DOCUMENT RESUME

ED 216 896 SE 037 873

AUTHOR Ekstrand, Judith M.
TITLE Methods of Validating Learning Hierarchies with Applications to Mathematics Learning.
PUB DATE Mar 82
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, March 18-23, 1982). Document is marginally legible.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Educational Research; *Elementary School Mathematics; Factor Analysis; Grade 5; Intermediate Grades; *Learning Processes; *Mathematics Achievement; Mathematics Education; Path Analysis; Quantitative Tests; *Vertical Organization
IDENTIFIERS *Learning Hierarchies; *Mathematics Education Research

ABSTRACT

The relationship between mathematics tests and the theoretical learning process was explored using alternative statistical methods and models. Data for over 1300 students in grade 5 using the mathematics subscales from the National Longitudinal Study of Mathematical Abilities (NLSMA) were analyzed. Results indicated that Bloom's taxonomy is weakly supported when a full model using both adjacent and non-adjacent paths (from path analysis) is examined. The factor analysis produced one factor, as Bloom's structure would dictate, but when the four factors accounting for the largest percentage of variance are examined, a factor structure was obtained which corresponded to only two factors that could be named, computation and problem solving. Analyses using the computer program LISREL IV, which accounts for large error components and sample sizes, supported Bloom's cumulative hierarchy. (MNS)

# Methods of Validating Learning Hierarchies With Applications to Mathematics Learning

by

Judith M. Ekstrand

Paper presented at the Annual Meeting of the American Educational Research Association. New York City, March 1982.

Mathematical Methods of Educational Research

Stanford University School of Education







## BACKGROUND AND PURPOSE OF THE STUDY

Mathematics tests yield direct measures of students' learning in mathematics. They, also, reflect unmeasurable cognitive processes, that is intellectual abilities such as remembering facts, analyzing components, integrating concepts, making assumptions, evaluating statements or solutions. In this study, the relationship between mathematics tests and the theoretical learning process is explored using alternative statistical methods and models to mathematically describe relationships between observed test scores and hypothetical constructs which we assume contribute to the observed test scores. The test results used are results of the mathematics scales (ie. subtests) from the National Longitudinal Study of Mathematical Abilities (NLSMA, 1962-1967) for grades 5, 8, 11, and 12.

The test items were constructed following a cumulative hierarchical structure based on the first four categories in Bloom's Taxonomy of Educational Objectives for the Cognitive Domain (Bloom et al, 1956). Bloom's classification system builds from simple behaviors to complex behaviors where simple behaviors are integrated with other simple behaviors to form a more complex one. The four classifications used here are:

| Classification | Behavior |
|---|---|
| 1. Computation (Knowledge) | A |
| 2. Comprehension | A + B |
| 3. Application | A + B + C |
| 4. Analysis | A + B + C + D |

A represents the behavior that is required to successfully answer a test-item requiring only Computation. That, in addition to another behavior, B, is required to succeed at a Comprehension-level item. Thus a higher level is build from the previous level plus one additional behavior in a model that is additive, rather than, say multiplicative.

Data that correspond to this cumulative structure have a correlation matrix that exhibits Guttman's (1954) simplex structure where adjacent categories are more highly correlated than non-adjacent categories. Thus, a perfect simplex has a correlation matrix such that the largest values lie along the first off-diagonal. The next largest values lie along the adjacent diagonals. The smallest values will then lie in the upper-right and lower-left corners of the matrix.

A similar hierarchy is found in mathematics learning. Except for some basic definitions and assumptions, mathematics is built upon other prerequisite mathematics regardless of the content area or grade level. It seems natural, then, to believe that in order to learn a particular mathematical concept or fact, a student needs to know all of the concepts below that one in the hierarchy. Thus, lower levels of mathematics are the building blocks of higher levels of mathematics. It is this idea that is the genesis of a hierarchy of learning in mathematics. See Appendix A for a review of the literature.

Previous attempts to validate Bloom's cumulative hierarchy of learning in several different subject areas have been inconclusive (Seddon, 1978). There is supportive evidence, however, particularly from the simplex analysis of Kropp and Stoker (1966), the multiple regression analysis of Madaus, Woods, and Nuttall (1973), and most recently the reanalysis by Hill and McGaw (1981) of the Kropp and Stoker data using Jöreskog's LISREL (1978) program.

Mathematics scales consisting of groups of similar items from one test battery are classified into one of the above cognitive levels. In order to explore the validity of Bloom's cumulative hierarchy path analysis, factor analysis, and the computer program LISREL IV are used to describe the relationships between these four cognitive levels.

The path analysis model assumes there should be direct links between mathematics scales from adjacent cognitve levels and no direct links from non-adjacent levels. The possible paths between the four cognitive levels are illustrated below where adjacent paths are shown as solid arrows and non-adjacent paths as broken arrows. In this model, the mathematics test scores equal the underlying cognitive levels.





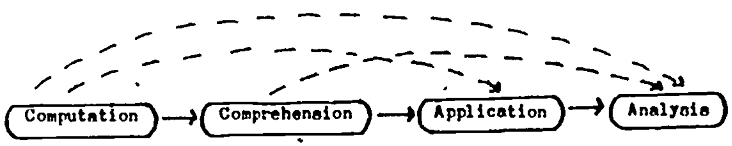



Figure 1. Path diagram illustrating adjacent and non-adjacent paths.

Path coefficients are estimated in linear structural equation models derived from relations among the observed mathematics scales in terms of cause and effect variables and their indicators. Here, the scales are grouped into those scales classified at the same category for a particular grade level.

Factor analysis investigates the theoretical latent structure that could have produced the observed correlation matrix of the mathematics scales. Here, it is the latent factor structure, in contrast to the relationships between observed test scores, that is manipulated. Assuming we have four tests -- one from each of the four cognitive levels under consideration -- a simplex structure in the correlation matrix of these tests would be represented in the factor structure shown below.

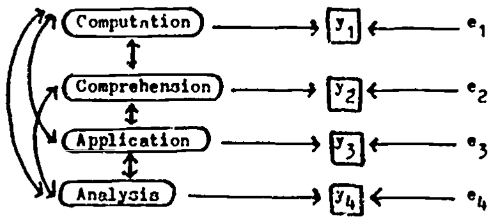
| Test: | Factor: $F_1$ | $F_2$ | $F_3$ | $F_4$ |
|---|---|---|---|---|
| $y_1$ | ---- | 0 | 0 | 0 |
| $y_2$ | ---- | ---- | 0 | 0 |
| $y_3$ | ---- | ---- | ---- | 0 |
| $y_4$ | ---- | ---- | ---- | ---- |

Figure 2. Factor structure of four tests whose correlation matrix exhibits a simplex structure.

For this study, there are usually several tests categorized at each cognitive level giving multiple indicators of each level. Bloom's

structure implies that those tests that were classified under the same cognitive level should exhibit the same factor structure. In this sense, certain mathematics tests go together. We then look for a factor structure that exemplifies the classifications of the tests and the cumulative ordering.

In another factor analytic approach, the computer program LISREL IV integrates linear structural equation models involving observed test scores with latent variables corresponding to the four cognitive levels. Again, we investigate how well the theoretical simplex structure exists in the observed data. However, in this analysis, the relationships between the four cognitive levels from Bloom's hierarchy as latent variables and the observed mathematics test results are separated and made explicit. Furthermore, two models are specified and tested for their goodness of fit using a chi-square statistic. The first model specifies the test classifications according to cognitive level but does not limit the relationships between cognitive levels to Bloom's cumulative ordering. In this model there is complete symmetry between the four cognitive levels in that each one is connected to all of the others. This first model is illustrated in Figure 3.



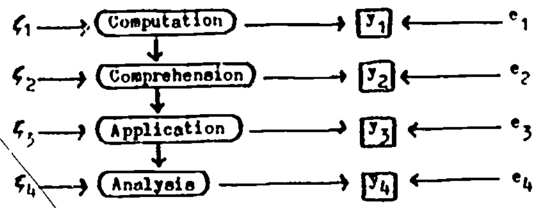

$y_i$ = observed test i (i=1, ... , 4)
$e_i$ = error in test i (i=1, ... , 4)

Figure 3. Hypothetical LISREL model showing all possible paths between the latent variables.





The first model is compared to a second model that differs from the first in that there is Bloom's hierarchical ordering in the four latent variables.




$y_i$ = observed test i (i=1, ... , 4)
$e_i$ = error in test i (i=1, ... , 4)
$\zeta_i$ = error in latent variable i (i= 1, ... , 4)

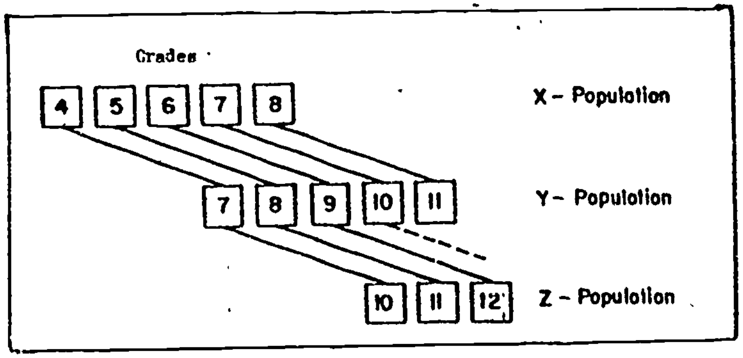
Figure 4. Hypothetical LISREL model showing cumulative hierarchy in the latent variables.

By exploring the application of these three various structural equation estimation techniques to the NLSMA data, each adds to our understanding of whether Bloom's cumulative hierarchical ordering is exhibited in the data. The important question is not whether Bloom's hierarchy gives the only valid explanation of the test results, but rather in exploring the validity of Bloom's hierarchy how do these statistical methods contribute to our understanding of the test results.

## THE NLSMA DATA

The National Longitudinal Study of Mathematical Abilities (NLSMA) began in 1962 to collect data on over 112,000 students from 1,500 schools from 40 states in the United States. Data was collected for 5 years on students in grades four through twelve. This continues to be the largest such study of mathematical achievement in this country.

Three populations of students were studied. The figure below and the accompanying quotation from NLSMA describes the design of NLSMA.




The figure above illustrates the design of NLSMA. A large population of students at each of three grade levels was tested in the fall and spring of each year, beginning with grades 4, 7, and 10 in the fall, 1962. The X-Population and Y-Population were tested for five years. The Z-Population was tested for three years and then followed with questionnaires after graduating from high school. The design stressed three features: (1) the long-term study of a group of students - up to five years, (2) study of the same grade level at different times - for instance, grades 7-8 in 1962-64 for the Y-Population and again in 1965-67 for the X-Population, and (3) extensive data on mathematics achievement for grades 4 through 12.

Figure 5. The design of NLSMA. (Reprinted from FOREWORD to all NLSMA Reports.)

Mathematical achievement was characterized in a matrix of three content areas that students typically covered in the fourth through twelfth grade curricula and four cognitive level categories taken from Bloom's taxonomy.

| | Number Systems | Geometry | Algebra |
|---|---|---|---|
| Computation | | | |
| Comprehension | | | |
| Application | | | |
| Analysis | | | |

Figure 6. NLSMA model for mathematics achievement.





Since the goal of the present study is to investigate the validity of the imposed taxonomic structure in the mathematics scales, only those populations that include at least one scale from each cognitive level category are considered. From the thirteen possible grades spread over the NLSMA X-, Y-, and Z-Populations, four samples were found that met this criteria: X-Population, Grade 5 and Grade 8; Y-Population, Grade 11 (group 1); and Z-Population, Grade 12 (group 2). Group 1 in the grade 11 Y-Population includes all students who had completed at least three years of college preparatory mathematics by the end of that year. Approximately 45 percent of the Y-Population, Grade 11 are in group 1. Group 2 in the grade 12 Z-Population includes all students who had not had at least one mathematics course more advanced than geometry. Approximately 30 percent of the Z-Population, Grade 12 are in group 2.

In general, the students were above average in mental ability, mathematics achievment, and socio-economic status. They came from schools from all five geographic regions in the United States -- North Atlantic, Southeast, Midwest, Great Plains and Rocky Mountains, and Far West. Statistical information is presented for a 5 percent stratified (by geography) random sample of each of the entire X-, Y-, and Z-Populations. This sampling procedure yields large sample sizes in the four samples under investigation -- 1776 students in the X-Population Grade 5; 1130 students in the X-Population Grade 8; 515 students in the group 1 Y-Population Grade 11; and 205 students in the group 2 Z-Population Grade 12.

In this paper, we will examine the results for the grade 5 population only. Descriptive statistics for this sample are given in Table 1 and Table 2. Complete test batteries and item statistics for each scale are found in NLSMA Reports Nos. 1 - 6 (Wilson, Cahen, Begle, 1968). Correlation matrices on all of the NLSMA scales are found in NLSMA Report No. 33 (Wilson, Begle, 1972).

TABLE 1

X-Population Mathematics Scales for Spring 1964, Grade 5, Sample Sizes Range From 1326 - 1330.

| Cognitive Level | Math Scale | Number of Items | Time* Allowed | Alpha** | Error of Measurement |
|---|---|---|---|---|---|
| Computation | X301 - Fractions 3 | 10 | 12 min. | 0.87 | 1.11 |
| | X302 - Decimals 2 | 7 | 7 min. | 0.76 | 1.03 |
| | X303 - Division Whole Numbers 2 | 8 | 15 min. | 0.82 | 1.08 |
| Comprehension | X304 - Decimal Notation | 8 | | 0.54 | 1.24 |
| | X305 - Translation | 7 | | 0.69 | 1.06 |
| | X306 - Geometric Figures | 4 | | 0.53 | 0.73 |
| Application | X307 - Working With Numbers | 12 | 20 min. | 0.68 | 1.49 |
| Analysis | X308 - Five Dots | 19 | 15 min. | 0.85 | 1.78 |

* If time information is not given, then the scale is part of a larger timed section.

** Alpha is an estimate of the internal consistency reliability of the scale.





TABLE 2

Means and Standard Deviations for X-Population Mathematics Scales, Spring 1964, Grade 5, N-1776.

| Cognitive Level | Math Scale | Mean | Standard Deviation | Percentile Mean |
|---|---|---|---|---|
| Computation | X301 - Fractions 3 | 5.97 | 3.07 | 0.60 |
| | X302 - Decimals 2 | 3.20 | 2.11 | 0.46 |
| | X303 - Division Whole Numbers 2 | 4.83 | 2.54 | 0.60 |
| Comprehension | X304 - Decimal Notation | 3.86 | 1.83 | 0.48 |
| | X305 - Translation | 4.82 | 1.87 | 0.67 |
| | X306 - Geometric Figures | 2.57 | 1.07 | 0.64 |
| Application | X307 - Working With Numbers | 5.78 | 2.59 | 0.48 |
| Analysis | X308 - Five Dots | 10.74 | 4.77 | 0.25 |

From Table 1 we note that the small number of items in each scale yield less reliable scale scores. Also, there are not equal numbers of items in each scale. There are not equal numbers of scales for each cognitive level category. In general, over the entire NLSMA population, there are more scales at the lower cognitive level categories in the lower grades and more scales at the higher cognitive level categories in the higher grades corresponding to the more abstract quality of the material being taught at the higher grades.

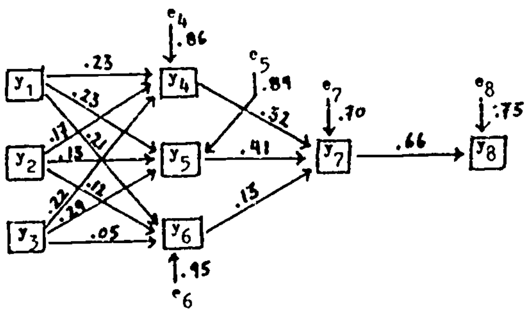
From Table 2 we note that the percentile means generally decrease from lower to higher cognitive level category which is consistent with Bloom's taxonomic structure.

## RESULTS FOR X-POPULATION, GRADE 5

### Results Using Path Analysis.

Using the NLSMA classification of the mathematics scales, a path diagram is used to display the pattern and strength of the causal relations between scale scores from adjacent cognitive levels. The variabilities of the independent scales at the lowest level, Computation, are all assumed determined by causes outside the model. The variabilities of the dependent scales at the other three levels are explained by the previous independent variables or by other dependent variables.

Each path diagram represents a model that consists of a set of equations each of which includes a disturbance term that summarizes the effect on the structure of the system of both measurement error for that equation and all other unknown variables. Each model is assumed to represent linear relationships between the variables and the disturbance terms where the disturbances are independent of each other and of all variables that precede them in the given causal ordering.




$y_1$=Fractions $y_2$=Decimals, $y_3$=Division of Whole Numbers, $y_4$=Decimal Notation, $y_5$=Translation, $y_6$=Geometric Figures, $y_7$=Working With Numbers, $y_8$=Five Dots.

Figure 7. Path diagram in a model using tests from Number Systems and Geometry, Grade 5, N=1776.





Path coefficients, $p_{ij}$, where $p_{ij}$ represents the path from test $y_j$ to test $y_i$ are given in Figure 7 above. This diagram represents the following five regression equations.

$$
(1) \qquad
\begin{aligned}
y_4 &= p_{41} y_1 + p_{42} y_2 + p_{43} y_3 + e_4 \\
y_5 &= p_{51} y_1 + p_{52} y_2 + p_{53} y_3 + e_5 \\
y_6 &= p_{61} y_1 + p_{62} y_2 + p_{63} y_3 + e_6 \\
y_7 &= p_{74} y_4 + p_{75} y_5 + p_{76} y_6 + e_7 \\
y_8 &= p_{87} y_7 + e_8
\end{aligned}
$$

Relationships between the regressed variables $y_1$, $y_2$, $y_3$ may be presented in a correlation matrix as shown in Table 3.

TABLE 3

Correlation Matrix For Computation Scales, Grade 5.

| Test: | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $y_1$ | 1.00 | | |
| $y_2$ | 0.49 | 1.00 | |
| $y_3$ | 0.59 | 0.45 | 1.00 |

Since the correlations are computed on standardized variables, the regression coefficients are identical to the path coefficients obtained from solving a series of linear structural equations. Using the model described in Figure 7, we obtain estimates for the path coefficients as solutions of three systems of equations. From the equations

$$
(2) \qquad
\begin{aligned}
r_{14} &= p_{41} + p_{42} r_{12} + p_{43} r_{13} \\
r_{24} &= p_{41} r_{12} + p_{42} + p_{43} r_{23} \\
r_{34} &= p_{41} r_{13} + p_{42} r_{23} + p_{43}
\end{aligned}
$$

we obtain estimates of $p_{41}$, $p_{42}$, $p_{43}$; from the equations .

$$
(3)\qquad
\begin{aligned}
r_{15} &= p_{51} + p_{52} r_{21} + p_{53} r_{13} \\
r_{25} &= p_{51} r_{12} + p_{52} + p_{53} r_{23} \\
r_{35} &= p_{51} r_{13} + p_{52} r_{23} + p_{53}
\end{aligned}
$$

we obtain estimates of $p_{51}$, $p_{52}$, $p_{53}$; from the equations

$$
(4)\qquad
\begin{aligned}
r_{16} &= p_{61} + p_{62} r_{12} + p_{63} r_{13} \\
r_{26} &= p_{61} r_{12} + p_{62} + p_{63} r_{23} \\
r_{36} &= p_{61} r_{13} + p_{62} r_{23} + p_{63}
\end{aligned}
$$

we obtain estimates of $p_{61}$, $p_{62}$, $p_{63}$. In each system of equations there are exactly three equations in three unknowns, and the quations are independent, so that these systems are identified.

As a consequence of the last two regression equations of equation (1) we obtain the equations

$$
(5)\qquad
\begin{aligned}
r_{17} &= p_{74} r_{41} + p_{75} r_{51} + p_{76} r_{61} \\
r_{27} &= p_{74} r_{42} + p_{75} r_{52} + p_{76} r_{62} \\
r_{37} &= p_{74} r_{43} + p_{75} r_{53} + p_{76} r_{63} \\
r_{47} &= p_{74} + p_{75} r_{54} + p_{76} r_{64} \\
r_{57} &= p_{74} r_{45} + p_{75} + p_{76} r_{65} \\
r_{67} &= p_{74} r_{46} + p_{75} r_{56} + p_{76}
\end{aligned}
$$

Here we have six equations in three unknowns, so that the system is overidentified. Solving this system algebraically cannot give unique estimates. However, the regression procedure does give unique regression coefficients which can be used as unbiased estimates for the path coefficients. Similarly, the last regression equation in (1) corresponds to an overidentified system of seven equations in one unknown.







All but two of the mathematics scales from the Grade 5 sample were originally classified under the content area Number Systems. The other two scales were classified under Geometry. By collapsing across these two content areas, and using ordinary least squares regression, we obtain the estimates given in Figure 7. Estimates for the disturbance terms were computed using $\sqrt{1 - R^2}$ for each regression.

For this sample, we note only one path coefficient ($p_{63}$=0.05) that is close to zero. The mathematics scale $y_6$=Geometric Figures consists of four multiple-choice questions dealing with naming figures that represent a square, circle, rectangle, and triangle. Working With Fractions ($y_1$) and Decimals ($y_2$) may be taught using geometric figures (pie charts, segmented rectangles, etc.) whereas Division of Whole Numbers ($y_3$) typically is not illustrated using geometric figures which may explain the difference in the strengths of these relationships.

None of the path coefficients is very large, which we would expect given the large sample size. We note, however, that the strength of the path coefficients generally increase as we go from low to high cognitive levels. Also, the disturbance terms are large and are inversely proportional to the number of items in the mathematics scale. That is, the largest disturbance terms correspond to those scales with the fewest numbers of items.

Because collapsing the mathematics scales across the two content areas may not be consistent with Bloom's taxonomic structure, similar regressions were carried out using only the scales from Number Systems. The path coefficients obtained are essentially the same.

Regressions were also completed on this sample for the full model, that is, including all of the non-adjacent paths. The non-adjacent paths were all smaller than the adjacent paths; however they were not all zero.

The large error terms in Figure 7, compared to the values of the path coefficients themselves, preclude any strong interpretation of the results of this analysis. Thus we turn to factor analytic techniques to aid us in our understanding of these results.

## Results Using Factor Analysis.

Principal components analysis was used as an exploratory technique to justify using models relating the four underlying hypothetical

cognitive levels of Bloom's Taxonomy as unobserved latent variables to the observed mathematics scale scores. First, corresponding eigenvalues greater than one were extracted. This procedure produced one principal component that accounted for approximately 53% of the total variance. Because the mathematics scales were categorized into one of four cognitive levels, the first four principal components were also examined and rotated hoping to find a structure consistent with the four hypothesized cognitive levels. The first four principal components accounted for approximately 81% of the total variance.

Bloom's cumulative hierarchy implies that one principal component should be sufficient to account for a large percentage of the total variance. Principal components analysis is also used to check the structure of each group of tests in a given cognitive level to determine in any particular scale loads very differently than all the others in that category. (In that case, the results would indicate a misclassification of that scale.) For this sample, all of the tests at a particular cognitive level had similar factor loadings.

Since principal components are orthogonal, and we expect an oblique factor structure, we do not expect to find a structure exemplifying Bloom's hierarchical structure using principal components. We do, however, want to account for as much of the total variance as possible. Therefore, principal components is used to determine the values of the first four eigenvalues in order to specify four factors in a classical factor analysis that will allow oblique factors.

Both orthogonal and oblique rotations are examined. The potentially low reliabilities of the scales limit the strength of the expected structure. The only interpretable structure that was found was to separate the tests into those at the Computation level and those at all the higher levels together, except the one geometry scale, Geometric Figures, produced its own unique factor. Because not all possible oblique factor structures were examined, this analysis does not give a definitive answer to whether or not Bloom's hierarchy is exhited in this data. It does indicate, though, that classifying mathematics tests into those tapping two factors (say, computation and problem-solving) is warranted.





Results Using LISREL IV.

Maximum likelihood factor analytic procedures are used in order to obtain initial values for the parameters required in the LISREL IV computer program. (We note that a new version of the program, LISREL V, produces its own initial values given one initial estimate.)

This computer program provides the analysis for a structural equation model that is used to specify explicitely the behavior of the NLSMA mathematics scales related to a cumulative hierarchical ordering of four cognitive levels. Jöreskog introduced a very general model in 1973. The program description of the specification, estimation, and testing of the model with illustrations from social science research is given in Jöreskog (1977).

The estimates obtained are based on the method of maximum likelihood. Models are examined containing directly observed variables or unobserved hypothetical construct variables. Latent variables are assumed to be related to other observed variables and to each other.

The general model for which LISREL IV was designed is described by the following matrix equations:

(6) $$\beta\eta = \Gamma\xi + \zeta$$

(7) $$y = \Lambda_y\eta + \epsilon$$

(8) $$x = \Lambda_x\xi + \delta$$

where $\beta, \Gamma$ are coefficient matrices; $\eta$, $\xi$ represent independent and dependent latent variables, respectively; $\zeta$ is a matrix of error terms; x, y represent independent and dependent observed variables, respectively, with corresponding matrices of error terms $\epsilon$ and $\delta$. $\Lambda_y$ and $\Lambda_x$ contain regression coefficients of y on $\eta$ and of x on $\xi$, respectively. An estimated covariance matrix is computed whose elements are functions of eight parameter matrices -- $\Lambda_y$, $\Lambda_x$, $\beta$, $\Gamma$, and the covariance matrices of $\xi$, $\zeta$, $\epsilon$, and $\delta$.

The program allows for both errors in the linear equations, including specification errors and disturbance terms, and errors in the observed variables, including measurement errors and observational errors. The program yields estimates of the residual covariance matrix and

and measurement error covariance matrix as well as estimates of the causal effects from the structural equations, provided all the parameters in the model are identified.

We use a special case of the structural equation model where the observed mathematics scales are represented by dependent variables and our hypothesized cognitive levels are represented by four latent independent variables. The structural equation model (6) then reduces to

$$\beta \eta = \zeta \tag{7}$$

and the only vector of observed variables is y.

The methodological goal is to reproduce a covariance matrix whose elements are funcions of the four parameter matrices: $\beta, \Lambda_y$, and the covariance matrices of $\zeta$, and $\epsilon$. Our special case is identical to a factor analysis with the following differences. There is no restriction that there be fewer factors than variables or that the covariance matrix of the residuals be diagonal. The only requirement is that the covariance matrix of dependent observed variables be nonsingular and that the model be identified.

In the identification of the parameters, the assumption is made that the distributions of the observed variables are described by their moments of first and second order. That is, the information in moments of higher order is ignored. This assumption is valid if the distributions are multivariate normal.

In the estimation and testing of the model, it is assumed that the distributions of the observed variables are described by a mean vector and covariance matrix. The problem is to fit the covariance matrix imposed by the model to the sample covariance matrix. In the process, maximum likelihood estimates emerge; such estimates are efficient for large sample sizes. A test of the model is made using the chi-square statistic under the assumptions of multidimensional normality and large sample sizes.

For the grade 5 sample, a comparison is made of how well two models fit the observed correlation matrix of the NLSMA mathematics scales. The first model includes the four cognitive levels but does not contain





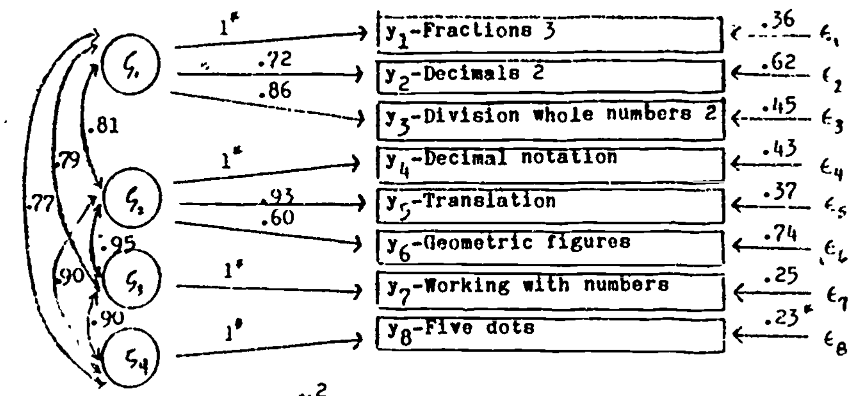
the theoretical cumulative hierarchy. That is, all possible paths between the cognitive levels are included. The second model is the same as the first but is restricted to only include the cumulative hierarchy. In both models, one parameter from each mathematics scale at a given cognitive level is restricted to equal one in order to assess the relative effects of each scale at that level. Chi-square statistics for both models are given. A large decrease in the chi-square statistic compared to a small change in the degrees of freedom indicates an improvement in the model. That is, there is supportive evidence that Bloom's cumulative hierarchical structure does exist in the empirical data.

Because of the stratified cluster sampling procedure used by the NLSMA investigators, these chi-square statistics are known to be too large. The critical issue is not the fit, however; the critical issue is the difference in chi-square statistics.

Figures 8 and 9 show the LISREL results for the Grade 5 sample. Model 1 is the model without the hierarchy and model 2 is the model with the hierarchy. Model 2 was unidentified without fixing the value of the disturbance term $\epsilon_8$. In order to compare these two models, model 1 was re-run restricting this estimate to the value 0.23 that had been obtained when it was allowed to be free. This did not change any of the estimates from model 1 and allowed model 2 to be identified by using this same value for $\epsilon_8$. The chi-square statistic dropped from $\chi^2_{19} = 192.94$ to $\chi^2_{18} = 39.43$ indicating a better model with Bloom's cumulative hierarchy.

Because of the large sample size, the probability value associated with the model including Bloom's structure is not statistically significant even though the residual differences are small indicating a good practical fit. The sampling method also implies that the probability value should also be higher giving an even better fit.

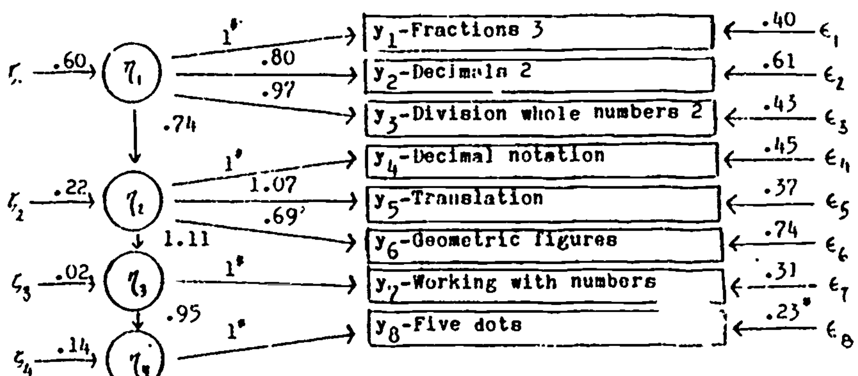
Figure 8. X-Population, Grade 5, N = 1776, Model 1.




$$\chi^2_{19} = 192.94$$

P = 0.0

Figure 9. X-Population, Grade 5, N = 1776, Model 2.




$$\chi^2_{18} = 39.43$$

P = 0.0025

* → fixed





## CONCLUSIONS

Analysis of the means for the Grade 5 sample are limited because of the differences in the numbers of items in each scale and because of the low reliabilities. These low reliabilities due mostly to the small number of items in each scale contribute to the large error terms in the path analysis of this data. However, even with the large error terms, Bloom's taxonomy is weakly supported in this analysis when a full model using both adjacent and non-adjacent paths is examined. The factor analysis produces one factor as Bloom's structure would dictate, but when the four factors accounting for the largest percentage of variance are examined, we get a factor structure corresponding to only two factors not four that could be named computation and problem-solving. Neither orthogonal nor oblique rotations yielded the desired factor structure. However, because any particular oblique factor structure may be difficult to find, the factor analysis here is inconclusive. On the other hand, the analysis using the computer program LISREL IV, that does account for large error components and large sample sizes, does support Bloom's cumulative hierarchy. (We note that in the other grade levels (8, 11, and 12) examined but not reported but not reported in this paper, the analyses using LISREL all gave results supporting Bloom's hierarchy, whereas the path analyses and factor analyses were not conclusive.)

## APPENDIX A

### Review of the Literature on Hierarchies of Learning

In this section we review some of the studies that have, in the last two decades, tried to validate learning hierarchies. Two types of studies emerge. First, especially in mathematics and science education studies, there are those that use Gagné's model of constructing a network of links between higher-order tasks that depend on the mastery of a set of lower-order tasks. The validation techniques are based on analyzing dependencies along the links. The second type of study deals with validating Bloom's taxonomy. In these studies various statistical techniques have been used to test whether a simplex structure exists in the data. There are indications that modifications to the original ordering and/or additional constructs are needed to adequately explain the results of the Bloom taxonomy-type tests.

Capie and Jones (1971) study a series of ratios derived from a phi-correlation matrix to validate a Gagné model. They suggest that the construction of a logical sequence is not necessary when all behaviors considered relevant are measured and each behavior paired with all others for analysis.

White (1973, 1974a, 1974b) discusses three major Gagné investigations as well as several later studies following the Gagné model. He identifies several weaknesses in all the previous studies including the lack of a statistical test that takes into account errors of measurement, small sample sizes, use of only one question per element, delays in testing, and imprecise specification of component elements. The later studies propose ways of correcting these weaknesses. They support the later postulate of Gagné (1968) that generalized intellectual skills are learned hierarchically whereas verbalized knowledge is not. The later studies also illustrate a statistical test developed by White and Clark (1973) for calculating the critical value of the maximum number that could occur in the crucial cell (number of subjects who answer correctly higher-order tasks but not lower-order tasks) for particular probabilities under the hypothesis that the connection between the two elements is hierarchical.





Kropp and Stoker (1966) offer the first major validation study of Bloom's hierarchy. They develop taxonomy-type tests with equal numbers of items for each class. A comparison of means for each class shows that higher means occur in the lower levels, in general, over four subject areas. Their simplex analysis gives some support for a cumulative hierarchical structure though they note problems in interpreting the category 'Knowledge' and have some reversals in the ordering of the categories 'Synthesis' and 'Evaluation.'

Madaus, Woods, and Nuttall (1973) reanalyze a subset of the Kropp and Stoker data with a causal model using multiple regression procedures. They also compute a general mental ability factor ("g"-factor) using principal components analysis on one standardized test. They find many indirect links that would invalidate the assumed hierarchy. All but one of these disappear, however, upon adding the "g"-factor. They also report a decline in the magnitude of direct links as processes progress from simple to more complex behaviors.

Seddon (1978) summarized many of the previous Bloom validation studies in terms of educational and psychological issues. The results he studies are inconsistent. He questions the correlational properties of composite scores used in many of these studies and suggests they could be avoided using factor analysis or smallest space analysis on the correlation matrix of individual items.

Miller (1979) uses path analysis, stepwise regression, commonality analysis and factor analysis in reanalyzing the Kropp and Stoker data. All methods reject a simple hierarchical structure. The factor analysis and commonality analysis suggest a two-factor model. The path analysis suggests a branching in the order of the levels though the ordering of Knowledge -- Comprehension -- Application -- Analysis remains in the original order.

Hill and McGaw (1981) use LISREL to reanalyze a modified version of the Kropp and Stoker data. Their results support the simplex assumption when the category Knowledge is deleted.

These previous studies yield conflicting results. They all, however, reiterate a need for more research into validating learning hierarchies and improvement of the statistical methodology needed to adequately interpret the results.

REFERENCES

Bloom, B.S. (ed.), Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R., Taxonomy of Educational Objectives Handbook One Cognitive Domain. New York and London: Longman, 1956.

Capie, W., & Jones, H.L. An assessment of hierarchy validation techniques, Journal of Research in Science Teaching, 1971, 8 (2), 137-147.

Gagné, R.M. Learning hierarchies, Educational Psychologist, 1968, 6, 1-9.

Guttman, L. A new approach to factor analysis; the radex. In Lazarsfeld, P. (ed.), Mathematical Thinking in the Social Sciences. Glencoe, Illinois: The Free Press, 1964.

Hill, P.W. & McGaw, B. Testing the simplex assumption underlying Bloom's taxonomy, American Educational Research Journal, 1981, 18 (1), 93-.

Jöreskog, K.G. A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (eds.) Structural Equation Models in the Social Sciences. New York: Seminar Press, 1973.

Jöreskog, K.G. & Sörbom, D. Statistical models and methods for analysis of longitudinal data. In D.J. Aigner & A.S. Goldberger (eds.) Latent Variables in Socioeconomic Models. Amsterdam: North Holland Publishing Co., 1976.

LISREL IV, Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources, 1978.

Kropp, R.P. & Stoker, H.W. The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives (Cooperative Research Project No. 2117) Institute of Human Learning and Department of Educational Research and Testing, Florida State University, February, 1966. (ERIC ED 010 044)

Madaus, G.F., Woods, E.M. & Nuttall, R.L. A causal model analysis of Bloom's taxonomy, American Educational Research Journal, 1973, 10 (4), 253-262.

Miller, W.G., Snowman, J., & O'Hara, T. Application of alternative statistical techniques to examine the hierarchical ordering in Bloom's taxonomy, American Educational Research Journal, 1979, 16 (3), 241-248.

Seddon, G.M. The properties of Bloom's taxonomy of educational objectives for the cognitive domain, Review of Educational Research, 1978, 48 (2), 303-323.

White, R.T. Research into learning hierarchies, Review of Educational Research, 1973, 43, 361-375.





White, R.T. The validation of a learning hierarchy, American Educational Research Journal, 1974a, 11 (2), 121-136.

Indexes used in testing the validity of learning hierarchies, Journal of Research in Science Teaching, 1974b, 11 (1), 61-66.

White, R.T., & Clark, R.M. A test of inclusion which allows for errors of measurement, Psychometrika, 1973, 38, 77-86.