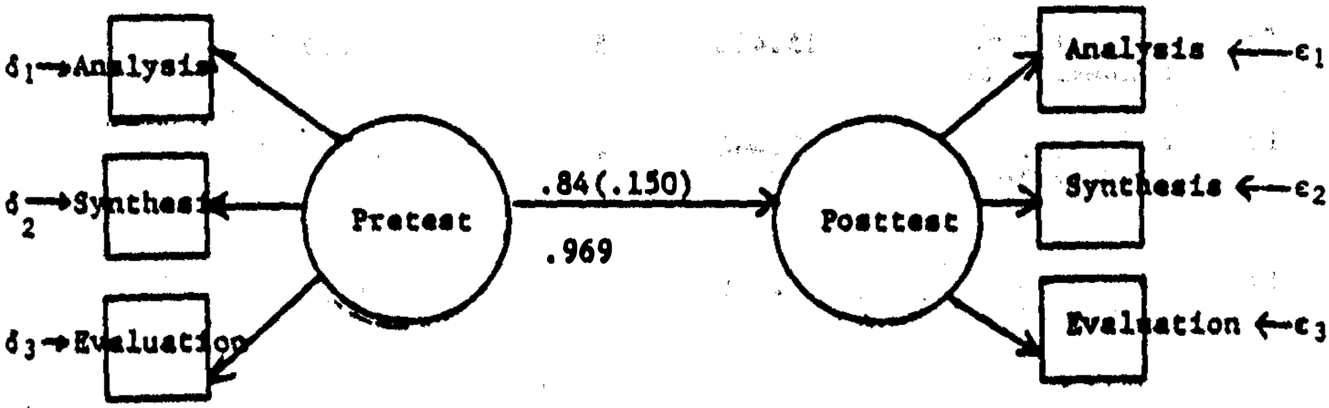
	Model	χ^2-value	df	p-value
A:	no correlations between errors	30.662	8	.0002
B:	error correlations within occasions	25.275	6	.0003
Error Correlations Across Occasions:				
C:	correlation between $\delta_1 \epsilon_1$	15.433	5	.0087
D:	add $\delta_2 \epsilon_2$ correlation	7.440	4	.1144
E:	add $\delta_3 \epsilon_3$ correlation	.1952	3	.9784

Model A



A: no correlations between measurement errors

Model B

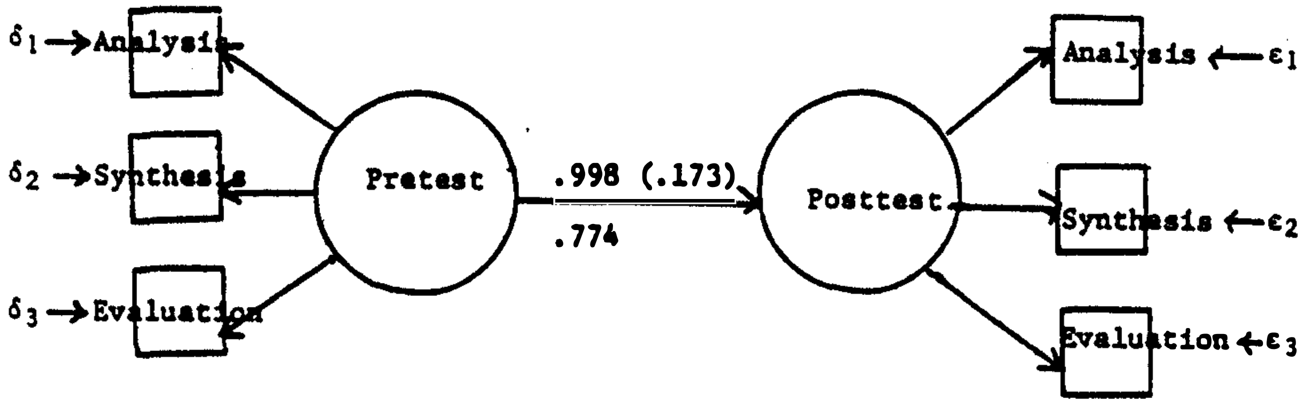


B: correlations between measurement errors within a given occasion

*Figure 2. Pretest-Posttest Structural Equation Models: A null model and an alternative to the null.

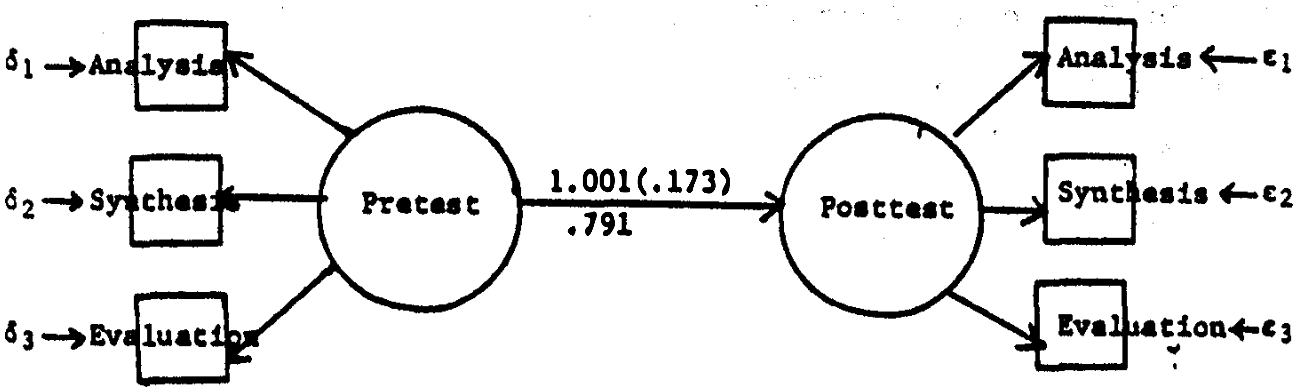
*The ability change path coefficient is shown. The value above the arrow is the unstandardized LISREL estimate and its associated standard error; the value below is the standardized estimate.

Model C



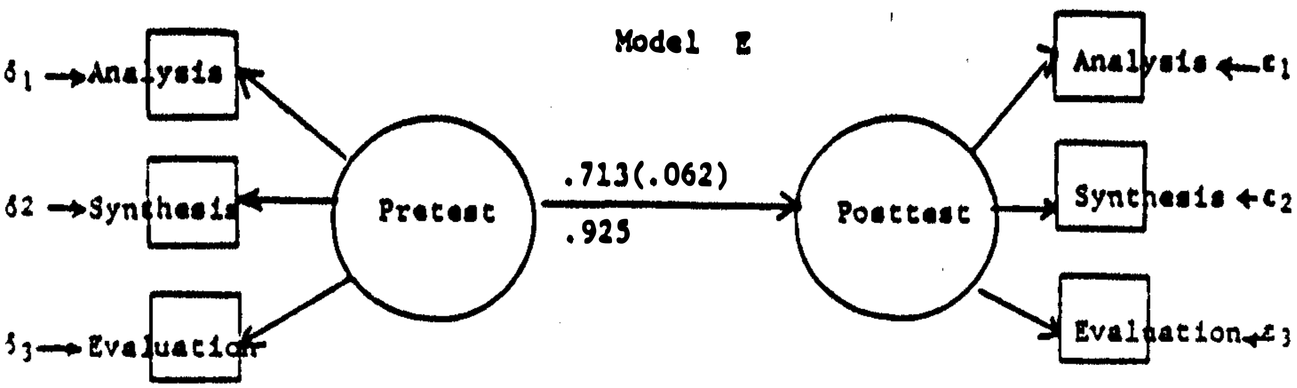
C: $\delta_1\varepsilon_1$ is correlated.

Model D



D: The $\delta_2\varepsilon_2$ correlation is added.

Model E



E: The $\delta_3\varepsilon_3$ correlation is added.

*Figure 3. Pretest-Posttest Structural Equation Models: Correlated Measurement Errors across the Two Occasions.

*The ability change path coefficient is shown. The value above the arrow is the unstandardized LISREL estimate and its standard error; the value below is the standardized LISREL estimate.

Table 6
Model Comparisons

Model Comparison	difference	df*
A - B	5.387	2
A - C	15.229	3
A - D	23.222	4
A - E	30.467	5
B - C	10.139	1
B - D	17.835	2
B - E	25.077	3
C - D	7.993	1
C - E	15.2378	2
D - E	7.2448	1

*$p < .01$, with the exception of the A-B comparison ($p < .05$)

for the model comparison tests.

6. Discussion

The study's findings lent support to Bloom's taxonomy of higher cognitive thinking skills as the underlying structure of the Ross test (as modified for the particular program). Further analyses should follow in an effort to obtain more efficient estimates.

Confirmatory factor analytic methods were applied to a pre-specified model representing the hierarchy of analysis, synthesis, and evaluation skills. The direct path between synthesis and evaluation was relatively large. The results gave credence to the theoretical notion about vertical learning (Gagné, 1970). That is, the learning of subordinate rules facilitate the occurrence of higher-order learning. More importantly, the results have implications for the importance of proper sequencing of instruction so that such transfer of learning does occur.

There was some departure, however, in support to the above premise. Reference is made here to the weak direct path between analysis to synthesis. It is possible that some modifications made in the synthesis component of the Ross test for the given sample could in part have contributed to the departures. Another plausible explanation is perhaps the implemanted curriculum did not necessarily precede synthesis activities with analysis activities.

Additional hypotheses were investigated regarding correlation between the errors in the longitudinal data set. Conventional models used in data analysis (i.e., classical factor analysis, single-equation regression) impose restrictions on the error terms. However, the general LISREL approach to covariance structural analysis can be used to detect nonzero covariations between error terms and consequently identify the measurement model of "best" fit. The evaluations of longitudinal data sets generally offer better representations of reality by allowing correlations across the observed occasions. That is, the effects of previous testing are likely to carry over to the posttest situation and contribute to nonrandom or systematic error. In the present models, the ability change path coefficient (standardized) ranged from a .774 to a .969. The model that allowed for correlations between all three similar measures across the two occasions yielded a relatively higher effects path than did the model with the zero correlation constraint. More importantly, the standard error associated with the alternative structure was smaller. The consideration of information that contributes to systematic error in the estimations should result in a more accurate assessment of ability change over time.

References

Bloom, B. S., ed. Taxonomy of educational objectives; The classification of educational goals, Handbook I; Cognitive domain. New York: Longmans, Green and Co., 1956.

CTB/McGraw-Hill. Comprehensive tests of basic skills. Monterey, CA: CTB/McGraw-Hill, 1975.

Davis, GA.A and Sullivan, M.I. Taxonomy of creative objectives: The Model AVTA. The Journal of Creative Behavior, 1980, 14, 149-160.

Gagne, R.M. The conditions of learning. New York: Holt, Rinehart, and Winston, Inc., 1970.

Ganopole, S.J. Measuring the educational outcomes of gifted programs. Roeper Review, September 1982, 5, 4-7.

Guilford, J.P. Three faces of intellect. In W.B. Barbe and J.S. Renzulli (Eds.). Psychology and Education of the Gifted (2nd ed.). New York: Irvington, 1975.

Joreskog, K.G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-201.

Joreskog, K.G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Joreskog, K.G. and Lawley, D.N. New methods in maximum likelihood factor analysis. The British Journal of Mathematical and Statistical Psychology, 1968, 21, 85-96.

Joreskog, K.G. and Sorbom, D. LISREL IV - A general computer program for estimating a linear structural equation system. Chicago, IL: International Educational Resources, 1978.

Joreskog, K. G. and Sorbom, D. LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods. Chicago: National Educational Resources, 1981.

Klein, R.D. An inquiry into the factors related to creativity. The Elementary School Journal, 1982, 82, 256-265.

Perrone, P.A. and Chen, F. Toward the development of an identification instrument for the gifted. Roeper Review, September 1982, 5, 45-48.

Ross, J.D. and Ross, C.M. Ross test of higher cognitive processes. Novata, CA: Academic Therapy Publications, 1976.

Renzulli, J.S. A guidebook for evaluating programs for the gifted and talented. Ventura, CA: Office of the Ventura County Superintendent of Schools, 1975.

Sorbom, D. Detection of correlated errors in longitudinal data. In Advances in factor analysis and structural equation models by R. G. Joreskog and D. Sorbom (eds.), Cambridge, MA: Abt books, 1979, 171-184.

Vermilyea, J. Common Sense in the identification of gifted and talented students who need alternative programming. G/C/T, January-February 1981, 11-16.