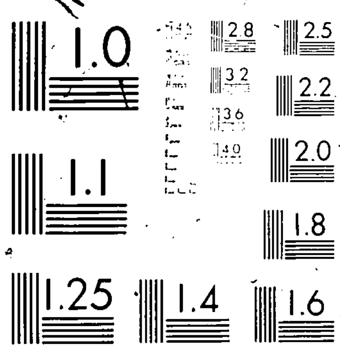
THRU



MICROCOPY RESOLUTION TEST CHART

DOCUMENT RESUME

ED 169 082 TM 008 391

AUTHOR Gustafsson, Jan-Eric; Lindstrom, Berner
TITLE Describing and Testing Aptitude-Treatment Interaction Effects with Structural Equation Models: Reanalysis of a Study by M. J. Behr. No. 73.
INSTITUTION Gothenburg Univ. (Sweden). Inst. of Education.
SPONS AGENCY National Swedish Board of Education, Stockholm.; Swedish Council for Social Science Research, Stockholm.
PUB DATE Dec 78
NOTE 50p.; For Behr's Study, see EJ 039 647 ; Parts may not reproduce clearly

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Aptitude; *Aptitude; *Computer Programs; *Data Analysis; Elementary School Teachers; *Factor Analysis; *Learning; *Mathematical Models; Mathematics Instruction; Multiple Regression Analysis; Predictor Variables
IDENTIFIERS *Aptitude Treatment Interaction; *LISREL Computer Program

ABSTRACT

The Joreskog and Sorbom LISREL (linear structural relations) method is investigated as an alternative to regression analysis in studies of aptitude-treatment interactions (ATI), to solve problems caused by unreliability of measurements and by large sets of variables. A study reported by M.J. Behr is reanalyzed. This study investigated relations between verbal and figural aptitude variables and outcome variables within one verbal and one figural treatment, teaching modulus seven arithmetic to prospective elementary teachers. The relations between latent aptitude variables and latent outcome variables are studied in four LISREL models with either one or two latent aptitude variables (interpretable as a general factor and verbal and figural ability, respectively), and with two sets of latent outcome variables (interpretable as learning/retention and computation speed/understanding). No significant interaction is found, but tendencies towards interaction are noted. It is concluded that LISREL has several advantages in the analysis of ATI studies, but also that the power of the test of ATI effects is low, particularly when there is a high correlation between the latent aptitude variables. (Author/GDC)

Describing and testing aptitude-treatment interaction effects with structural equation models: Reanalysis of a study by M.J. Behr.

Jan-Eric Gustafsson
Berner Lindström
University of Göteborg

The research presented in this report has been supported financially by the Swedish Council for Research in the Humanistic and Social Sciences and by the National Board of Education in Sweden. We want to express our gratitude to Dag Sörbom and Bengt Muthén for giving much instruction and advice in the use of LISREL; we assume, however, complete responsibility for any misuse.

ABSTRACT

The Jöreskog and Sörbom LISREL method is investigated as an alternative to regression analysis in studies of aptitude-treatment interactions (ATI), to solve problems caused by unreliability of measurements and by large sets of variables. A study reported by M.J. Behr is reanalyzed. The study investigated relations between verbal and figural aptitude variables and outcome variables within one verbal and one figural treatment teaching modulus seven arithmetic. The relations between latent aptitude variables and latent outcome variables are studied in 4 LISREL models with either 1 or 2 latent aptitude variables, interpretable as a general factor and verbal and figural ability, respectively, and with 2 sets of latent outcome variables, interpretable as learning/retention and computational speed/understanding. No significant interaction is found, but tendencies towards interaction are noted. It is concluded that LISREL has several advantages in the analysis of ATI studies, but also that the power of the test of ATI effects is low, particularly when there is a high correlation between the latent aptitude variables.

INTRODUCTION

In research on aptitude-treatment interactions (ATI) the interest is focussed on joint effects of instructional treatment and individual differences. The standard procedure for analyzing data from ATI studies is to regress outcome variables one at a time on one or more aptitude variables within treatment groups and to test for homogeneity of the within treatment regression slopes (Cronbach & Snow, 1977, ch. 3).

However, the ordinary regression analysis of ATI data is not free from problems. The regression on an observed aptitude variable is different from the regression on the true aptitude variable when it is not perfectly reliable; errors of measurement in the aptitude variables thus bias the tests and descriptions of ATI effects. Cronbach and Snow (1977) stated that "Ideally, every ATI study would examine the regression of outcome onto the true aptitude score... This... regression will have a different slope than the observed-score regression, and under some circumstances the apparent interaction may be radically altered. Important as this matter is, it has been ignored in ATI research to date..." (p.34).

The problem is severe enough when only one aptitude variable is considered but it is aggravated in multiple regression (Cronbach & Snow, 1977, p. 36). Furthermore, an additional complication arises in the regression based approach to the analysis of ATI data when there are many variables. Then a great many regression coefficients are estimated and tested, which makes for chance significancies and tends to give rise to complex patterns of results which are hard to interpret.

In this paper an alternative approach to the analysis of ATI data will be illustrated -- the LISREL method of

Jöreskog and Sörbom (see Jöreskog, 1970, 1973, 1974; Jöreskog & Sörbom, 1976, 1977, 1978). LISREL (LInear Structural RELations) is a method, and a computer program, for analysis of linear structural relations between variables, which may be either observed or latent. When latent variables are studied a factorial structure is imposed on the observed variables, which structure serves to identify the latent variables and allows estimation of the error variances in the observed variables. In this way several observed variables may be reduced to fewer latent variables and relations between true variables rather than observed variables can be studied.

It cannot be taken for granted, however, that LISREL only has advantages when applied to analysis of ATI studies: The method is built on strong assumptions concerning the nature of the data; not only consistency of the estimates is of importance but also their variance, just to mention two possible sources of problems. There seems thus to be a need for empirical studies of the applicability of this alternative method in the analysis of ATI data.

One purpose of the present study is to make such an appraisal of the problems and virtues of LISREL. The study is a reanalysis of a study presented by Behr (1967), which was chosen because it included several aptitude and outcome variables and because the present authors are familiar with the substantive problem studied by Behr.

Behr investigated the hypothesis that tests of verbal ability are more highly correlated with achievement in a verbal treatment than in a figural treatment and that tests of figural ability are more highly correlated with achievement in a figural treatment than in a verbal treatment (cf. Gustafsson, 1976). This was one of the first ATI hypotheses to be suggested and several of the early ATI studies, of which the Behr study is one, investigated this hypothesis.

Behr used 14 aptitude variabels; about half of them were verbal and half of them were figural. One group of subjects studied a verbal-symbolic (VS) programmed teaching material and another group of subjects studied a figural-symbolic (FS) teaching material. The study included 7 outcome variables (for a more detailed description of the Behr study, see below).

Behr investigated interactions by comparing the within-treatment regressions of each of the dependent variables on each of the aptitude variables. About a dozen significant interactions were found, and almost all interactions were due to a steeper regression on verbal tests in the VS treatment than in the FS treatment.

In their review of ATI research Cronbach and Snow (1977, p. 286) made a simple reanalysis using figures presented by Behr. For one outcome they summed the regression coefficients of all the figural tests separately for the two treatments. This procedure was repeated for the verbal tests. The sum for both groups of tests was higher in the VS treatment, even though the difference was smaller for the figural tests. Cronbach and Snow drew the conclusion that a general factor was more associated with achievement in the VS than in the FS treatment.

Gustafsson (1974, pp, 15-16; cf. Gustafsson, 1976, p.56) presented another simple reanalysis of the Behr study. The most reliable tests were selected and one verbal and one figural test at a time were entered into separate multiple regression equations for the two treatments. It was found that one figural test that had a higher zero-order regression coefficient in the FS treatment, in the multiple analysis came out with a much higher partial regression coefficient in the FS treatment than in the VS treatment. This result was interpreted as a very weak tendency towards interaction with figural ability and since the interaction with verbal ability seemed relatively well established it was concluded that the Behr study gave some support to the verbal/figural ATI hypothesis. Three different conclusions have thus been drawn from the Behr study: Behr himself stressed the results obtained with each particular aptitude variable; Cronbach and Snow interpreted the effects found as being accountable for by general ability; and Gustafsson, finally, saw some value in the distinction between verbal and figural abilities. In addition to the methodological purpose of studying the applicability of LISREL, another purpose will be to see which, if any, of these conclusions receives support.

METHOD

The Behr study

The subject matter taught in the Behr study was modulus seven arithmetic. In both the VS and FS treatments algebraic symbols were used but in the VS treatment this information was supplemented with verbal information. In the FS treatment figural information was added to the symbolic presentation. The figural material was of low complexity, consisting for example of a circle with which the arithmetic operations were illustrated as movements along the circle.

The aptitude varibles were selected to correspond to cells in the Guilford (1967) "Structure-of-Intellect (SI)" model. Of the tests, 6 had a figural content (-F-), 5 had a semantic (verbal) content (-M-) and 2 had a symbolic (numeric) (-S-) content. In addition there was a test, called Integration, which was not classifiable in the SI structure. This test gave verbally formulated instructions concerning directions. It can be hypothesized that this test measures figural ability; it comes quite close to the task which Brooks (1967) in experimental studies has shown to be of a spatial/figural kind. Two kinds of operations, cognition (C--) and memory (M--), were represented among the aptitude variables. Most of the product categories were also represented among one or more of the tests: units (--U), classes (--C), relations (--R), transformations (--T) and implications (--I). Table 1 displays the tests used, their coding in the SI system, the number of items, and the time limits.

Insert Table 1 about here

It can be observed that many of the tests are quite short, with a very limited testing time, which is partly due to the fact that Behr in most cases used shortened versions of the tests.

Three criterion measures were determined: Time used to study the program (TP), a Learning Test (LT) score and a Retention Test (RT) score. The LT was administered two days after the instruction and the RT two weeks after the instruction.

The LT was written in five parts, Part I was a speed test of modulus seven addition and subtraction. Parts II, III, IV and V all dealt with structural properties of the modulus seven system, but contained different kinds of items. Two sub-test scores were derived from the LT: one consisting of the score on Part I (LA) and one consisting of the sums of scores on parts II-V (LB). Behr used the two sub-test scores as dependent variables, along with the total LT score.

The RT was a parallel form of LT and was written with other modulus seven numbers in the items or with changes in the order of the questions or responses. Also from the RT two sub-test scores (RA and RB), corresponding to those in the LT, were derived and used as dependent variables along with the total RT score.

Subjects in the study were prospective elementary school teachers, there being 120 and 108 subjects in the VS and

FS groups, respectively. The subjects were randomly assigned to treatments.

Through assembling information presented by Behr in tables and appendices it was possible to construct the within treatment covariance matrices. It must be mentioned, however, that one of the correlations presented by Behr was changed. For the FS group the correlation between CFR and MMU was reported to be .71. This correlation is by far the highest correlation for any two aptitude variables and is probably due to a typing error; it was therefore replaced with the value .17.

LISREL

The LISREL approach, and related ones, has been described in several publications (e.g. Jöreskog, 1970, 1973, 1974; Jöreskog & Sörbom, 1976, 1977, 1978). Here only a very sketchy description can be afforded.

The LISREL model consists of two parts: the measurement models for the dependent and independent variables, in which latent variables (common factors) are defined in terms of the observed variables, and the linear structural equation model, in which the relations between the latent variables are specified.

There are two sets of observed variables $\underset{\sim}{y}' = (y_1, y_2, \ldots, y_p)$ and $\underset{\sim}{x}' = (x_1, x_2, \ldots, x_q)$, corresponding to dependent and independent observed variables respectively, and two sets of latent variables $\underset{\sim}{\eta}' = (\eta_1, \eta_2, \ldots, \eta_m)$ and $\underset{\sim}{\xi}' = (\xi_1, \xi_2, \ldots, \xi_n)$, corresponding to dependent and independent latent variables respectively. There are also vectors specifying the unique parts of the y and x variables, $\underset{\sim}{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)$ and $\underset{\sim}{\delta}' = (\delta_1, \delta_2, \ldots, \delta_q)$ and a vector specifying the residuals in the structural equations system: $\underset{\sim}{\zeta}' = (\zeta_1, \zeta_2, \ldots, \zeta_m)$.

In LISREL the relations between the latent variables and the observed variables, and the relations between the latent variables are specified in up to eight parameter matrices:

Λ_x is a factor loading matrix of order q x n, for the regression of the x variables on the ξ variables.
Λ_y is a factor loading matrix of order p x m, for the regression of the y variables on the η variables.
Θ_δ is a diagonal or symmetric matrix of order q x q containg the covariance matrix for the unique parts of the x variables.
Θ_ε is a diagonal or symmetric matrix of order p x p containg the covariance matrix for the unique parts of the y variables.
Φ is a diagonal or symmetric matrix of order n x n containg the covariance matrix of the ξ variables.
Ψ is a diagonal or symmetric matrix of order m x m for the residuals (disturbance terms or errors in equations).
Γ is a coefficient matrix or order m x n for the structural relations between the ξ and the η variables.
β is a coefficient matrix of order m x m for the structural relations among the η variables.

The measurement models for the x variables is written:

$$x = \Lambda_x \xi + \delta,$$

and for the y variables it is written:

$$y = \Lambda_y \eta + \varepsilon.$$

The system of linear structural equations has the form:

$$\beta \eta = \Gamma \xi + \zeta.$$

Specifying a LISREL model involves specification of the nature of each element in the parameter matrices (the elements will be referred to with small Greek letters).

The elements can be of three kinds: a fixed parameter, i.e. the parameter is assigned a given value; a free parameter, i.e. the parameter is to be estimated; and a constrained parameter, i.e. the parameter is to be estimated but it is constrained to be equal to one or more other parameters.

If the model is identified, i,e, if there is a unique estimate of each non-fixed parameter, the parameters can be estimated with maximum likelihood methods from the sample covariance matrices, using the Jöreskog and Sörbom (1978) LISREL4 program, for example. Each analysis of a fully identified model not only yields estimates of parameters but also an overall χ^2 test of the goodness of fit to the model, along with standard errors of the estimated parameters. Through computing the differences between the values of the test statistics obtained with more and less constrained models, i.e. models differing as to the number of parameters estimated, it is also possible to test the signigicance of subsets of parameters.

So far the presentation of the LISREL model has only dealt with the case when there is one group only. However, the LISREL4 program handles several groups as well, and through constraining parameters to be equal in several groups it is possible to test the equality between groups, either of all the estimated parameters or of subsets of the parameters (cf. Sörbom, 1978).

In ATI applications of LISREL the main interest is of course centered on the $\underset{\sim}{\Gamma}$ coefficients, which roughly correspond to the within-treatment regression coefficients and it is a rather straightforward process to estimate these within the treatment groups and to test them for equality between groups.

RESULTS

It was suggested by Jöreskog and Sörbom (1977) that it may be a good strategy to construct the full LISREL model in steps, starting with the measurement models and then only at a later step fit these together through the linear structural equations. This is the strategy followed here, and we will start with the measurement model for the aptitude variables.

Measurement models for the aptitude variables

Each of the cells in the Guilford SI-system supposedly defines a unique factor and since in the Behr study only one test was sampled from each cell it could, from the SI point of view, be argued that no attempts should be made to account for these with a smaller set of common factors. However, apart from the fact that the inter-correlations among the factors defined in the SI-system have not been much studied, it could be argued that the SI-system is so elaborate as to be impractical. Cronbach and Snow said:

> "if as many abilities as Guilford recognizes must be recognized, hypotheses about ATI will have to be finely differentiated and very large samples will be needed to establish weightings for separate abilities. The prospects for sucessful ATI research would be much enhanced if it were decided that a system simpler than Guilford's accounts for the ability differences of long-run practical importance." (Cronbach & Snow, 1977, p. 155, emphasis in original).

Cronbach and Snow (1977, pp. 155-160) made a partial assessment of the SI system and presented some reanalyses of correlation matrices. They concluded, among other things, that tests with a similar "content" tend to inter-correlate but that tests calling for the same kind of "product" are not functionally similar.

There are thus strong reasons to investigate the possibilities to reduce the many observed aptitude variables in the Behr study to a smaller set of common factors.

When LISREL is applied to develop the measurement models a sequence of confirmatory factor analyses is carried out. To get some information about the dimensionality of the latent space one ordinary exploratory factor analysis was first performed within each treatment group. These analyses indicated, in the first place, that the two -S- tests had very little in common with the other tests, and they were therefore excluded. The sequence of eigenvalues indicated, secondly, that there in the groups were 3 or 4 common factors. There were, however, some differences in the size and pattern of loadings for the two groups.

It seemed a good strategy to start with few factors in the confirmatory analyses. and then add more factors if necessary. A 1-factor model was fitted, and also a 2-factor model, with one factor defined by all the -M- tests, and the other factor defined by the -F- tests. For each of these cases two solutions were made; one in which the parameters were specified to be the same in the two treatment groups and one in which no constraint of equality was imposed. The results from the tests of goodness of fit of these four solutions have been entered in Table 2, along with the appropriate differences to test equality of the solutions in the groups and to test whether the 2-factor solution is significantly better than the 1-factor solution.

Insert Table 2 about here

No significant difference is found between the solutions for the treatment groups. The tests of the differences between the 1- and 2-factor solutions indicate a significantly better fit for the 2-factor solution. Furthermore, the overall evaluation of fit of the 2-factor solution gives a value of the test statistic which is slightly lower than the critical value. Thus, even though the fit of the verbal/figural solution is far from perfect it is for the present purposes considered acceptable.

Had higher standards of fit been set it would of course have been natural to [illegible] 3-factor, and perhaps a 4-factor model. However, even though it will not be illustrated here we would like to mention the availability of another strategy to obtain fit without adding further factors. Often the specific parts of some tests correlate because the tests share certain characteristics, such as being administrered at a common occasion, or having similar content, or requiring the same type of answers and so on, In contrast with ordinary factor analysis the correlations between the errors of measurement need in LISREL not be assumed to be zero, but can be set as free parameters to be estimated (cf. Sörbom, 1975), which may at times be a better strategy than to introduce additional factors.

In Table 3 the standardized factor loadings estimated within each treatment group are presented. The factors will be labeled the figural and the verbal factor respectively.

Insert Table 3 about here

It is, however, difficult to characterize the nature of the factors more closely on the basis of the size of the loadings, since the differences between the groups tend to be almost as large as the differences between the tests. There is a tendency for the figural factor to display higher loadings in the VS group than in the FS group. Tests of the equality of each loading show that among the figural tests there is one significant difference between the treatment groups;

for MFU a t-value of 2.63 is obtained. The differences are smaller and less systematic for the verbal factor, but there is a significant difference for the MMR test (t= 1.97).

It can be observed that for many of the tests the communality is very low. One reason for this is that the reliability of the tests must be low since they are often very short. Another reason might be that each test brings much specificity; this is of course to be expected on the basis of SI-theory. With LISREL, however, it is not in a model like this one possible to obtain separate estimates of the specificity and the errors of measurement as such.

The factors are very highly intercorrelated; in the VS group the correlation is .81 and in the FS group it is .80. Even though these correlations are significantly different from unity they are so high that there is very little information in the verbal/figural ability distinction.

Two reasons can be cited why in this case the correlation between the factors is so high. It will be recalled that the sample was drawn from a population of prospective elementary school teachers; thus it is likely to have been all, or almost all, female and it has repeatedly been found that the figural/spatial factors are weaker among females than among males (e.g.Werdelin, 1961). Secondly, it has been shown that certain figural tests are quite amenable to solution with verbal/reasoning kinds of processing (cf. Gustafsson, 1976, ch 2). Most of the figural tests used by Behr are likely to be of that type; the tests on which verbal/reasoning strategies are less successful seem to be those with less complex stimulus configurations and which place higher demands on speed (such tests are classified as CFS in the SI system).

Furthermore it appears that the choice of tests and the choice of subjects interact to produce a high correlation between the factors, since females in particular appear to resort to verbal/reasoning strategies whenever possible (Gustafsson, 1976, ch. 6).

In spite of the high correlation between the factors the 2-factor model will be used as a measurement model for the aptitude variables. However, the 1-factor model will be used as well, to contrast interpretations in terms of general ability with interpreations in terms of the verbal/figural distinction.

Measurement models for the dependent variables

In the measurement models for the dependent variables only the sub-test scores (LA, LB, RA, and RB) will be used; the total scores can of course not be included since they are linearly dependent on the sub-test scores, and the TP measure cannot alone define a "time-factor."

With 4 dependent variables it is possible to define a 2-factor model and still have 1 degree of freedom left to test goodness of fit. It should be possible, however, to define two potentially meaningful 2-factor solutions: either one learning factor defined by LA and LB and one retention factor defined by RA and RB; or one computational speed factor defined by LA and RA and one factor defined as understanding of the structural properties of the system, measured by LB and RB. Both these possibilities will be investigated.

Basing the analysis on the correlation matrices the learning/retention factors were first postulated. In both groups an exceedingly poor fit was obtained (VS: χ^2 = 49.9, df = 1; FS: χ^2 = 29.0). Testing the equality of the estimated parameters for the two groups, no significant difference was found (χ^2 = 6.0, df = 9).

The poor fit within both treatment groups may in this case be due to correlated errors of measurement; the common content of the LA and RA tests on the one hand and of the LB and RB tests on the other can be suspected to cause $\theta_{\varepsilon_{LA,RA}}$ and $\theta_{\varepsilon_{LB,RB}}$ to be larger than zero. Since the model has only 1 degree of freedom it is not possible to estimate these two correlated errors here; when ξ variables are added this is, however, possible (Jöreskog & Sörbom, 1978, pp. 22-28).

It is nevertheless possible to obtain an indirect appraisal of the effects of the correlated errors of measurement. A maximum likelihood estimate of the correlation between η_L and η_R can easily be obtained (Jöreskog & Sörbom, 1977, p. 293). If this estimate is lower than the estimate obtained when the assumption is made that the errors of measurement are uncorrelated, this can only be because the correlated errors contribute to the latter estimate. Allowing for possibly correlated errors the estimates of the correlation between η_L and η_R were .79 and .91 in the VS and FS groups, respectively; assuming uncorrelated errors the corresponding estimates were .99 and 1.08. Thus, when uncorrelated errors of measurement are assumed the learning and retention factors collapse, but when the effects of common test content are partialled out the factors appear to be distinct, at least within the VS group.

Within the measurement model only it is not possible to make proper tests of whether one or both of the sub-tests contribute with correlated errors. Within the full LISREL model, where the aptitude factors have been added it is, however, possible to obtain t-tests of the estimates of these parameters. Bringing in these results already here it is found that the t-value for the estimate of the covariance between the errors in LA and RA is 5.06 and 2.92 in the VS and FS groups, respectively; for LB and RB the corresponding figures are 1.65 and 1.05. Thus, the correlated errors are higher for computational speed, and particularly so in the VS group.

Behr found a significantly higher variance for RA and RB in the VS treatment (Behr, 1967: pp. 42-43). It can thus be expected that differences are found between the treatment groups when the covariance matrices are analyzed instead of the correlation matrices. This is indeed the case, there being a highly significant value for the test statistic ($\chi^2 = 31.9$, df = 9). It would of course also be possible to apply LISREL to test subsets of the parameters to get information about which variances differ between the treatments; this is hardly necessary, though, since the conclusions from such analyses are not likely to differ from the conclusions drawn by Behr.

The correlated errors of measurement found above are clear indications that LA and RA form one factor and that LB and RB form another factor. Testing the hypothesis of this 2-factor structure among the dependent variables a somewhat better fit than when the learning/retention factors were investigated is obtained in both groups (VS: $\chi^2 = 21.0$, df = 1; FS: $\chi^2 = 4.1$, df = 1, $p < .04$).

The fit is obviously better in the FS group than in the VS group; still a test of the equality of the estimated parameters (from the correlation matrices) is far from significant ($\chi^2 = 4.6$, df = 9). Since it was shown above that the learning/retention factors are identifiable, at least in the VS group, the reason for the poor fit of the speed/understanding factor structure must be correlated errors of measurement.

Allowing for correlated errors of measurement the estimates of the correlation between the factors were .65 and .72 in the VS and FS groups, respectively. Assuming the errors to be uncorrelated the estimates were .70 and .75. These figures indicate that the speed/understanding factors are stronger than are the learning/retention factors, and that the differentiation between the factors is more clear in the VS than in the FS group. Using results from the full LISREL model the t-value for the covariance between the errors in LA and LB is found to be 1.44 and 1.39 in the VS and FS groups, respectively; for RA and RB the corresponding figures are 3.15 and .20. Again it is found that the correlated errors are more important in the VS group.

The analyses of the dependent variables thus have shown that it is possible to define two measurement models; one with the speed/understanding factors, with the effects of the two occasions of measurement taken into account; and one with the learning/retention factors, with the effects of the common content in the test taken into account. Both these measurement models will be used in the analysis of the structural relations within treatments.

Structural relations within treatments

Having developed the measurement models for the aptitude variables and for the outcome variables it is now time to fit these together into the full LISREL model and to study the structural relations within treatments.

It was decided to use two different measurement models for the aptitude variables, one 1-factor model only representing a general factor and one 2-factor model with the verbal/figural factors. Also for the dependent variables two measurement models were defined, one containg the learning/retention factors and one containing the speed/understanding factors. Combining these measurement models we get 4 full LISREL models.

In Figure 1 the LISREL model for the verbal/figural and speed/understanding factors is shown, using the following symbols: Latent variables are enclosed in circles, observed variables are enclosed in squares, and errors of measurement and disturbance terms are included without being enclosed. A straight one-way arrow indicates a causal influence of one variable on another and curved two-way arrows indicate correlation between variables without any causal interpretation (cf. Jöreskog & Sörbom, 1978).

Insert Figure 1 about here

Only one figure is shown but since there are two treatment groups it must be imagined that there is one graph for each treatment. The models for the two groups can have more or less in common, however. One possibility is of course to estimate all the parameters within the groups without constraining any of them to be equal. However, in this case it is reasonable that the same measurement model for the aptitude variables (SM) is used in both groups. Should the results from such an analysis differ from the results when different measurement models (DM) are used this can only be because differences between the treatment groups with respect to the aptitude variables spuriously affect the relations between aptitude and outcome within treatments. Results obtained under both the DM and SM models will be presented.

The unstandardized coefficients of the relations between latent aptitude variables and the speed/understanding outcome variables are presented in Table 4, for both the 1-factor and the 2-factor aptitude models. Along with the coefficients, t-values for pairwise tests of their similarity in the treatments have been entered, while results from overall tests of fit are presented in Table 5.

Insert Tables 4 and 5 about here

None of the tests of interaction even approaches significance. However, a partial explanation of this is found if the standard errors of the estimates are considered. These are particularly in the verbal/figural model so high that only few of the within-treatment relationships are significant. Since the coefficients presented are the unstandardized ones their absolute level cannot be judged from these figures alone. However, the standardized coefficients, interpretable as partial correlations, were quite high, with the values often being around .50.

When the general aptitude variable is included in the model the standard errors generally are smaller and the within-treatment coefficients are all highly significant. This indicates that the standard errors are so large when the verbal/figural variables are used because of the high correlation between the latent aptitude variables, and from the formula for the standard errors of estimates of the γ coefficients it is clear that these are a quadratic function of the correlation.

Looking at the descriptive pattern of results for the verbal/figural aptitude variables it can be observed that with respect to understanding the figural aptitude variable has a higher coefficient in the VS group than in the FS group and that the verbal aptitude variable has a slightly higher coefficient in FS; with respect to speed the pattern of differences is completely reversed. It can also be observed that there are some differences in the descriptive patterns of results under the DM and SM models: With respect to understanding the SM model yields smaller differences between the treatments than does the DM model, while the opposite is the case for speed.

For the general aptitude variable there is for both the outcome variables a higher coefficient in the VS than in the FS group, but it must also be observed that there especially with respect to understanding is a smaller difference between the treatments under the SM model than under the DM model.

No graph is shown for the model with the learning/retention outcome variables. This LISREL model is quite similar to the one for speed/understanding, the only differences being that there is a unidirectional influence of learning on retention and that LA and RA now have correlated errors of measurement, as have LB and RB.

In Tables 6 and 7 the results obtained with the learning/ retention factors are shown. The tests of interaction yield even lower values of the test statistics than was

Insert Tables 6 and 7 about here

the case for the other outcome variables, and descriptively there are for the verbal/figural variables only very small differences between the coefficients for the treatments.

For the general aptitude variable there is with respect to both the outcome variables a somewhat higher coefficient in the VS group than in the FS group, but again the difference is smaller under the SM model than under the DM model.

To summarize, no significant interaction is found in the LISREL analyses even though there descriptively are some differences between the treatments: For computational speed a pattern of differences between treatments conforming to that specified in the original verbal/figural ATI hypothesis is found, but for understanding the reverse pattern of differences is found.

With respect to the learning/rentention factors there are even descriptively very few signs of interactions. It can be noted, however, that treatment obviously in some ways had differential effects on outcome: The variance in the variables measuring retention was higher in the VS treatment than in the FS treatment and there is a lower correlation between learning and retention in the VS group than in the FS group. These differential effects may of course be related to aptitude variables, but if that is so it must be other variables than the verbal/figural ones.

DISCUSSION AND CONCLUSIONS

Before discussing the methodological and substantive conclusions to be drawn from this reanalysis it should be pointed out that a replication of the Behr (1967) study has been presented by Behr and Eastman (1975). In the replication a sample of about the same size was drawn from the same population as in the original study. Some changes were made of the instructional materials in that the figural treatment was made more inductive and the verbal treatment more deductive. In the replication only two outcome measures were used, a retention and a transfer test. Both these tests were administrered one week after the learning session. Seven aptitude variables were used; among those were the tests that had shown interactions in the original study.

In the Behr and Eastman (1975) analysis several different methods were used in the study of ATIs:Correlation coefficients were compared, and a large number of multiple-regression analyses were performed. The dependent variables also were regressed on two Varimax factors. In no analysis any significant interaction was found and on the basis of these negative results Behr and Eastman warned against accepting the results from the original study.

However, the replication did not assess computational speed and understanding as separate outcomes, and the reanalysis presented above indicates both that different patters of results are obtained with respect to these outcome variables and that the strongest effects are found with respect to computational speed. Therefore it does not seem that the replication has much to tell about the results obtained in the reanalysis.

As was mentioned earlier there exist at least three analyses and interpretations of the results in the Behr (1967) study. These will now be examined in the light of the results obtained using the LISREL approach.

Behr himself stressed, in the Guilford (1967) tradition, the uniqueness of each single aptitude variable and refrained from relating these to each other. In contrast with the LISREL analysis Behr did find some significant interactions -- some more actually than can be expected from chance at the significance level chosen, if independence is assumed. The tests entering interactions with one or more of the post-tests were CMU, MMR and CMC, which is a thoroughly mixed set, and it is hard to find any reason why these verbal tests and not the others should enter into interactions.

Using large sets of aptitude variables the ATI researcher is exposed to very great risks of false postive and false negative conclusions. Thus with such an approach it is even in the long run almost impossible to sort out the dependable findings from the undependable ones and to know whether it is the unique or the common parts of the variables that enters into interactions.

There is of course a risk that important interactions are missed if the specific parts of the aptitude variables are left out. However, if it cannot be shown on theoretical grounds that it is the specific parts of the aptitude variables which are likely to enter into interactions with the treatment variables there are strong reasons to reduce large sets of aptitude variables to a smaller set of latent variables.

In order not to create any misunderstanding, we should point out that this recommendation of course only applies when it is reasonable to impose a factorial structure on the aptitude variables. If the tests all measure different factors such an approach is not reasonable. However, even in such cases LISREL can be used to study relations between latent variables, if each test is entered as two half-tests or if previously obtained estimates of the reliability is entered into the model.

Cronbach and Snow concluded, as described in the Introduction, that a general factor of ability was more associated with achievement in the VS treatment than in the FS treatment. The LISREL results show that the coefficient for the structural relations between the general aptitude factor and the outcome variables tends to be considerably higher in VS than in FS. However, a part of the interaction found by Cronbach and Snow can be accounted for by differences between the treatment groups with respect to the aptitude variables, since when the same measurement model is used for both treatments the difference between the coefficients is lower. With respect to computational speed there is even under the SM model a considerably higher coefficient in the VS group, but when the verbal/figural aptitude variables are analyzed instead this difference is found to be wholly accounted for by the verbal ability variable. Thus, the LISREL results give very little support to the Cronbach and Snow conclusion.

Gustafsson (1974) concluded that the interaction with verbal ability seemed well established in Behrs study and suggested also that there may be a very weak interaction with figural ability. Gustafsson (1976) also pointed out that the tests of verbal ability were not highly correlated with achievement in the VS treatment, but that they had no or a negative relationship with achievement in the FS treatment (cf. Behr & Eastman, 1975, p. 156). It was concluded that this indicated that "pupils with a high verbal ability did poorly in the FS treatment rather than well in the VS treatment" (Gustafsson, 1976, p. 56).

The results for the understanding variable certainly do not give any support to such a conclusion but the results with respect to computational speed do to some extent: there is a negative coefficient for the relation between verbal ability and computational speed within the FS treatment, while at the same time there is a weak positive relationship within the VS treatment.

To the best of our knowledge computational speed and similar types of outcome have not been much studied in ATI research studying the verbal/figural aptitude variables, in spite of the fact that many studies have used mathematics as subject matter. It is thus difficult to compare this result in the Behr study with results in other studies. However, the lack of interaction with respect to understanding is in agreement with the results in other studies, including the Behr and Eastman (1975) replication and studies by Bracht (1969), Carry (1967), Webb and Carry (1975), Hancock (1975), and Gustafsson (1976).

No attempt will here be made to interpret the tendency towards interaction with respect to computational speed since it is so utterly weak. However, should further research provide supporting evidence, it seems reasonable that the interpretation should be couched in what Gustafsson (1976, p. 81) labelled interferential terms, i.e. that a treatment has negative effects on the learning and processing of subjects high in an ability.

So far we have mainly discussed the results from a substantive point of view; now we will turn to a discussion about the methodological aspects. Before scrutinizing the advantages and disadvantages of LISREL as a method for analyzing ATI data, it must, however, be pointed out that suggestions for methods of analysis purporting to solve some of the problems which can be approached with LISREL have been offered earlier.
Cronbach and Snow (1977), p. 39) argued that when there are several aptitude variables a reduced rank analysis may be performed through regressing outcome on aptitudes formed as composites, either on the basis of factor analysis or on the basis of judgement. And in the context of an illustrative analysis of a study including measurement of achievement at several points in time Cronbach and Snow (1977) claimed that methods such as path analysis and structural equation models "are becoming increasingly significant in educational research and that they may be peculiarly valuable in studies of learning and ATI" (p.94).

Some of the problems which may be solved with LISREL can certainly be solved with other methods. In order to reduce several observed variables to few latent variables component or factor analysis can be used; to study relationships between true variables rather than observed variables correction for attenuation may be employed; and to study relations between measurements with an intrinsic causal ordering path analysis can be used. But with LISREL it is possible to specify models including all these features, which results in a more parsimonious and often more efficient analysis. Furthermore, LISREL offers capabilities not to be found in any other method of analysis, such as the possibility of treating correlated errors of measurement, and to test goodness of fit.

In the context of ATI research the LISREL approach also has some more specific advantages. Even though it is seldom investigated, it is surprisingly often found that there are differences between the treatment groups with respect to the level and structure of the aptitude variables (cf. Cronbach & Webb, 1975; Cronbach & Snow, 1977, p.38; Gustafsson, 1976, 1977, 1978) and such differences often result in spurious ATIs (and can for that matter also be suspected to conceal ATIs at times). The necessity of formulating explicit measurement models for the aptitude variables in LISREL makes it, however, natural to investigate the similarity of the structure of the aptitude variables in the treatment groups. LISREL also allows investigation into the effects of the differences between the treatment groups on the structural relations within treatments through comparing the results when the same measurement model is used and when different measurement models are used, even though great caution is always necessary when large and systematic differences are found between the structure of the aptitude variables for the treatment groups.

So far we have only dealt with the possible advantages of LISREL as a method for analyzing ATI data, but there are disadvantages as well. For one thing the method is built on rather strong assumptions: multivariate normality is assumed. The same assumption is found in regression analysis, but it is probably fair to say that any advantage that LISREL may have is reduced when this assumption is not fulfilled. It must also be pointed out that the goodness of fit test is a large sample test. Unfortunately very little is known about at what sample sizes the test sufficiently well approximates its asymptotic properties, but obviously a warning must be made against using LISREL for very small samples.

The greatest problem, however, in applying LISREL to analyze ATI data is that the standard errors of the estimates of the structural relations within treatments are large. One reason why these standard errors are so large is that there tends to be an inverse relationship between the degree of consistency and the variance of an estimator, so there is a price to be paid to obtain the unbiased estimates. (Parenthetically, it can be pointed out that even within regression analysis the inverse relationship between consistency and variance has been studied, see Winer, 1978). Another reason why the standard errors are so extremely large in this particular case is that the correlation between the latent aptitude variables is so high.

Even when regression analysis is used the power of the tests of interaction mostly is too low with the sample sizes which are feasible in ATI research (Cronbach & Snow, 1977). If LISREL is to be used instead of regression analysis this problem is aggravated to become an obstacle against ever finding a significant interaction, at least within areas of study where there tend to be high correlations between the aptitude variables.

What can be done to reduce the large standard errors of the estimates? One solution is to increase sample size. However, this solution may be simple in theory but need not be so in practice. Since ATI research is often experimental the data are cumbersome and costly to collect so it may simply be impossible to obtain the large samples needed.

A more feasible strategy is to select the sample, the tests and the model so as to reduce the correlation between the latent variables. Reasons were stated above why with this sample and with these tests a large correlation can be expected with the verbal and the figural aptitude variables, and in retrospect it can be concluded that neither were well chosen to study the verbal/figural ATI hypothesis. However, with the importance of having a low correlation between the latent variables in mind the researcher can often chose tests and subjects so as to minimize it.

It is also possible to specify LISREL models which reduce the correlation between the aptitude factors. One possibility is to allow correlated errors of measurement instead of invoking additional factors. Another way to lower the correlations between the latent variables is to allow some of the observed variables to load in more than one factor.

Using the six tests in Behr's study with the highest communality it was in fact possible to define an orthogonal 2-factor model using such methods. However, the results from application of this measurement model have not been presented in full for two reasons. In the first place the factors were quite difficult to interpret; some of the verbal tests had, for example, their highest loadings in the figural factor. Secondly, when this measurement model was used together with observed dependent variables (the measurement models used here for the dependent variables were not identified with this measurement model for the aptitude variables) the verbal factor disappeared in the sense that it had a non-significant variance.

These experiences indicate that in this case the information in the observations about the verbal/figural aptitude variable distinction is too limited, whichever model is chosen to describe the data.

A more drastic solution is of course to specify an orthogonal model without bothering about the resulting poor fit when the variables are in fact correlated. To study the effects of chosing an orthogonal model instead of the proper oblique model the computational speed/understanding outcome variables were analyzed together with the verbal/figural aptitude variables, defined by the 12 tests but with a zero covariance specified.

For computational speed the difference between the within-treatment coefficients for the VS and FS groups changed from 1.25 to .98 for verbal ability and from -.12 to .01 for figural ability. Thus, specification of the wrong model did not affect the estimates too seriously. However, the standard errors were reduced with a factor of about 2.5 so in the orthogonal model there was in fact a significant difference between the within-treatment coefficients for the relation between computational speed and verbal ability ($t = 2.08$).

In spite of the fact that it is even possible to find a significant interaction in this case we certainly cannot recommend that poor-fitting orthogonal models are used. The estimates of the parameters need not be too wrong, but the standard errors will be seriously underestimated, so the tests of interaction cannot be trusted.

As was pointed out previously the problem of the low power of the statistical tests in ATI research is certainly not unique to LISREL; it is just more pronounced in that method of analysis than it is in regression analysis. Cronbach and Snow (1977) concluded that the low power of the test of interaction in regression analysis makes it necessary to place lower weight on formal statistical tests and instead consider the descriptive results. Cronbach (1975) even claimed that: "The time has come to exorcise the null hypothesis. We cannot afford to pour costly data down the drain whenever effects present in the sample 'fail to reach significance'."(p. 124).

We strongly agree that less emphasis should be placed on statistical inference, and that greater importance should be attached to description of the effects in the sample (cf. Gustafsson, 1976).

LISREL does offer some advantages in a research strategy based on description. For one thing, the latent aptitude variables can be supposed to be more or less invariant over different studies, which is important when the results from different studies are brought together. Furthermore, the description is based on consistent estimates of the parameters and it is generally very parsimonious.

However, there are disadvantages as well. As was mentioned above the specific parts of the variables are not studied which of course is serious when the possibility of local effects is entertained. It can also be claimed that the LISREL approach, which involves estimation of parameters in a model, tends to be quite remote from the observations themselves. But more important is the fact that LISREL in its present version (LISREL 4) does not allow hypotheses on means; consequently it is notpossible to investigate the ordinality/disordinality of an interaction, let alone to describe the interaction effects as they appear in the sample. It should be pointed out, however, that LISREL can be complemented with analyses using the COFAMM program (Jöreskog & Sörbom,1976b) which does allow estimation of the intercept as long as there are no constraints on the structural relation coefficients.

But even so the description is based on rather constrained models, involving strong assumptions about the nature of the interaction effects and it can be claimed that the description of interaction effects should be as unconstrained as possible (cf. Gustafsson, 1976; Gustafsson & Svensson, 1978). Thus it can be argued that even as simple methods as descriptions of cell means for levels of ability within treatments for descriptive purposes do have some advantages over more constrained statistical models, either these are regression models or structural equation models.

Therefore we think that LISREL analyses of ATI data often ought to be complemented with close descriptions of effects in the sample using as unconstrained a model for the description as possible, even though the latter type of analysis has not been possible in this reanalysis for lack of raw data.

REFERENCES

Behr, M.J. (1967) A study of interactions between "structure-of-intellect" factors and two methods of presenting concepts of modulus seven arithmetic. Unpublished doctoral dissertation. Florida State University. Shorter version: Journal for Research in Mathematics Education, 1, 1970, 29-42.

Behr, M.J. & Eastman, P.M. (1975) Interactions between structure-of-intellect factors and two methods of presenting concepts of modulus seven arithmetic -- a follow up and refinement study. Journal for Research in Mathematics Education, 6, 150-157.

Bracht, G.H. (1969) The relationship of treatment tasks, personological variables and dependent variables to aptitude-treatment interactions. Unpublished doctoral dissertation, University of Colorado.

Brooks, L.R. (1967) The suppression of vizualization by reading. Quarterly Journal of Experimental Psychology, 19, 289-299.

Carry, L.R. (1967) Interaction of visualization and general reasoning in curriculum treatment in algebra. Unpublished doctoral dissertation, Stanford University.

Cronbach, L.J. (1975) Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Cronbach, L.J., & Snow, R.E. (1977) Aptitudes and instructional methods. A handbook for research on interactions, New York: Irvington.

Cronbach, L.J., & Webb, N. (1975) Between-class and within-class effects in a reported aptitude x treatment interaction: Reanalysis of a study by G.L. Anderson. Journal of Educational Psychology, 67, 717-724.

Guilford, J.P. (1967) The nature of human intelligence. New York: McGraw-Hill.

Gustafsson, J.-E, (1974) Verbal versus figural in aptitude-treatment interactions. Review of the literature and an empirical study. Reports from the Institute of Education, University of Göteborg, no 36.

Gustafsson, J.-E. (1976) Verbal and figural aptitudes in relation to instructional methods. Studies in aptitude-treatment interactions. Göteborg: Acta Universitatis Gothoburgensis.

Gustafsson, J.-E. (1977) Differential effects on imagery instructions on pupils with different abilities. Scandinavian Journal of Educational Research, 21, 157-179.

Gustafsson, J.-E. (1978) A note on class effects in aptitude x treatment interactions. Journal of Educational Psychology, 70, 142-146.

Gustafsson, J.-E.,& Svensson, A. (1978) Family size, social class, abilities and achievement: A study of interactions. Forthcoming.

Hancock, R.R. (1975) Cognitive factors and their interaction with instructional mode. Journal for Research in Mathematics Education, 6, 37-50.

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

Jöreskog, K.G. (1973) A general method for estimating a linear structural equation system. In A.S. Goldberger & O.E. Duncan (Eds.): Structural equation models in the social sciences. New York: Seminar Press, 85-112.

Jöreskog, K.G. (1974) Analyzing psychological data by structural analysis of covariance matrices. In R.C. Atkinson, D.H. Krantz, & R.D. Suppes (Eds.): Contemporary developments in mathematical psychology, Volume II, San Francisco: W.H. Freeman & Col., 1-56.

Jöreskog, K.G. & Sörbom, D. (1976) Statistical models and methods for test-retest situations. Im D.N.M. de Gruitjer, L.J.Th. van der Kamp & H.F. Crombag (Eds.): Advances in psychological and educational measurement. London, 135-137.

Jöreskog, K.G. & Sörbom. D. (1976) COFAMM -- confirmatory factor analysis with model modifications. A computer program manual. Chicago: International Educational Services.

Jöreskog, K.G.,& Sörbom, D. (1977) 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., 285-325.

Jöreskog, K.G. , & Sörbom, D. (1978) LISREL IV. A general computer program for estimation of linear structural equation systems by maximum likelihood methods. University of Uppsala, Department of Statistics.

Sörbom, D. (1975) Detection of correlated errors in longitudinal data. _British Journal of Mathematical and Statistical Psychology_, _28_, 138-151.

Sörbom, D. (1978) An alternative to the methodology for analysis of covariance. _Psychometrika_, _43_, 381-396.

Webb, L,F,. & Carry, L.R. (1975) Interaction of spatial visualization and general reasoning abilities with instructional treatment in quadratic inequalities: A follow-up study. _Journal for Research in Mathematics Education_, _6_, 132-140.

Werdelin, I. (1961) _Geometrical ability and the space factors in boys and girls_. Lund: Gleerups.

Winer, B.J. (1978) Statistics and data analysis: Trading bias for reduced mean squared error. In M.R. Rosenzweig, & L.W. Porter. (Eds.) _Annual Review of Psychology, Volume 29_. Palo Alto: Annual Reviews Inc.

Figure 1 The full LISREL model for the verbal/figural aptitude variables and the speed/understanding dependent variables.

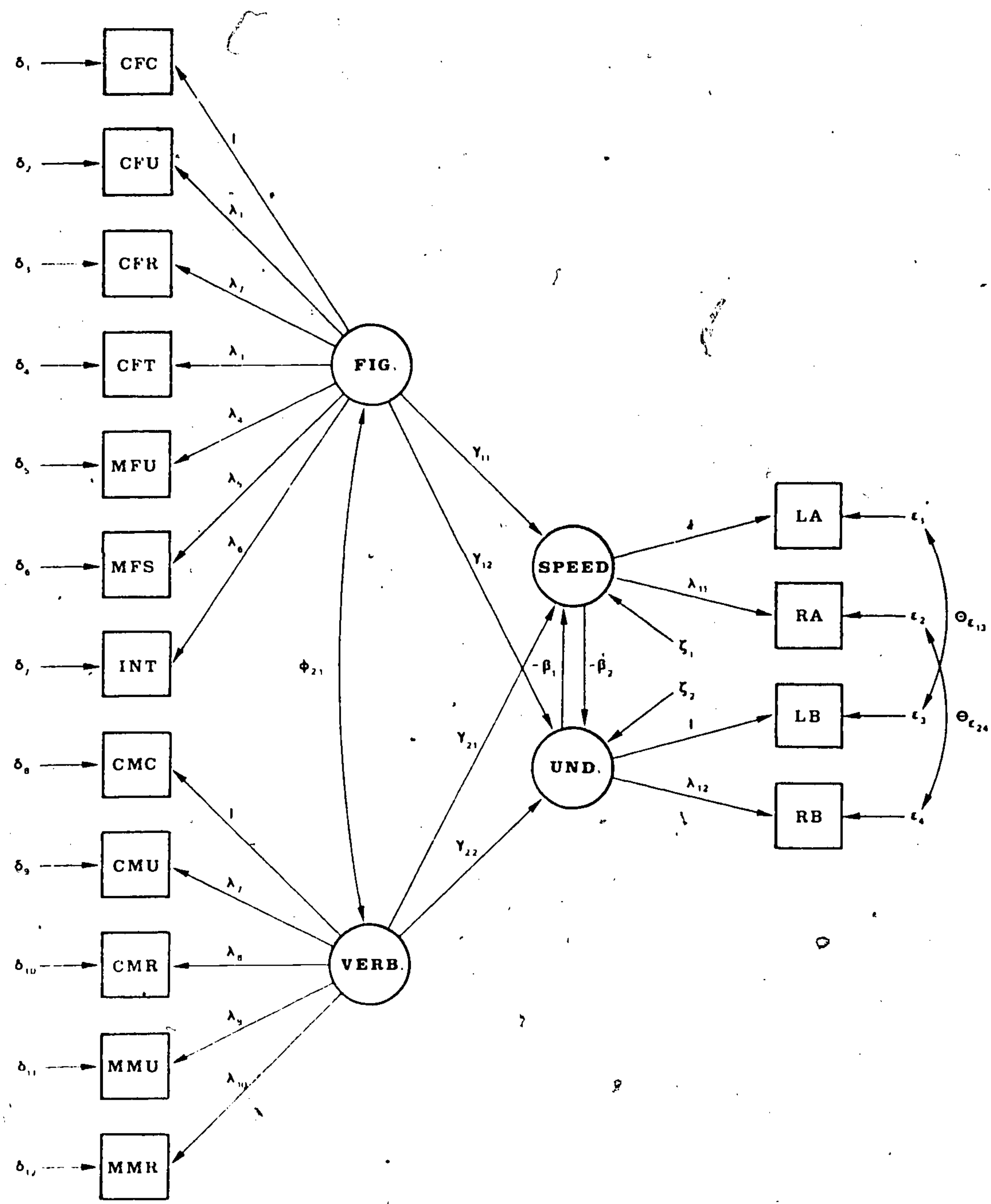


Table 1 The aptitude variables in the Behr study

Test	SI-cell	No of items	Time limit
Gestalt Completion Test	CFU	20	6
Figure Classification	CFC	14	3
Figure Matrix	CFR	15	7
Paper Folding Test	CFT	20	6
Map Memory	MFU	19	10
Object Memory	MFS	30	3 + time for recall
Wide Range Vocabulary Test	CMU	24	6
Word Classification	CMC	20	4
Verbal Analogies	CMR	30	9
Memory for word Meanings	MMU	30	5
Sentence Completion	MMR	35	10
Object-Number Test	MSR	15	5
Addition/Subtraction Test	MSI	120	4
Following Direction Test	(INT)	8	5

Table 2 Results from tests of fit of the 1- and 2-factor solutions for the aptitude variables

	No. of factors								
	1			2			Difference		
	χ^2	df	p	χ^2	df	p	χ^2	df	p
Equality 1) over groups	168.1	132	.02	158.4	131	.05	9.7	1	.00
No equality 1) over groups	139.0	108	.02	128.0	106	.07	11.0	1	.00
Difference	29.1	24	.22	30.4	25	.22			

1) The parameters are constrained (or not constrained) to be equal in the treatment groups.

Table 3 The factor loadings (standardized) in the 2-factor solutions for the aptitude variables

	Factor				Communality	
	1		2			
Test	VS	FS	VS	FS	VS	FS
CFU	.41	.19			.17	.04
CFC	.60	.57			.36	.32
CFR	.74	.51			.55	.26
CFT	.69	.52			.48	.27
MFU	.67	.19			.45	.04
MFS	.51	.20			.26	.04
INT	.50	.55			.25	.30
CMU			.58	.48	.34	.23
CMC			.44	.62	.19	.38
CMR			.66	.72	.44	.52
MMU			.42	.52	.18	.27
MMR			.39	.11	.15	.01

Table 4 Coefficients of structural relations between the latent aptitude variables and the speed/understanding outcome variables within treatments

	Speed			Understanding		
	VS	FS	t	VS	FS	t
DM:						
Verbal	.45(.76)	-.55(.74)	.94	.74(.62)	1.60(.92)	-.78
Figural	.29(.23)	.29(.15)	-.02	.39(.20)	.04(.19)	1.27
General	.43(.14)	.19(.08)	1.49	.61(.15)	.37(.10)	1.33
SM:						
Verbal	.24(.85)	-1.01(.75)	1.10	.63(.70)	1.20(.85)	-.52
Figural	.32(.23)	.44(.20)	-.39	.37(.19)	.14(.23)	.77
General	.37(.11)	.19(.08)	1.32	.53(.11)	.45(.11)	.51

Standard errors are shown in parentheses.

The t-values refer to approximate tests of equality of the within-treatment coefficients.

In the DM models the parameters in the measurement model for the aptitude variables are not constrained to be equal in the treatment groups, while in the SM models they are.

Table 5

Results from tests of fit of the full LISREL models with the learning/retention outcome variables.

	Verbal/figural aptitudes			General aptitude		
	χ^2	df	p	χ^2	df	p
DM:						
γ-parameters equal in the groups	243.95	196	.01	258.00	200	.00
γ-parameters not equal in the groups	242.29	192	.01	255.79	198	.00
Difference	1.66	4	.80	2.21	2	.33
SM:						
γ-parameters equal in the groups	275.29	221	.01	287.08	224	.00
γ-parameters not equal in the groups	274.25	217	.01	286.12	222	.00
Difference	1.04	4	.85	.96	2	.62

Table 6 Coefficients of structural relations between the latent aptitude variables and the learning/retention outcome variables within treatments.

	Learning			Retention		
	VS	FS	t	VS	FS	t
DM:						
Verbal	.50(.59)	.46(.73)	.04	.47(.48)	.56(.47)	-.12
Figural	.39(.19)	.24(.15)	.49	.07(.17)	-.03(.10)	.5[illegible]
General	.55(.14)	.33(.08)	1.36	.22(.15)	.10(.06)	.67
SM:						
Verbal	.37(.69)	-.19(.76)	.55	.46(.56)	.51(.45)	-.07
Figural	.39(.18)	.42(.20)	-.07	.06(.18)	-.02(.14)	.35
General	.48(.10)	.36(.09)	.89	.18(.13)	.13(.07)	.34

Standard errors are shown in parentheses.

The t-values refer to approximate tests of equality of the within-treatment coefficients.

In the DM models the parameters in the measurement model for the aptitude variables are not constrained to be equal in the treatment groups, while in the SM models they are.

Table 7 Results from tests of fit of the full LISREL models with the speed/understanding outcome variables.

	Verbal/figural aptitudes			General aptitude		
	χ^2	df	p	χ^2	df	p
DM:						
γ-parameters equal in the groups	244.54	196	.01	258.83	200	.00
γ-parameters not equal in the groups	240.43	192	.01	255.79	198	.00
Difference	4.11	4	.39	3.04	2	.22
SM:						
γ-parameters equal in the groups	275.17	221	.01	287.99	224	.00
γ-parameters not equal in the groups	272.01	217	.01	286.12	222	.00
Difference	3.16	4	.53	.87	2	.65

Reports from the Institute of Education, University of Göteborg

1. Andersson, Bengt-Erik & Nilsson, S-G: An application of the critical incident technique to the study of job and training requirements of shop mangers. June 1962.

2. Bullock, Donald, H: Research in programmed learning: Description of and Rationale for A Program of Experimental Analysis. November 1965.

3. Härnqvist, Kjell: Social Factors and Educational Choice. 1965.

4. Härnqvist, Kjell: Relative changes in intelligence from 13 to 18. December 1967.

5. Dahllöf, Urban & Lundgren, Ulf P: A project concerning macromodels for the curriculum process. A short presentation. Project COMPASS 12. April 1969.

6. Andersson, Bengt-Erik: Project YG (Youth in Göteborg) - A presentation of background, design, instruments and populations. Project UG 9. June 1969.

7. Dahllöf, Urban: Ability grouping, content validity and curriculum process analysis. Project COMPASS 13. June 196'

8. Dahllöf, Urban: Research on oral examination within the International Baccalaureats: Some general points of view and some inferences from the psychology of interviewing. Project COMPASS 20. January 1970.

9. Dahllöf, Urban: The materials and methods of implementation in the development of the curriculum. Outline of a model and some illustrations from Sweden. Paper read at the Conference of Comparative Education Society in Europe, Prague, June 1969. Project COMPASS 21. January 1970.

10. Dahllöf, Urban & Lundgren, Ulf P: Macro and micro approaches combined for curriculum analysis: A Swedish educational field project. Paper read at the annual meeting of the American Educational Research Association in Minneapolis, March 2-6, 1970, Project COMPASS 23. April 1070.

11. Andersson, Bengt-Erik: Actual and perceived attitudes among adolescents and adults toward each other. A study of a group of Swedish teenagers and their parents. The Stug-project 3. April 1971.

12. Andersson, Bengt-Erik: Parental and peer influences and adolescent peer orientation. Project UG 12. April 1971.

13. Andersson, Bengt-Erik & Ekholm, Mats: The generation gap. Frame of reference and design of the Stug-project. The Stug-project 4. May 1971.

14. Andersson, Bengt-Erik & Ekholm, Mats: Swedish Youth of To-day. Description of a Research Project About the Generation GAp. Paper presented to the Interantional conference on "Socialization Processes in the Contemporary European Youth" in Bratislava June 14-17 June, 1971. The Stug-project 5. June 1971.

15. Marton, Ference & Sandqvist, Gjertrud: Learning while typing. TIPS 3. September 1971.

16. Dahllöf, Utban, Lundgren, Ulf P, & Siöö, Margareta: Reform implementation studies as a basic for curriculum theory: Three Swedish approaches. Reprint from curriculum Theory Network 7, 1971: Monograph Supplement: Elements of Curriculum Development, edited by F. Michael Connely with the assistance of John Herbert and Joel Weiss. Project COMPASS 39. December 1971.

17. Lundgren, Ulf P: An empirical study of the teaching process. Theory design and methodology. Project COMPASS 40. January 1972.

18. Härnqvist, Kjell: Training and career structure of educational researchers. Reprint of paper read at Colloquium of Directors of Educational Research Organizations, London 10-12 November 1971, arranged by Council of Europe. March 1972.

19. Härnqvist, Kjell: Canonical analyses of mental test profiles. Project MID 2. April 1972.

20. Härnqvist, Kjell & Bengtsson, Jarl: Educational reforms and educational equality. Contribution to a Reader on Social Stratification edited by Scase. April 1972.

21. Bradley, Gunilla: Women's interest in promotion in relation to job satisfaction and home and school background. September 1972.

22. Åsberg, Rodney: An attempt at a Pedagogical Evaluation of the Development Studies in Lund 1969-70: Housing, Building and Planning. May 1971.

23. Marton, Ference, Fransson, Anders, Jonsson, Barbro, Klenell Ann-Charlotte & Roos, Birgitta: Differential effects of stress-including instructions on anixiety. TIPS 7. October 1972.

24. Lundgren, Ulf P: Educational Process Analysis. Two articles. Project COMPASS 48. November 1972.

25. Lundgren, Ulf, P: Pedagogical frames and the teaching process. A report from an empirical curriculum project. Paper read at the AERA meeting in New Orleans, 26 Feb - 1 March 1973. Project COMPASS 51. February 1973.

26. Andrae, Annika: Late-bloomers and alternative courses of study at the senior level of the compulsory comprehensive school and qualification for academic education at the upper secondary level. Project COMPASS 54. February 1973.

27. Lundgren, Ulf P: Pedagogical roles in the classroom. A contribution to Eggleson, J. 1973: Contemporary Research in the Sociology of Education. London: Methusen. Project COMPASS 56.

28 Andersson, Bengt-Erik: The Generation Gap - Imagination or Reality? Paper presenteted to the annual meeting of the International Society for the Study of Behavioral development. The Stug-project 21. May 1973.

29. Andersson, Bengt-Erik: Project YG. (Youth in Göteborg) - A Study of a group of Swedish adolescents. Project UG 13. June 1973.

30. Dahllöf, Urban: Data on curriculum and teaching process: Do they make any difference to non-significant test differences - undr what conditions? Feb 27 1973. Project COMPASS 60. August 1973.

31. Andrae, Annika & Dahllöf, Urban: Process Analysis of Non-Graded Rural Schools in Sweden. Outline of an Evaluation Project. Project PANG 5. Fabruary 1973.

32. Lundgren, Ulf P: & Wallin, Erik: Some NOtes on Education as a Science. September 1973.

33. Norvell, Bo: The Class Size and the Teaching Process. Part 1. October 1973.

34. Härnqvist, Kjell: Individual differences in higher education Selection or Adaption? Plenary lecture at the Congress on Methodology of Research in Higher Education, Rotterdam, 3-5 December 1973. Project MID 9. December 1973.

35. New Trends in Evaluation. Report from a conference on educational evaluation with an introductory paper by professor Robert E. Stake, University of Illinois. January 1974.

36. Gustafsson, Jan-Eric: Verbal versus Figural Aptitude-Treatment Interactions. Review of the litterature and an empirical study. Project MID 10. March 1974.

37. Lindblad, Sverker: Simulation and guidance. Experiences from planning and evaluation teaching with simulation designed for grade 6 in Swedish compulsory schools. Project-group UPLM 24. April 1974.

38. Gustafsson, Jan-Eric: Implications of interactions for the experimental research on teaching methods. Project MID 12. June 1974.

39. Marton, Ference: On non-verbatim learning: 1. Level of processing and level of outcome. December 1974.

40. Marton, Ference: On non-verbatim learning:11. The erosion effect of a taskpinduced learning alorithm. December 1974.

41. Marton, Ference & Dahlgren, Lars-Owe: On non-verbatim learning: iii. The outcome space of some basic concepts in economics. December 1974.

42. Marton, Ference: On non-verbatim learning: VI. Some theoretical and methodological notes. December 1974.

43. Lundgren, Ulf P: Analyse von Curriculumprozessen - der engere Ansatz: Curriculum-prozess als instruktionsgeschehen. Research Report from the MAP-group 1975:1. January 1975.

44. Marton, Ference, Dahlgren, Lars-Owe, Säljö, Roger & Svensson, Lennart: The Göteborg Project on Nonverbatim Learning. Four invited papers presented at the American Educational Research Association, 30 March - 3 April, Washington, DC. April 1975.

45. Kilborn, Wiggo & Lundgren, Ulf P: A Contribution to the Analysis of arithmetic Teaching and Learning. A paper presented at the first DIMO-workshop, 31st of August - 5th of September, 1975, in Wasserlos, BDR. Research Report from the MAP-group 1975:14. December 1975.

46. Gustafsson, Jan-Eric: Inconsistencies in aptitude-treatment interactions as a function of producers in the studies and methods of analysis. Project MID 16. March 1976.

47. Gustafsson, Jan-Eric: Differential effects of imagery instructions on pupils with different abilities. Project MID 17. April 1976.

48. Ekholm, Mats: Social development in school. Summary and excerpts. Project SOS 23. May 1976.

49. Stangvik, Gunnar: Approaches to the analysis of learner-task interactions and some implications for the study of pedagogical processes. Project YP 7. January 1976.

50. Andrae, Annika: Non-graded instruction in small rural lower secondary schools. A presentation of the PANG-project. Paper read at the INTERSKOLA conference in Sveg, July 1976. Project PANG 20. July 1976.

51. Patriksson, Göran: Attitudes toward olympic games of Swedish adolescents. Paper presented at the international congress of physical activity sciences in Quebec City 11-16 July 1976. September 1976.

52. Gustafsson, Jan-Eric: A note on the importance of studing class effects in aptitude-treatment interactions. Project MID 19. September 1976.

53. Gustafsson, Jan-Eric: Spatial ability and the suppression of visualization by reading. Project MID.September 1976.

54. Dahlgren, Lars-Owe & Marton, Ference: Investigations - into the learning and teaching of basic concepts in economics - a research project on higher education. Paper presented at the Congress of the European Association for Research and Development in Higher Education, August 30 - September 3, 1976, Louvain la Neuve, Belgium. September 1976.

55. Marton, Ference: Skill as an aspect of knowledge. Some implications from research on students conceptions of central phenomena in their subjects. Paper presented at the Second International Conference on Improving University Teaching, July 13-16, 1976, Heidelberg, F.R. Germany. September 1976.

56. Andrae Annika (Ed): Non-graded instruction. Research organization and design. Administration and daily teaching experiences in small rural lower secondary schools. Experiences from the PANG-project. Paper read at the INTERSKOLA conference in Sveg, July 1976. Project PANG 23, August 1976.

57. Sandgren, Björn & Åsberg, Rodney: On cognition and social change. A report from a pilot study regarding the effect of schooling on cognitive growth and attitudes towards social change in Pakistan.October 1976.

58. Sandgren, Björn: Relation between cognition and social development. January 1977.

59. Entwistle, Noel: Changing approaches to research into personality and learning. February 1977.

60. Härnqvist, Kjell: Primary Mental Abilities at Collective and Individual Level. October 1977.

61. Härnqvist, Kjell: Enduring Effects of Schooling - A Neglected Area in Educational Research. October 1977.

62. Härnqvist, Kjell: A Note on the Correlations between increments, Cumulated Attainment and a Predictor. October 1977.

63. Gustafsson, Jan-Eric: The Rasch Model for Dichotomous Items: Theory, Applications and a Computer Program. December 1977.

64. Härnqvist, Kjell & Stahle, Gun: An Ecological Analysis of Test Score Changes over Time. December 1977.

65. Dahlgren, Lars Owe: Effects of University Education on the Conception of Reality. Paper to be presented at the Fourth International Conference on Improving University Teaching. July 26-29, 1978, Aachen, F.R. Germany. May 197[illegible]

66. Marton, Ference: Describing conceptions of the world around us. Sept 1978.

67. Gustafsson, Jan-Eric & Lindblad, Torsten: The rasch model for dichotomous items: A solution of the conditional estimation problem for long tests and some thoughts about item screening procedures. June 1978.

68. Marklund, Ulla: Smoking in comulsary schools. Interviews and improvisations. Project Behavioral science tobaco research. September 1978.

69. Svensson, Lennart: Some notes on a methodological problem in the study of the relationship between thought and language. - Describing the thought content in terms of different conceptions of the same phenomenon. September 1978.

70. Härnqvist, Kjell & Gustafsson, Jan-Eric: International references to Swedish educational research: A citation study. October 1978.

71. Lybeck, Leif: A Research Approach into science education at Göteborg. Paper presented at the Joint ICMI/ICPE/CTS/UNESCO/IDM-conference on 'Cooperation between Science Teachers and Mathematics Teachers', September 17-23, 1978, Institut für Didaktik der Mathematik, Bielefeld, FRG. October 1978.

72. Lybeck, Leif: Studies of mathematics in science lesson in Göteborg. Paper presented at the Joint ICMI/ICPE/CTS/UNESCO/IDM-conference on 'Cooperation between Scince Teachers and Mathematics Teachers', September 17-23, 1978, Institut für Didaktik der Mathematik, Bielefeld, FRG, October 1978.

73. Gustafsson, Jan-Eric & Lindström, Berner: Describing and testing aptitude-treatment interaction effects with structural equation models: Reanalysis of a study by M.J. Behr. December 1978.