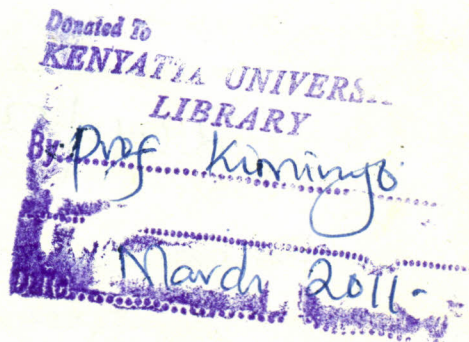


A CAUSALITY AND PREDICTIVE VALIDITY STUDY OF
O-LEVEL IN RELATION TO THE PERFORMANCES IN
PRIMARY TEACHERS' COLLEGES IN KENYA

KENYATTA UNIVERSITY LIBRARY

BY: HANIEL NYAGA GATUMU

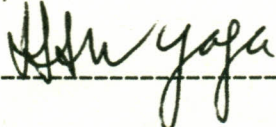


"A THESIS SUBMITTED TO THE DEPARTMENT
OF EDUCATIONAL PSYCHOLOGY FACULTY OF EDUCATION
IN FULFILMENT FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY IN THE
KENYATTA UNIVERSITY".

KENYATTA UNIVERSITY LIBRARY

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.



HANIEL NYAGA GATUMU

"The thesis has been submitted for examination with our approval as University Supervisors."

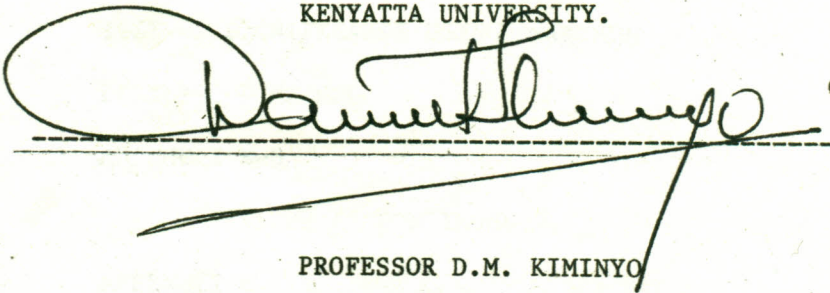


DR. F.O. INGULE

LECTURER

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

KENYATTA UNIVERSITY.



PROFESSOR D.M. KIMINYO

PROFESSOR

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

KENYATTA UNIVERSITY.

TABLE OF CONTENTS

	PAGE
DECLARATION	ii
TABLE OF CONTENT	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
CHATER I	1
INTRODUCTION	
CHAPTER II	
REVIEW OF RELATED LITERATURE	34
CHAPTER III	
DESIGN OF THE STUDY	71
CHAPTER IV	
THE ANALYSIS AND DISCUSSION OF RESULTS	83
SUMMARY AND IMPLICATION	126
REFERENCES	128
APPENDICES	
APPENDIX A	
TEST OF SIGNIFICANCE USING MAXIMUM	135
LIKELIHOOD METHOD	
APPENDIX B-D	
TABLE II CONTINUED	136
APPENDIX E	
OTHER ASSUMPTIONS OF PATH	140
ANALYSIS MODEL CONSIDERED	
APPENDIX F	
FURTHER DISCUSSION OF	
MULTIPLE CORRELATION ANALYSIS	141

ACKNOWLEDGEMENT

I wish to express my great gratitude and appreciation to the sponsors the joint International Development Research Centre (IDRC)/Ford Foundation for their generous financial assistance for this study.

I am also indebted to Dr. E. Maritim of Kenya Examination Council for making the Preservice examinations' raw scores for those students who finished in Primary Teachers' Colleges in 1985 and 1986, available to me.

Many thanks also go to the Deans of Curriculum and administrators in the Primary Teachers' Colleges used in the study who greatly assisted in making available the O-level results for the students (subjects). A job which was very demanding for the results were in separate files for every individual student.

Many gratitude also go to my professors at Michigan State University, U.S.A. who exposed me to other realms of Education in different culture and setting. Special gratitude to Dr. Schmidt who taught me path analysis and research methods, Dr. Stapleton - several courses in advanced statistics, Dr. Mehrens and Dr. Ebel - measurement just to mention a few.

Great indebtedness is expressed to the members of department of Educational Psychology, Kenyatta University who made many useful suggestions for improvement. I particularly appreciate

the criticism of my colleagues who pointed out numerous shortcomings in the approaches much so in path analysis. I do admit this mode of attack, or paradigm as called by Kuhn, like any other is not entirely adequate. Each one of them explains just limited amount of reality.

My special appreciation to Prof. Kiminyo and Dr. Ingule, my supervisors, whose guidance, comments and suggestions were very valuable throughout the study; to Dr. Karugu for inspiration and encouragement, the list is too long to mention.

Sincere appreciation is expressed to all who typed this project Miss. Rose Ndovo, Miss. J. Gichia etc. Also who coded and entered the data, assisted in debugging computer programmes used in the data analysis as well as those who gave moral encouragement.

I am also greatly indebted to my wife Jane Ciumwari and our children Karimi and Wambu for their patience, encouragement and understanding.

Thanks to everyone who contributed in one way or another to the completion of this study. May God bless them all.

Nairobi, June 1989

Haniel Nyaga Gatumu

ABSTRACT

The study was attempting to answer the following questions:-

- (i) How well does O-level predict the performance in Primary Teachers' Training Colleges ?
- (ii) Is there any causal relationship between O-level grades with the performance in Primary Teachers' Training Colleges ?

In attempt to answer these questions a large sample of 1622 subjects was obtained by method of cluster sampling. All the Primary Teachers Training Colleges in Kenya were considered as clusters and 5 cluster (colleges) were randomly selected. Thus 1622 subjects were from 5 different colleges.

The independent variables were O-level grades. Dependent variables were scores of final year college examinations obtained from Kenya Examination Council. A predictive validity using both intercorrelation and multiple regression analyses was done for this sample, using path analysis a causality study was also done for the same sample, trying two models in the latter analysis.

The findings were, O-level grades predict very well performance in final year college examination. It was also observed that definitely O-level grades have a causal relationship with performance in final year college examinations.

OK

KENYATTA UNIVERSITY LIBRARY

CHAPTER I

INTRODUCTION

This is organized into 3 major sections:-

Section One deals with general statements about Causality.

Section Two deals with general statements about predictive validity.

Section Three deals with the definition of terms and included here are the mathematical definitions too.

SECTION ONE

I CAUSALITY:

Anyone who has studied correlation (defined below) is familiar with the statement that correlation does not imply causation. In other words a high correlation coefficient does not necessarily indicate a cause-and-effect relationship between the two variables. Then one wonders whether causal inference is really possible.¹ Indeed it is possible through the application of multivariate statistical methods, known as "Path analysis", to a stated structural model or causal model.

Correlation means a statistical relationship between a set of variables, one of which have been experimentally manipulated. Although correlations (covariances) can be computed from experimental data, usually the term correlation (covariances) is reserved for a relationship between unmanipulated experimental variables. Causal inferences are straight forward with data obtained through "the experimental designs". More often random assignment of subjects to experimental treatment and control conditions the backbone of what is referred to as the true 'experimental design' may not be possible and consequently we have correlational data in our sense. We still

¹ The method discussed here of studying a cause and effect relationship is analytical in nature. The method of using experimental design to study causality cannot be underscored. It is highly recommended method, by the author, if feasible. The author agrees in total that it is by controlling all sources of fluctuation within the setting, time and subjects, except for the experimental treatment variable, the researcher can be absolutely certain that results are due to the experimental treatment and not other confounding variables. In other words, it is under true experimental conditions are causality statement absolutely justifiable.

may wish to make causal inferences with such kind of data.

Inferences means confirming or disconfirming of a scientific hypothesis by the use of data. Confirming hypothesis or else referred to as "failing to reject the the null hypothesis" does not imply in any other way that the hypothesis is proven true.

To interpret data, we must have a set of reasonable assumptions about how the data were generated and additional assumptions about how the data can be summarized. The set of assumptions about how the data is generated gives the model while the data is summarized by statistical (analytical) methods.

Inference is usually evaluated statistically. At best one knows only the probability the results would have obtained by chance given the model. There is no certainty on the conclusions, it is only probability statements being made. Thus, this is why we may say that although statistical analytical methods may be used to study causality, it is necessary to carry out true experimental design in order to establish beyond any reasonable doubt that there is a cause-and-effect relationship or causality.

Another thing about analytical method is that the model or the assumptions on which the inference is based can always be questioned. If the assumptions and model are false or one of them is false, our analyses may be quite irrelevant to the experiment that was actually performed. Consequently the assumptions used here for both statistical or causal models will be those which can be supported by theoretical considerations.

CAUSATION: a causal statement has two components, conditions must hold for a scientist to claim that X causes Y

- (1) time precedence
- (2) relationship
- (3) nonspuriousness

I will need to explain what we mean by the last two terms, relationship and nonspuriousness, in details otherwise time precedence is self explanatory. X must occur before Y so that we can talk of X causing Y.

Relationship To measure a relationship between two variables, we need define the meaning of no relationship between variables, or as it is sometimes called independence (zero correlation between the two variables).

Two variables X and Y, say, are independent if knowing the value of one variable, say, X provides no information about the value of Y'. More technically X and Y are independent if and only if the conditional distribution of Y does not vary across X. If variables are not independent, then they are said to be related (correlated). In judging whether two variables are related, it must be determined whether the relationship could be explained by chance.

Nonspuriousness:- We say the relationship between variables X and Y is nonspurious if there exist ^S no variable Z which causes both X and Y such that once variable Z is controlled, relationship between X and Y disappears. Here distinction should be made between a spurious

variable and intervening variable². A variable Z intervenes between X and Y if X causes Z and Z in turn causes Y. Thus an intervening variable elaborates the causal chain. Observe controlling for either a spurious variable or an intervening variable makes the relationship between X and Y vanish. The issue of spuriousness, is really a stumbling block in causal analysis and it may be referred to as the excluded variable.

A causal law is of the form for all Q, X causes Y, Q referring to the subjects or objects to whom the causal law applies, X refers to the cause and Y the effect.

Both X and Y are variables and consequently they can put in a form of functional equation of form:-

$Y=f(X)$ which we call structural equation or causal equation.

The typical structural equation is a linear one of form

$$Y=a_0 + a_1X$$

or $Y=a_0 + a_1X_1 + \dots + a_n X_n$

the a_1, \dots, a_n or a_i are/is called the causal parameter.

For a set of variables in a causal model, most if not all information can be summarized in the covariances (correlations) between variables. Covariances (correlations) structure is sufficient statistics in the statistical sense. The information about the set of

² Confounding variable - Here will be used to mean an extraneous variable that is any variable not considered explicitly in a design but which is correlated with an independent variable in the design which the researcher would like to argue has a causal relationship with the dependent variable. Hence note the difference between intervening and confounding variables.

parameters of linear model, causal or otherwise, is all contained in the covariances (correlations) structures between the variables.

There is difficult in translating theory into equations. This process of translating theory into equations is called 'specifications'. Most theories in the social science are not strong enough to elaborate the exact form of equations, and so the structural modeler must make a number of theoretical assumptions or specifications. Each specification should have some justification drawn from the theory.

Besides specification being based on substantive theory there are two other sources of specification:-

- (i) classical measurement theory and
- (ii) the experimental design.

The measurement theory is well developed and is useful in specifying structural equations. For example, classical test (measurement) theory posits that errors of measurement are uncorrelated with true score. This is often helpful in formulating structural models.

Factorial design and random assignment to experimental treatment and control conditions (true experimental design e.g Solomon four group design) yield independent causes that are uncorrelated with the unmeasured causes (confounding variables) of the effect. Likewise, longitudinal designs may bring with the specification that certain parameters do not change over time. Usually a structural modeler should exploit all the 3 types of specifications namely, substantive

theory, measurement theory and true experimental design - accordingly to the suitability of the problem under consideration. This may be one or more of the specification.

Specification error is something a structural modeler should seriously consider, that is, one of the assumptions of the model is incorrect. It is not enough to criticize a model just because it has a specification error. One must show that it seriously biases estimates of the parameters of the model. It may be that the whole model or parts of it are very robust even with specification error.

SECTION TWO

II PREDICTIVE VALIDITY

Validity of a test is high if the test measures what it is supposed to i.e. if it gives the information the decision maker needs.

No matter how satisfactory it is in other respects, a test which measures what it is not supposed to is worthless.

Validity can be classified as:-

- (i) Predictive
- (ii) Concurrent
- (iii) Content
- (iv) Construct

The instrument's validity is seldom a problem when we are dealing with physical measurements such as weight, length etc. It is obvious, weight is measured with a balance (provided the balance is functioning correctly) etc. However, with methods used for measuring psychological variables, it is necessary to test empirically whether the instrument is valid in every case. When we construct a questionnaire to obtain a measure of attitudes we must show that the scores really distinguish between degrees of attitude and not other differences. Often it might appear, obvious that a test measures some trait, while empirical testing, shows that it measures something entirely different.

The application of a psychological instrument must not be based on subjective confidence that the instrument works in practical situations. Empirical studies show that in clinical practice impressions of that type are very unreliable bases for judging the dependability of a given method (Magnusson, 1959). It is necessary that every method's validity is tested empirically in all the different situations in which it will be used.

When a test is constructed the primary interest should be whether it has high validity. Reliability is mainly of interest as a necessary condition for valid measurements. It should be noted that high reliability is a necessary but not sufficient requirement for high validity. However, it happens occasionally that high reliability, for instance, in the form of

agreement among makers or judges giving subjective ratings in essays, say, is taken to be a sign of ratings' validity. Such an agreement is not a sufficient basis for concluding high validity. It may arise because the makers have the same bias in common. Note the narrow difference which exists between this kind of reliability and concurrent validity (predictive) defined later.

What
is
sufficient
for
validity

Validity is usually computed as a validity coefficient (correlation coefficient) for the relationship between test data and criterion data. It should be always be remembered that a high coefficient for the relationship between test and criterion does not necessarily mean that the test measures what we believe it to measure. Let us suppose that we have constructed a test in logical aptitude for predicting success in engineering studies, assuming that high logical aptitude is a necessary requirement for success in such studies. If we now obtain a high correlation coefficient for the relationship between test scores and grades in engineering studies, we cannot be certain that the test measures logical ability. It may be infact it measures verbal ability, since comprehending each individual item places great demands on this ability. But, since verbal ability also is related to success in engineering studies, we still obtain a relatively good correlation. We can cite many examples of such kind which show cases where it was believed a test measured certain trait, and where this supposition was later shown to be false.

One example of this is, a number of tests which were supposed to measure pure spatial aptitude, more careful control has shown certain individuals solved the items correctly using purely logical methods instead.

which ones

When we are computing validity coefficients, it is necessary that test data and criterion data be determined independently.

Prediction studies provide the researcher with 3 types of information (Borg & Gall, 1971):-

1. The extent to which a criterion behaviour pattern can be predicted.
2. Data for theory building about possible determinants of the criterion behaviour pattern.
3. Evidence regarding the predictive validity of the test or tests that are correlated with the criterion.

Prediction studies can be differentiated in terms of which of these types of information the researcher is most interested in obtaining. In some studies the emphasis is on a particular criterion behaviour and one or more examination grades or personality and aptitudes tests are used to predict this criterion. In other studies, a researcher might be primarily interested in theoretical significance of his findings.

Prediction research has made a major contribution to educational practice. Many prediction studies have been aimed at short term prediction of students' performance in a specific course of study. Others have aimed at long term prediction of general academic success.

The findings of the studies have greatly aided school and University personnel in choosing students most likely to succeed in a particular academic environment or course of study. Also prediction studies provide the scientific basis for the teachers' efforts to help students plan their academic future.

Educational researchers have been searching for indicators of success other than grades and test scores, thus non-intellective predictors, and some have been found. However none has yet achieved enough demonstrated success to be widely adopted (Watkins & Astilla, 1980). Some educators are beginning to feel that the student who makes the best grades is not necessarily the most valuable student. Stalnaker (1965) nicely put it "..... we want to find students who will succeed in college - but much more important, will also use their college education in some socially desirable, productive way after graduation. How relevant are grades to this goal?..... Do you inquire of your accountant, your doctor or your lawyer the grades he received in college?".

Predicting grades has little social significance. Hollard and Richard (1965) have pointed out that a student's extra-

curricular achievement may be similar to achievement after graduation than is the academic achievement represented by grades. They demonstrated that academic and extracurricular achievement in college are not highly correlated and urged greater use of extracurricular achievement as an alternative criterion to grades in development of selection devices. Hoyt (1966) concluded from a review of relevant studies that there is little relationship between college grades and post college achievement.

The validity is often not the only issue of concern in a selection program. The use of non-intellective scales would certainly raise very many problems which although always present are hidden when intellectual predictors are used. Stalnaker (1965) stated the issue well when he said "In a program very well in public eye, predictive validity alone cannot rule..... Suppose there develop sound evidence that among the highly intelligent, the most conforming, compulsive, dependent, unoriginal individuals do best in college, would we then try to limit our selection to students having these characteristics?"

There is an acute lack of opportunities in secondary and tertiary education. For instance the number of pupils qualified for high school in our country, Kenya, considerably exceeds the number of high school places available, and as a consequence the

number of pupils who can be admitted³. As a result admission policy for secondary and higher education is usually very 'strict'.

A view has sometimes been put forward that once a student has become qualified for a certain type of school he has a right to be admitted. If for some reason only a restricted number of applicants can be accepted, this right should lead to selection on the basis of drawing lots among the group of qualified applicants. This point of view is bound to be challenged on educational, ethical, and economical grounds especially in developing countries where the need for optimal use of scarce educational and financial resources as well as the urge for quick development require that some form of rational selection should be adopted. Hence in general principle, those who have the biggest chance of successfully completing the course within given limits of time should be admitted.

How can we determine one's chance, of future school success in a valid, reliable and efficient way? Though in this study the concern is a previous performance for predicting future success, it should be remembered that estimates of previous school performance and maximum performance tests are not the only categories of possible predictors, as indicated earlier.

3 A rejection rate of about 80 percent is not all rare, rejection rates were 72, 77 and 63 percent for years 1975, 1976 and 1983 respectively for high schools. The figures for 1975 and 1976 were obtained from official reports from examination office in Ministry of Education, Nairobi; for 1983 from Daily Nation No.7174 30th December 1983.

When one does a study on predictive validity, one is confronted with a far reaching difficulty of considering the quality of education exposed to each subject (in this study the factor will not be considered, another study could be planned for looking at the factor and its effect). The school performance is dependent upon at least two important variables; the student's level of scholastic aptitude and the quality of education. Where the quality of education is more uniform (as assumed in this study), the variance in school grades can be considered as a reflection of student's ability for learning. But when the quality of education varies widely the school performance is not a valid and reliable indicator of student's capabilities. This implies, in those cases where school performance is used as a predictor for future school success, and for selection and admission purposes, that students from poorer schools as well poorer background are discriminated against.

Many others

A number of environmental factors influence school performance in varying degrees namely, socioeconomic status, cultural and linguistic differences just to mention a few. Whenever a "nonability" factor discriminates between students with better and poorer school performance, the use of the school performance variable as a predictor seems less justified. Though this might be "less justified" it does not necessarily mean that past school performance does not correlate with future performance. Hence

it is worthwhile appreciating that what we measure by our tests is always a result of learning process influenced by many factors including quality of education and the diversity of stimuli in the individual environment.

Also that we measure is not free from "measurement error". If each individual score is regarded as the sum of two components, a true score and an error score, then the "measurement errors" refer to errors produced by factors which result in individual scores differing from one parallel test to the other, even though in the study some variables have been assumed not to have any measurement error. This is clearly indicated in the models in each case but usually the predictor variables have been assumed to have no measurement error while criterion variables are assumed to have measurement error.

Section Three

Definition of Terms

0-level This consists of achievement tests done after four years of Secondary Education. The candidates offer 6 to 9 subjects each tested by one or more examination tests. The subjects range from Social Science (English, Kiswahili, Religion Education, Georgraphy etc.) to Natural Sciences (Biology, Mathematics, Chemistry etc.) The candidates are classified into Division I (best), II, III, IV or Fail which is judged by their aggregate performance of their best six subjects which need come from certain laid groups of subjects. The examinations

are annually and centrally administered by the Examinations Council to all Secondary Schools in Kenya. Some natural science subjects do have a practical component in them but more weight is on the achievement test. For samples considered the number of those who had offered technical subjects (also offered for O-level) was negligible - hardly anyone.

P1 - This is Primary Teachers' status determined by the performances in O-level and in final year Primary Colleges examination. Those with Division III or better are admitted for P1 course and have to pass the final year college examination to be awarded the P1 certificate.

Validity - In general, the validity of a method is the accuracy with which meaningful and relevant measurements can be made with it, in the sense that it accurately measures the traits it was intended to measure. Validity is one aspect of dependability in testing procedures.

Other dependability being reliability. When we deal with reliability we are not interested in what the test measures, we only find whether it gives the same results on repeated measurement. When testing validity, we investigate whether the test whose reliability is known measures what it has been constructed to measure.

I distinguish here among four concepts of validity in accordance with the recommendations on nomenclature presented in "Technical Recommendations for Psychological Tests and Diagnostic Technique" (1954). These are:-

- (i) Predictive validity
- (ii) Concurrent validity
- (iii) Content validity
- (iv) Construct validity

(i) Predictive Validity:-

Very often especially in selection or classification the decision is based on a person's expected future performance as predicted from the test score. If these expectations are confirmed the test has given highly useful information but if the predictions do not correspond to what happens later, the test was worthless. To know how validly the test predicts, a follow-up study is required. For instance a predictor such as a student's rank in his secondary school graduating class, predicts

a criterion such as his average university grade during the freshman year (first year).

(ii) Concurrent Validity:-

In many situations for which tests are developed some more cumbersome method for collecting information is already in use. For instance, tests intended for clinical diagnosis of brain damage, on which a number of neurologists agree, can be for example be used as a criterion for the validity of a brain damage test. This kind of check in which two sources of information are obtained at nearly the same time is the one called "concurrent validity". Concurrent and predictive validities differ essentially in time factor otherwise they are the same.

There is no differences in principle between the methods of computing predictive and concurrent validities. Both are determined by computing the correlation between test scores and measures of a criterion variable, and the validity is expressed in both cases as a correlation coefficient. The coefficient of validity gives the test's validity with respect to the variable which is defined by criterion measurement. For concurrent validity, the question is whether the test measures what it is intended to measure, and the validity coefficient indicates how adequate the test data are as a basis for diagnosis, in the widest sense of the word. For predictive validity, the question is the accuracy

with which the test predicts what it is intended to predict - a question which especially important for vocational guidance, selection, and classification. It should be noted here that good predictive ability presupposes good diagnostic ability. If something cannot be measured accurately, it cannot be used as a basis for valid prediction.

(iii) Content Validity:-

This type of validity is applicable when we are estimating extent to which a school test, for example, covers some field of study. If a course is supposed to teach a unit e.g. 'probability' it would not be fair to measure its effectiveness by a test on a different unit e.g. 'calculus'.

The test items can be regarded as a sample from a population representing the content and the aims of the course. Content validity is determined by the extent to which the sample of items in the test is representative of the total population. Before content validity can be estimated it is necessary to define explicitly the aims of the instruction given in the field to which the test refers, the material which the pupils should have grasped, the relative importance of different parts of the course etc. Unlike predictive or concurrent validity, content validity cannot be expressed as a validity coefficient.

(iv) Construct Validity:-

The tester is interested in what the scores mean psychologically or what causes a person get a certain test score. Construct validity, unlike predictive and concurrent validity, is not expressed as a single coefficient representing the correlation between test and criterion measurements. The concept of construct validity is especially useful with reference to tests measuring traits for which external criteria are not available.

The procedure for testing construct validity is the same as the deductive method currently applied in all scientific research. The concept of construct validity is an application of this method to the problem of evaluating the accuracy of predictions based upon a test. This evaluation takes place in accordance with the classical procedure: theory-deduction-hypthesis-experimental testing-data which falsify or verify the hypothesis.

Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned, and must use indirect measures. The trait or quality underlying the test is of central importance, rather than either the test behaviour or the scores on the criteria.

Construct validity would be involved in answering such questions as: To what extent is this test of intelligence culture free? Does this test of "interpretation of data" measure reading ability, quantitative reasoning, or response sets? How does a person with A in one subject and B in another differ from a person who has these scores reversed?

Related Mathematical definitions

Consider a situation where $p + q$ quantities $\eta_1, \eta_2, \dots, \eta_p$ and $\xi_a, \xi_b, \dots, \xi_q$ are thought of as measurable where p causal relations determine the η 's, absolutely and uniquely by the ξ 's. Let the equations which express these causal interrelations be:-

$$\eta_1 = F_1(\xi_a, \xi_b, \dots, \xi_q, \alpha_{11}, \dots, \alpha_{1k})$$

$$\eta_2 = F_2(\xi_a, \xi_b, \dots, \xi_q, \alpha_{21}, \dots, \alpha_{2k})$$

$$\eta_p = F_p(\xi_a, \xi_b, \dots, \xi_q, \alpha_{p1}, \dots, \alpha_{pk})$$

These equations are referred to as "structural equations or causal equations". Each variable η is assumed to be absolutely and uniquely determined (or 'caused') by the set of ξ variables.

Usually the quantities which can be measured are not η 's and ξ 's but rather Y 's and X 's which satisfy the error equations:-

$$Y_1 = \eta_1 + \epsilon_1$$

$$Y_2 = \eta_2 + \epsilon_2$$

⋮

$$Y_p = \eta_p + \epsilon_p$$

$$x_a = \xi_a + \delta_a$$

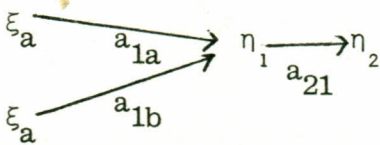
$$x_b = \xi_b + \delta_b$$

⋮

$$x_q = \xi_q + \delta_q$$

where the ϵ ' and δ 's are distributed independently of the η 's and ξ 's. The y ' and x 's are unbiased estimates of the η 's and ξ 's since the expected values of ϵ 's and δ 's are zero (i.e. the errors are all assumed to have means of zero).

By device due to Wright, we represent causal pathways by single headed arrows connecting cause (tail) to corresponding effect (head).



The path diagram may be interpreted in terms of the structural equations. The variable at the head of one or more arrows is interpreted as being a function of just those variables at these same arrows.

Assuming linearity (important) then, upon inspection write down the structural equations which are implied by any path diagram. For the above example we have:-

$$\eta_1 = \alpha_1 + \alpha_{1a} \xi_a + \alpha_{1b} \xi_b$$

$$\eta_2 = \alpha_2 + \alpha_{21} \eta_1$$

as the structural equations.

α_{1a}, α_{1b} and α_{21} are referred to as "path coefficients".

It is often assumed errors in x 's are negligible such that

$$x_i = \xi_i \quad \text{for all } i$$

and errors in y 's are additive such that

$$y_i = \eta_i + \varepsilon_i \quad \text{for all } i$$

so for our example the model would be

$$y_1 = \alpha_1 + \alpha_{1a}x_a + \alpha_{1b}x_b + \varepsilon_1$$

$$y_2 = \alpha_2 + \alpha_{21}(y_1 - \varepsilon_1) + \varepsilon_2$$

If η_1 is eliminated from the second structural equation by substitution of the right member of the first structural equation we get a modified pair of equations which is referred to as the reduced structural equations.

Thus the structural equation in our case are:-

$$\eta_1 = \alpha_1 + \alpha_{1a}\xi_a + \alpha_{1b}\xi_b$$

$$\eta_2 = \alpha_2 + \alpha_{21}\eta_1$$

and the reduced structural equations are:-

$$\eta_1 = \alpha_1 + \alpha_{1a}\xi_a + \alpha_{1b}\xi_b$$

$$\eta_2 = \alpha_2 + \alpha_1\alpha_{21} + \alpha_{1a}\alpha_{21}\xi_a + \alpha_{1b}\alpha_{21}\xi_b$$

Substituting x's for ξ 's and y for η 's on the reduced structural equations and attaching the additive errors as below we get reduced regression equation:-

$$y_1 = \alpha_1 + \alpha_{1a}x_a + \alpha_{1b}x_b + \varepsilon_1$$

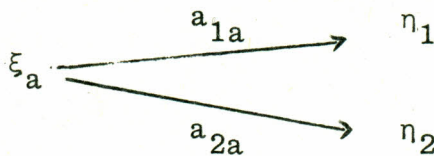
$$y_2 = (\alpha_2 + \alpha_{1a}\alpha_{21}) + \alpha_{1a}\alpha_{21}x_a + \alpha_{1b}\alpha_{21}x_b + \varepsilon_2$$

$$= \alpha_3 + \alpha'_a x_a + \alpha'_b x_b + \varepsilon_2$$

The example discussed here is overidentified. 'overidentification' occurs when there are more regression coefficients. In our example there are 3 path coefficients and 4 regression ($\alpha_{1a}, \alpha_{1b}, \alpha'_a$ and α'_b). 'Underidentification' can also occur as well as "Just identification". Just identification occurs when there is same number of path coefficients as there are simple or partial regression coefficient (a necessary but not sufficient condition for just identification).

A set of structural equations is called the causal or structural model. η 's or y are usually referred to as endogenous variables and ξ 's or x's as exogenous variables.

The path diagram



give an example of simultaneous models. The structural equations are:-

$$\eta_1 = \alpha_1 + \alpha_{1a} \xi_a$$

$$\eta_2 = \alpha_2 + \alpha_{2a} \xi_a$$

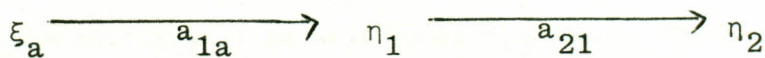
and the model equations are

$$Y_1 = \alpha_1 + \alpha_{1a} X_a + \epsilon_1$$

$$Y_2 = \alpha_2 + \alpha_{2a} X_a + \epsilon_2$$

the errors are independently distributed and in effect the maximum likelihood or least squares estimators can be found separately for each equation.

If a primary factor determines an effect and if this effect in turn determines still another effect, we speak of the network as being a chain regression. The path diagram of the simplest example is:-



The reduced structural equations would be:-

$$\eta_1 = \alpha_1 + \alpha_{1a} \xi_a$$

$$\eta_2 = \alpha_2 + \alpha_1 \alpha_{21} + \alpha_{1a} \alpha_{21} \xi_a$$

Reduced regression equations are then:-

$$Y_1 = \alpha_1 + \alpha_{1a} X_a + \epsilon_1$$

$$Y_2 = (\alpha_2 + \alpha_1 \alpha_{21}) + \alpha_{1a} \alpha_{21} X_a + \epsilon_2$$

$$= \alpha_3 + \alpha_2 X_a + \epsilon_2$$

and regression coefficients can be found using the usual procedures (least square method, maximum likelihood method etc.)

It is rare, indeed that theory specifies all causes of a variable. Therefore, usually another unspecified of the endogenous variables.. This residual term is often called the disturbance or error or stochastic term. The error or disturbance (here symbolised as ϵ or δ) represents the effect of causes that are not specified. Some of these unknown causes may be potentially specifiable while others may be essentially unknowable.

We define a matrix A

$$A = \begin{bmatrix} a_{11} & \dots & a_{1c} \\ \vdots & & \\ a_{r1} & \dots & a_{rc} \end{bmatrix}$$

as a rectangular ordered array of the elements. The general term of the matrix will be written as a_{ij} where the first subscript will always refer to the i th row, and the second to the j th column. A matrix with r rows and c columns will be referred to as of order $r \times c$ (read as r by c).

A vector is a matrix with a single row or column.

n -component column vector will be written as:-

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

a row vector x' (transpose of x) is

$$x' = \text{transpose of } \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [x_1 \dots x_n]$$

we can obtain also transpose of any matrix. In our case the transpose of A written as A' is

$$A' = \begin{bmatrix} a_{11} & \dots & a_{r1} \\ \vdots & \ddots & \vdots \\ a_{1c} & \dots & a_{rc} \end{bmatrix}$$

If a matrix is square (i.e number of rows is equal to the number of columns, $r = c$) and equal to its transpose, it is said to be symmetric. The elements a_{ii} of a square matrix occupy what are called the main diagonal positions. The sum of these diagonal elements is called the trace of matrix. In our case the trace of A ($\text{tr } A$) would be denoted by:-

$$\text{tr } A = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii}$$

Some matrices of interest that are discussed or implied are:-

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix}$$

which is identify matrix. It is a square matrix with one in each main diagonal position and zeros elsewhere. The $p \times p$ triangular matrix has pattern

$$T = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1p} \\ 0 & t_{22} & \dots & t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_p \end{bmatrix}$$

i.e. below the main diagonal of T is all zeros. T is the triangular matrix.

The $r \times c$ null matrix or else called zero matrix has zero entries or elements everywhere on the matrix:-

$$O = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

The inverse of the square matrix A is that unique matrix A^{-1} with elements or entries such that

$$A A^{-1} = A^{-1} A = I$$

(analogue of scalar division e.g.
 $5 \times \frac{1}{5} = \frac{1}{5} \times 5 = 1$)

A quadratic form in the variables x_1, \dots, x_n is an expression of the type

$$\begin{aligned} f(x_1, \dots, x_n) &= a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{nn}x_n^2 \\ &\quad + 2a_{12}x_1x_2 + \dots + 2a_{1n}x_1x_n + \dots \\ &\quad \quad \quad + 2a_{(n-1)n}x_{n-1}x_n \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \end{aligned}$$

where $a_{ji} = a_{ij}$. Some of the a_{ij} may be zero.

Quadratic form can be written in matrix notion as:-

where $X' = \begin{matrix} X' A X \\ x_1, x_2, \dots, x_n \end{matrix}$

and A is the $n \times n$ symmetric matrix of the coefficients.

ie. $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{nn} \end{bmatrix}$

A symmetric matrix A and its associated quadratic form are called positive definite if $X' A X > 0$ for all non null X . If $X' A X \geq 0$, the form and its matrix are called positive semidefinite.

Associated with every square matrix is a unique scalar number called determinant. The determinant of A will be written as $|A|$

The eigenvalues* (also called characteristic roots) of the $p \times p$ matrix A are the solutions to the determinantal equation:-

$$|A - \lambda I| = 0$$

The determinant is a p th degree polynomial in λ and thus A has just p characteristic roots.

The following are the results which can be obtained from mathematical theory of polynomial equations:-

1. The sum of the eigenvalues of A is equal to the trace of A i.e.

$$\sum_{i=1}^p \lambda_i = \text{tr. } A$$

2. The product of the eigenvalues of A is equal to the determinant of A i.e.

$$\prod_{i=1}^p \lambda_i = |A|$$

The following are properties of eigenvalues:-

1. The eigenvalues of a symmetric matrix with real elements are all real.
2. The eigenvalues of a positive definite matrix are all positive.
3. The eigenvalues of a diagonal matrix are the diagonal elements themselves.

Associated with every eigenvalue λ_i of the square matrix is a eigenvector (also called characteristic vector) X_i whose elements satisfy the homogenous system of equations

$$[A - \lambda_i I] X_i = 0$$

Denoting X a n - dimensional random vector, $X = (x_1, \dots, x_n)'$ is said to have a n -variate normal distribution, if its probability density function (p.d.f)⁴ can be written as:-

$$(2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (X-\mu)' \Sigma^{-1} (X-\mu) \right\}$$

where $\mu = (\mu_1, \dots, \mu_n)$

4 p.d.f. may be seen as frequency distribution of which N is very large and the intervals are very small.

and Σ is a positive definite symmetric matrix of dimension $n \times n$.

Thus a random vector X is said to have a n -variate normal (or n -dimensional multinormal or multivariate normal) distribution with mean vector μ and variance covariance matrix Σ if its p.d.f. is as given above.

We write $X \sim N_n(\mu, \Sigma)$ which is a way of symbolising the above. It is read as random vector X has a multivariate normal distribution with mean vector μ and covariance matrix Σ .

The mean is the same as expected Value or Expectation. Thus denoting mean as \bar{X} we have:-

$$\bar{X} = E(X) \text{ where } X \text{ is a random vector.}$$

The symbol E denotes the operation of computing the expected value, and is called the expectation operator.

Let X and Y be q and p dimensions random variables (vectors) then the covariance matrix of X and Y is

$$C(X, Y) = (\text{covariance } (x_i, Y_j))$$

$p \times q$ matrix

The covariance matrix of X (X with it self) denoted as

$$D(X) \text{ is equal to } C(X, X)$$

i.e

$$D(X) = C(X, X)$$

$q \times q$ matrix

- ① Problem not clear
- ② Theoretical framework OK
- ③ Chapter left hanging.

CHAPTER II

REVIEW OF RELATED LITERATURE

The review is divided into two Sections. Section One deals with topics and works related to Causality i.e. path analysis while Section Two deals with some works related to predictive validity.

SECTION ONE

REVIEW OF LITERATURE ON CAUSALITY

Path analysis, due to Wright (1918), is a technique sometimes used to assess the direct causal contribution of one variable to another in a nonexperimental situation. The problem, in general, is that of estimating the parameters of a set of linear structural equations, representing the cause and effect relationships hypothesized by the investigator.

Such a system of equations involves variates of two kinds: Independent variables or cause variables or exogenous variables, ξ_1, \dots, ξ_m and dependent variables or effect variables or endogenous variables, η_1, \dots, η_n

The technique consists of solving the structural equations for the dependent variables in terms of the independent to obtain the reduced form of the equations and then estimating the regression of the dependent variables on the independent from this reduced form. There seem to be 3 types of difficulties associated with these techniques, namely:-

1. If there are errors of measure in the independent variates, these errors will give rise to inconsistent regression estimates.

How do you identify ind. Var.?

2. If the parameters in the structural equations and the parameters in the reduced form are not in a one-to-one correspondence, one or more of the parameters of structural equations may be over-identified or underidentified.
3. Since the regression technique is applied to each equation separately, one does not get an overall test of the entire causal structure.

These 3 difficulties may be eliminated in some models if analyses is done by means of a covariance structure approach. I now discuss what is covariance structure approach (Joreskog, 1970):-

Consider a data matrix $\chi = \{\chi_{\alpha_i}\}$ of N observations on p response variables and the following model. Rows of \bar{X} are independently distributed each having a multivariate normal distribution with the same variance covariance (usually referred to as covariance) matrix Σ of form:-

$$\bar{X} = \beta(\Lambda \Phi \Lambda + \Psi^2)\beta + \theta^2 \quad (1)$$

and mean vector given by

$$E(\bar{X}) = A E \beta$$

where

$A = \{\alpha_{\alpha_s}\}$ is a $N \times g$ matrix of rank⁵ g and

$P = \{\rho_{ti}\}$ is a $h \times p$ matrix of rank⁵ h both

being fixed matrices with $g < N$ and $h < p$. The following are parameter matrices:-

$$E = \{\epsilon_{st}\}$$

$$\beta = \{\beta_{ik}\}$$

$$\Lambda = \{\lambda_{km}\}$$

$$\Phi = \{\phi_{mn}\} \quad \text{which is symmetric matrix}$$

$$\Psi = \{\delta_{kl} \psi_k\} \quad \text{diagonal matrix}$$

and $\theta = \{\delta_{ij} \theta_i\}$

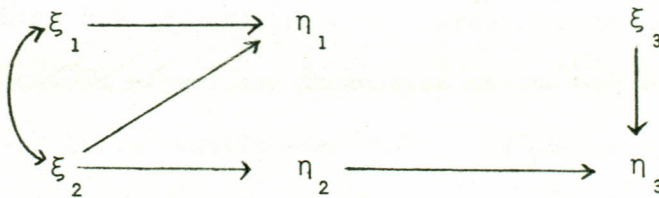
5 The rank of an $m \times n$ matrix A , is the number of columns (or rows) of the matrix that cannot be exactly obtained as linear combinations of other columns (or rows). The rank of a matrix can never exceed the smaller of its two dimension i.e. rank of $A < \min(m, n)$.

Thus the covariance structure approach (general model) is one where means, variances and covariances are structured in terms of other sets of parameters that are to be estimated. Any one of the parameters in $\bar{\xi}$, $\bar{\beta}$, $\bar{\Lambda}$, Φ , Ψ and Θ will be assumed known prior and for one or more subsets of remaining parameters to have identical but unknown values.

Thus parameters are of 3 kinds:-

- (i) fixed parameters that have been assigned given values (completely specified).
- (ii) constrained parameters that are unknown but equal to one or more other parameters (partially specified).
- (iii) free parameters that are unknown and not constrained to be equal to any other parameter (unspecified).

Below here I will take a path analysis model and try to show it can be expressed in form of (1) consequently indicating covariance structure approach is feasible with this model below:-



Note: double headed arrow means the two variables at the heads are correlated thus in our case ξ_1 and ξ_2 are correlated.

structural equation for this path diagram is

$$\eta_1 = \gamma_{11} \xi_1 + \gamma_{12} \xi_2$$

$$\eta_2 = \alpha_{21} \eta_1 + \gamma_{22} \xi_2$$

$$\eta_3 = \alpha_{32} \eta_2 + \gamma_{33} \xi_3$$

There are six parameters $\gamma_{11}, \gamma_{12}, \alpha_{21}, \gamma_{22}, \alpha_{32}$, and γ_{33} in these equations to be estimated.

The reduced form equations are:-

$$\eta_1 = \gamma_{11} \xi_1 + \gamma_{12} \xi_2 + 0 \xi_3$$

$$\eta_2 = \alpha_{21} \gamma_{11} \xi_1 + (\alpha_{21} \gamma_{12} + \gamma_{22}) \xi_2 + 0 \xi_3$$

$$\eta_3 = \alpha_{32} \alpha_{21} \gamma_{11} \xi_1 + \alpha_{32} (\alpha_{21} \gamma_{12} + \gamma_{22}) \xi_2 + \gamma_{33} \xi_3$$

There are 9 regression coefficients in these equations, two of which are constrained to be zero, the remaining seven being functions of the six parameters of the structural equations. It is readily seen that α_{32} is overdetermined. If we estimate the regression coefficients in the reduced form equations by the regression method, assuming ξ_1, ξ_2, ξ_3 without

error, there is no guarantee that the estimates

$$\hat{\beta}_{31} | \hat{\beta}_{21}$$

and $\hat{\beta}_{32} | \hat{\beta}_{22}$ of α_{32} are the same.

Defining $\mu = \alpha_{21}$, $v = \alpha_{32} \alpha_{21}$, $\delta = \gamma_{11}$, $\varepsilon = \gamma_{12}$

$$\phi = \gamma_{12} + (\gamma_{22} | \alpha_{21}) \text{ and}$$

$$\kappa = \gamma_{33} | \alpha_{32} \alpha_{21}$$

we may easily verify that there is one-one transformation between

$(\alpha_{21}, \alpha_{32}, \gamma_{11}, \gamma_{22}, \gamma_{33})$ and $(\mu, v, \delta, \varepsilon, \phi, \kappa)$ and

$$\text{that } \alpha_{21} = \mu$$

$$\gamma_{12} = \varepsilon$$

$$\alpha_{32} = \frac{v}{\mu}$$

$$\gamma_{22} = \mu(\phi - \varepsilon)$$

$$\gamma_{11} = \delta$$

$$\gamma_{33} = v\kappa$$

let

$$\beta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 & 0 \\ 0 & 0 & v & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \delta & 0 & 0 & 0 & 1 \end{bmatrix}$$

(m)

$$\Lambda \begin{bmatrix} \delta & \epsilon & 0 \\ \delta & \phi & 0 \\ \delta & \phi & \kappa \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (M)$$

Now suppose further that $\eta_1, \eta_2, \eta_3, \xi_1, \xi_2$ and ξ_3 cannot be observed directly but instead

$$y_i = v_i + \eta_i + \epsilon_i \text{ and } x_i = \mu_i + \xi_i + \delta_i \quad (i = 1, 2, 3) \quad (N)$$

are observed, where v_i and μ_i are the means of y_i and x_i respectively and ϵ_i and δ_i are errors of measurement. Assuming that errors of measurement are uncorrelated with the true variates and among themselves (classical measurement theory), we find from (M) and (N) that the covariance matrix of y_1, y_2, y_3, x_1, x_2 and x_3 is

$$\Sigma = \beta \Lambda \Phi \Lambda' \beta' + \Theta^2$$

where Φ is the covariance matrix ξ_1, ξ_2, ξ_3

i.e. $C(\xi_i, \xi_j)$

and Θ^2 is diagonal matrix containing six error variances.

This is of form (1) with $\Psi = 0$ and with certain elements of Λ constrained to be equal as in (M). Thus

indicating covariance structure approach is feasible with causal or structural models.

The sample variance - covariance (or simple covariance) matrix S of Y_1, Y_2, Y_3, X_1, X_2 and X_3 provide information for estimating the free parameters β, Λ, Φ and Θ . In large samples the χ^2 statistics may be used to test the overall goodness of fit of the causal structure. In our case there are 6 parameters in β and Λ , 6 in Φ and 6 in Θ to be estimated, so that degrees of freedom are $21 - 18 = 3$ in our case.

Estimation of the Parameters and Testing of the Model:-

To determine the estimates of the unknown parameters, two different methods of fitting the model to the observed data may be used. One is the generalized least squares method (GLS) that minimizes

$$G = \text{tr} (I - S^{-1} Z)^2$$

the other is the maximum - likelihood method (ML) that minimizes

$$M = \log |Z| + \text{tr}(S Z^{-1}) - \log |S| - p$$

where p is the total number of variables both endogenous and exogenous variables.

The sample covariance matrix \mathbf{s} of y 's (endogenous variables) and $\mathbf{x}'\mathbf{s}$ (exogenous variables) provide information for estimating the parameters. The method employed for the estimation of the parameters is most commonly maximum likelihood method (ML) and it is the one discussed here. For the discussion below we symbolise both endogenous and exogenous variables by Y . This Y will be

$$Y = \begin{bmatrix} y_{11} & \dots & y_{1p} \\ \vdots & & \vdots \\ y_{N1} & & y_{Np} \end{bmatrix} = \begin{bmatrix} y'_1 \\ \vdots \\ y'_N \end{bmatrix}$$

This denotes all data for each of the three sample to be considered.

The likelihood of the observations, Y ,

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$\text{is } L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}NP} |\Sigma|^{\frac{1}{2}N}} \exp \left[-\frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i - \mu)' \Sigma^{-1} (\mathbf{y}_i - \mu) \right]$$

If we introduce the sample mean vector

$$\begin{aligned} \bar{y} &= \frac{1}{N} \sum_{h=1}^N y_h \\ &= \hat{\mu} \end{aligned}$$

and matrix of sums of squares and cross products

$$\begin{aligned} \Lambda &= \sum_{h=1}^N (y_h - \bar{y})(y_h - \bar{y}) \\ &= \sum_{h=1}^N y_h y_h - N\bar{y}\bar{y} \end{aligned}$$

the natural logarithm of the likelihood can be written as

$$\begin{aligned} \ell(\mu, \Sigma) &= \ln L(\mu, \Sigma) \\ &= -\frac{1}{2} N \ln(2\pi) - \frac{1}{2} N \ln |\Sigma| - \frac{1}{2} \text{tr} \Sigma^{-1} \\ &\quad - \frac{1}{2} N (\mathbf{y} - \mu)' \Sigma^{-1} (\mathbf{y} - \mu) \end{aligned}$$

since Σ^{-1} is a symmetric positive definite matrix, the quadratic form term will be a minimum only when μ is equal to \mathbf{y} , so that the maximum likelihood estimate of the mean vector is $\bar{\mathbf{y}}$.

The estimate of the covariance matrix is more easily found by computing the maximum likelihood estimate of Σ^{-1} and applying the below useful result:-

Let the distribution function of a random variable (univariate or vector valued) depend upon the parameters.

$$\begin{aligned} \theta_1 &= g_1(\lambda_1, \dots, \lambda_k) \\ \vdots & \\ \theta_k &= g_k(\lambda_1, \dots, \lambda_k) \end{aligned}$$

which are in turn unique functions of other parameters $\lambda_1, \dots, \lambda_k$.
 The transformation from the θ_i set to the λ_i is one to one,
 with inverse functions.

$$\lambda_1 = h_1(\theta_1, \dots, \theta_k)$$

$$\lambda_k = h_k(\theta_1, \dots, \theta_k)$$

then, if the maximum likelihood estimates of the θ_i are $\hat{\theta}_1, \dots, \hat{\theta}_k$, those of λ_i are given by $\lambda_i = h_i(\hat{\theta}_1, \dots, \hat{\theta}_k)$ $i=1, \dots, k$

In our context, $k = \frac{1}{2} p(p+1)$, the θ_i are the elements θ_{ij} of Σ^{-1} and the λ_i are those of Σ . To find the maximum likelihood estimate of $\theta \equiv \Sigma^{-1}$ we compute the derivatives

$$\frac{\delta l(\mu, \theta)}{\delta \theta_{ii}} = \frac{1}{2} N \theta^{ii} - \frac{1}{2} a_{ii}$$

$$\frac{\delta l(\mu, \theta)}{\delta \theta_{ij}} = N \theta^{ij} - a_{ij} \quad i < j$$

where θ^{ii} , θ^{ij} are the elements of θ^{-1}

Hence, the estimate is

$$\hat{\theta} = \left(\frac{1}{N} \Lambda \right)^{-1}$$

Thus the maximum likelihood estimate of β is

$$\hat{\beta} = \frac{1}{N} \Lambda \quad \text{by above result.}$$

It is easily shown by an application of the expectation operator that $\hat{\beta}$ is biased estimate. The unbiased estimate is

$$S = \frac{1}{N-1} \Lambda$$

Thus though Σ may not be known its estimates $\hat{\Sigma} = \frac{1}{N} \Lambda$ can be used.

$$\text{Here } \Lambda = \sum_{h=1}^N \mathbf{y}_h \mathbf{y}_h' - N \bar{\mathbf{y}} \bar{\mathbf{y}}' \quad \text{as seen above.}$$

The statistics used to test the fit of the model is a function of M and is given by:-

$$f = N M = N \{ \log |\hat{\Sigma}| + \text{tr}(S \hat{\Sigma}^{-1}) - \log |S| - p \}$$

which is asymptotically distributed as a chisquare variable with $(p/2)(p+1) - m$ degrees of freedom. Where m is the number of parameters to be estimated in the model.

The model once hypothesized, it has to be tested to find whether it fits the data. The following is an illustrative example for the model as carried out by Sorbom (1975). At every stage in search of the model, the model has to be examined and perhaps reformulated to result in a model with more interpretable parameters. Since a nearly similar analysis except for search of other models was done in the study consequently the study would be reviewed in great details:-

The objective of the search procedure was to detect covariation between errors (measurement errors or residuals). Sorbom supposed that he wanted to study the change in ability between two occasions for a group of individuals. He let ξ denote the random variable representing ability at the first occasion for a random sample from the group, and η the ability for same individuals at the second occasion.

He supposed further the structural equation was:-

$$\eta = \alpha + \beta\xi + r \quad (1)$$

The above equation (1) describes the connection between final and initial status. The residual r is supposed to be uncorrelated with ξ . The main concern is the parameters α and β in above equation.

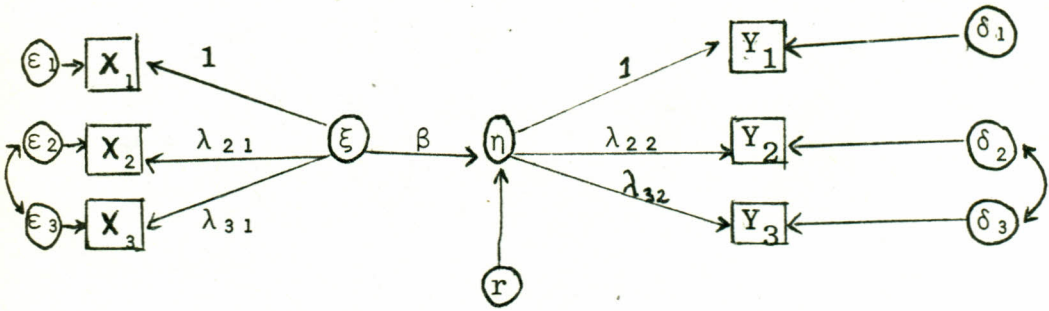
To estimate these parameters three tests were administered at each occasion. He let X denote the measurements at the first occasion and Y the measurements at the second occasion. The following model for the measurement was measured:-

$$\begin{array}{l}
 X_1 = \lambda_{11} \xi + \epsilon_1 \\
 X_2 = \lambda_{21} \xi + \epsilon_2 \\
 X_3 = \lambda_{31} \xi + \epsilon_3
 \end{array}
 \quad \text{and (2B)} \quad
 \begin{array}{l}
 Y_1 = \lambda_{12} \eta + \delta_1 \\
 Y_2 = \lambda_{22} \eta + \delta_2 \\
 Y_3 = \lambda_{32} \eta + \delta_3
 \end{array}$$

ϵ and δ are vectors of errors, supposed to be uncorrelated with ξ and η . It assumed also the variables in ϵ and δ are uncorrelated and that

$$E(\epsilon) = E(\delta) = 0$$

Sörbom also assumed that x_1 and y_1 represented very similar tests administered at the two occasions. He argued that there might have been covariation between x_1 and y_1 not only because of covariation between the abilities at the two occasions, ξ and η but also because of incidental features arising from the construction of the tests or because of memory effects. Thus it might not be true that $\text{cov}(\epsilon_1, \delta_1) = 0$. He further assumed that the tests x_2 and x_3 measure some traits, which are additional but are uncorrelated with ξ . Thus covariation between x_3 and x_2 for a given value of ξ would be expected not to be zero and also $\text{cov}(\epsilon_2, \epsilon_3) \neq 0$. Assuming that the errors for Y_2 and Y_3 were correlated too, he described the measurement model as in the figure below



The true model. Circles denote unobservable variables and squares denote observable variables.

The parameters λ_{11} , λ_{21} , λ_{31} and σ_{ξ}^2 (variance of ξ) in (2A) cannot be identified simultaneously. It is observed that, for example, multiplying ξ by a constant and dividing λ_{11} , λ_{21} and λ_{31} by the same constant does not change the measurements of X . Thus it is possible to fix at least one of these parameters. Here the choice made was $\lambda_{11} = 1$ and using the same argument the other choice was $\lambda_{12} = 1$.

Let

$$\theta = E \begin{bmatrix} \xi \\ \eta \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{bmatrix}$$

$$\Phi = E \left[\begin{array}{c} (\xi - \theta_1) \quad (\xi - \theta_1 \quad \eta - \theta_2) \\ (\eta - \theta_2) \end{array} \right]$$

$$= \left[\begin{array}{c} \sigma_{\xi}^2 \\ \sigma_{\xi\eta} \\ \sigma_{\eta}^2 \end{array} \right]$$

$$\Psi = E \left[\begin{array}{c} (\epsilon) \\ (\delta) \end{array} (\epsilon, \delta) \right] = \left[\begin{array}{c} \sigma_{\epsilon_1}^2 \\ \sigma_{\epsilon_2}^2 \\ \sigma_{\epsilon_3}^2 \\ \sigma_{\delta_1}^2 \\ \sigma_{\delta_2}^2 \\ \sigma_{\delta_3}^2 \\ \sigma_{\epsilon_1\delta_1} \\ \sigma_{\epsilon_2\delta_2} \\ \sigma_{\epsilon_3\delta_3} \\ \sigma_{\delta_1\delta_2} \\ \sigma_{\delta_2\delta_3} \\ \sigma_{\epsilon_1\delta_2} \\ \sigma_{\epsilon_2\delta_1} \\ \sigma_{\epsilon_1\delta_3} \\ \sigma_{\epsilon_3\delta_1} \\ \sigma_{\epsilon_2\delta_3} \\ \sigma_{\epsilon_3\delta_2} \end{array} \right] \quad (3)$$

Sörbom considered the following parameter values for a hypothetical population:

$$\Lambda_0 = \left[\begin{array}{cc} 1.0 & 0.0 \\ 0.8 & 0.0 \\ 0.5 & 0.0 \\ 0.0 & 1.0 \\ 0.0 & 1.0 \\ 0.0 & .9 \\ 0.0 & .7 \end{array} \right] \quad \Phi_0 = \left[\begin{array}{cc} 16.0 & \\ 12.0 & 25.0 \end{array} \right]$$

$$\theta_0 = \left[\begin{array}{c} 10 \\ 20 \end{array} \right]$$

$$\sigma_{\epsilon_0}^2 = \begin{bmatrix} 4.00 \\ 2.56 \\ 1.00 \end{bmatrix} \quad \sigma_{\delta_0}^2 = \begin{bmatrix} 6.25 \\ 5.0625 \\ 3.0625 \end{bmatrix}$$

These parameters were chosen so that the reliability of each test equalled 0.8. This meant for example, that the variance of ϵ_2 is $\lambda_{21}^2 \sigma_{\xi}^2 (1 - 0.8) / 0.8 = 2.56$

It was further assumed that the population correlation between ϵ_1 and δ_1 , $\rho_{\epsilon_1 \delta_1} = 0.5$

and that $\rho_{\epsilon_2 \epsilon_3} = 0.3$ and $\rho_{\delta_2 \delta_3} = 0.4$

From the path diagram above (in this section) there is:-

$$\eta = \beta \xi$$

$$Y_1 = \eta + \delta_1$$

$$x_1 = \xi + \epsilon_1$$

$$Y_2 = \lambda_{22} \eta + \delta_2$$

$$x_2 = \lambda_{21} \xi + \epsilon_2$$

$$Y_3 = \lambda_{32} \eta + \delta_3$$

$$x_3 = \lambda_{31} \xi + \epsilon_3$$

Thus

$$Y_1 = \beta \xi + \delta_1$$

$$Y_2 = \lambda_{22} \beta \xi + \delta_2$$

$$Y_3 = \lambda_{32} \beta \xi + \delta_3$$

Thus

$$C(Y, X) = \mathbf{\Gamma} \text{ is}$$

$$\beta^2 \sigma_\xi^2 + \sigma_{\delta 1}^2$$

$$\lambda_{22} \beta^2 \sigma_\xi^2 \quad \lambda_{22}^2 \beta^2 \sigma_\xi^2 + \sigma_{\delta 2}^2$$

$$\lambda_{32} \beta \sigma_\xi^2 \quad \lambda_{32} \lambda_{22} \beta^2 \sigma_\xi^2 \quad \lambda_{32}^2 \beta^2 \sigma_\xi^2 + \sigma_{\delta 3}^2$$

$$\beta \sigma_\xi^2 \quad \lambda_{22} \beta \sigma_\xi^2 \quad \lambda_{32} \beta \sigma_\xi^2 \quad \sigma_\xi^2 + \sigma_{\epsilon 1}^2$$

$$\beta \lambda_{21} \sigma_\xi^2 \quad \lambda_{22} \lambda_{21} \beta \sigma_\xi^2 \quad \lambda_{32} \beta \lambda_{21} \sigma_\xi^2 \quad \lambda_{21} \sigma_\xi^2 \quad \lambda_{21}^2 \sigma_\xi^2 + \sigma_{\epsilon 2}^2$$

$$\beta \lambda_{31} \sigma_\xi^2 \quad \lambda_{22} \lambda_{31} \beta \sigma_\xi^2 \quad \lambda_{31} \lambda_{32} \beta \sigma_\xi^2 \quad \lambda_{31} \sigma_\xi^2 \quad \lambda_{31} \lambda_{21} \sigma_\xi^2 \quad \lambda_{31}^2 \sigma_\xi^2 + \sigma_{\epsilon 3}^2$$

$$\frac{\lambda_{21} \sigma_\xi^2}{\sigma_\xi^2 + \sigma_{\epsilon 1}^2} = \lambda_{21} \frac{\sigma_\xi^2}{\sigma_\xi^2 + \sigma_{\epsilon 1}^2} = \lambda_{21} \rho_{x_1 x_1'}$$

where $\rho_{x_1 x_1'}$ is the reliability of
test x_1

The parameters of (1) are obtained by

$$\beta = \frac{\sigma_\xi \eta}{\sigma_\xi^2} = \frac{12}{16} = 0.75$$

$$\text{We has also } \theta_0 = \begin{pmatrix} 10 \\ 20 \end{pmatrix} = E \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

$$\text{Thus } E(\xi) = 10 \text{ and } E(\eta) = 20$$

$$(1) \text{ has form } \eta = \alpha + \beta \xi + r$$

$$\text{Thus } E(\eta) = \alpha + \beta E(\xi) + 0$$

$$\Rightarrow \alpha = 20 - \beta \times 10$$

$$= 20 - 0.75 \times 10 = 12.5$$

Thus $\beta = 0.75$ and $\alpha = 12.5$

and the mean values of the observed measurement are given by

$$E(X) = \begin{bmatrix} \theta_1 \\ \lambda_{21}\theta_1 \\ \lambda_{31}\theta_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ 5 \end{bmatrix}$$

and

$$E(Y) = \begin{bmatrix} \theta_2 \\ \lambda_{22}\theta_2 \\ \lambda_{32}\theta_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 18 \\ 14 \end{bmatrix}$$

The population variance - covariance matrix (dispersion matrix) for observed measurement

$C(X, Y) = \Sigma_0$ (subscript, 0, indicates we are referring to the population dispersion matrix).

$$\begin{aligned} \Sigma_0 &= E \left[\begin{pmatrix} X - E(X) \\ Y - E(Y) \end{pmatrix} \begin{pmatrix} X - E(X), Y - E(Y) \end{pmatrix} \right] \\ &= E \left[\Lambda \begin{pmatrix} \xi \\ \eta \end{pmatrix} - \Lambda \theta + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix} \right] \left[\Lambda \begin{pmatrix} \xi \\ \eta \end{pmatrix} - \Lambda \theta + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix} \right] \\ &= \Lambda \Phi \Lambda' + \Psi \end{aligned}$$

To illustrate the procedure Sörbom obtained a sample variance covariance matrix S and a vector of sample means $Z = (X, Y)$ by generating 1000 subjects normal variates according to a normal distribution with mean equal to $\Lambda_0 \Theta$ and variance covariance matrix equal to Σ_0 .

Thus it was as though for six tests he obtained independent measurements from a sample of 1000 individuals belonging to the population of individuals under study. Three of the tests were administered at a later occasion. At each occasion the tests were measuring approximately the same ability.

He was mainly concerned with development of this ability between those two occasions. He postulated that due to influence of memory effects and similar matters, the estimates of the parameters in (1) might be distorted. He presented a procedure which is not given here, on how to detect these distorting effects, and how to take them into account.

With the data from the above simulated study, he started with a model in which the errors are uncorrelated, that is

$$\Sigma = \Lambda \Phi \Lambda + \Psi$$

where Ψ is a diagonal matrix and Λ is structured as in (3). Maximum likelihood estimates of the parameters of the model

were obtained by minimizing

$$f = \log |\hat{\Phi}| + \text{tr}(\hat{\Phi}^{-1}S) - \log |S| - p$$

and a χ^2 measure of goodness of fit was obtained (as usual) as N times the minimum value of f. There was a computer program to do this.

A Precaution Note on Maximum likelihood Estimation

First, I will briefly say what maximum likelihood method is. The method of maximum likelihood estimation involves estimating population parameters such that they have the higher joint probability of yielding the sample data, given the constraints of the prior probability distributions of the variables. As applied to the common factor model or other models maximum likelihood estimators are chosen in the orthogonal case or otherwise to obtain the best estimate of the population common factor loading matrix. In other words the function M,

$$M = \log |\hat{\Phi}| + \text{tr}(S \hat{\Phi}^{-1}) - \log |S| - p$$

is minimized numerically with respect to the independent parameters using a modification of the iterative method of Fletcher and Powell. The minimization method makes use of the first order derivatives and large sample approximations to the elements or entries of the matrix of second order derivatives. The computer program LISREL VI is the newest version for obtaining the estimators of the model as well as testing the fit of it (the model); thus accomplishing the above.

Jackson and Chan (1980) had the following comments to make on maximum likelihood method as a procedure of estimating parameters:-

"One difficulty besetting all methods of finding maximum likelihood estimates is that the likelihood function, M , may not have any true maximum within the region for which all the uniquenesses are positive. An estimate of a population parameter of zero unique variance or negative unique variance appears to be unacceptable, both in terms of its inconsistency with the model and in terms of common sense expectations that error variance is virtually always present in some degree in empirical data. These so called improper solutions are commonly found. One interpretation of the improper solutions is that the model does not fit the empirical data, contention that should be tested. This interpretation does not seem to prevent some investigators from publishing their results anyway".

They (Jackson and Chan) had the following observations to make on the problems with interpreting significance tests in factor analysis, an observation which is true to a certain extent in all covariance structure approach:-

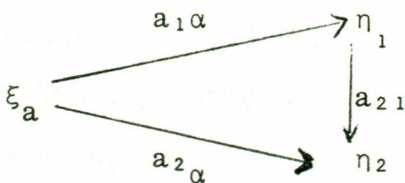
"Such statistical tests as significance tests can be useful and should be applied to check whether the model with a fixed number of factor is overfitted. They can, as can most statistical tests, gull a user into believing that something of importance in a practical or theoretical sense has been confirmed. At best, statistical significance is a minimum requirement, a necessary but not a sufficient condition for factor interpretability".

Thus in our case of Path analysis, the significance tests of the model can not be overemphasized. It should always be remembered that a statistical significance is a minimum requirement, a necessary but not a sufficient condition for interpreting that the data fits the model⁶.

6 In terms of hypothesis testing it means "failing to reject a null hypothesis" just tells us that the data gives the minimum requirement, a necessary but not a sufficient condition for the hypothesis to be true.

Relationship Between Factor Analysis and Path Analysis:-

Factor analysis may be thought of as the ultimate in under-identification, no regression information being available at all. A path diagram in which we have underidentification would look like:-



our maximum likelihood estimators would be

$$b_{1a} = a_{1a} \quad \text{and} \quad b_{2a} = a_{2a} + a_{1a} a_{21}$$

This presents the impossible task of solving for 3 quantities knowing only two relations.

Thus underidentification in general occurs when a set of observable variables whose true values are denoted by η_1, \dots, η_p are considered to be linear functions of some common factors $\xi_a, \xi_b, \dots, \xi_q$ where $p > q$. These common factors are either unknown or unmeasurable. The path coefficients $\alpha_{1a}, \alpha_{1b}, \dots, \alpha_{pq}$ are referred to as factor loadings.

The path coefficients may just as well be thought of as defining the factors in terms of the knowables. In any case,

the problem is to estimate the path coefficients or factor loadings where no information is provided by regression. This means that *a priori* restrictions must be supplied. There are infinite way of doing (restricting) that. In general, however, methods have been advocated which maximize the contribution of a single factor and then of a second factor, and so on until the least important is reached. The method of "Principal components" advocated by Hotelling⁽¹⁹³⁵⁾ adds to this ordering the restriction that the factors are orthogonal to one another.

Thus what we are saying is that under certain restrictions factor analysis and path analysis are the same. Factor analysis is a path analysis design in which the structural model is under-identified.

SECTION TWO

There are numerous studies which have been done on predictive validity of which previous examination performances have been used to predict future (or present) performance. Many of such studies have revealed that previous examination performances have tended to predict very well the future (or present) performance in both tertiary institutions or secondary institutions. In such study, usually the researchers carry out correlation analysis (simple correlation and/or multiple correlation analysis or the former one only) between the predictor variables (previous examination performance) and criterion variables.

One such study was carried out by Akeju and Michael (1970). The purpose of their study was to determine the degree of validity of several selection devices in the prediction of success of a sample of 109 college students at Federal School of Science, Lagos, Nigeria. The selection devices used were scores on six essay type achievement tests very much like 0-level and the success they were trying to predict was of General certificate of Education (Advance level) another achievement test done after two years of schooling. Both 0-level type examination and A-level were external examinations. The 0-level grades were used as predictor variables together with aptitude tests developed for African Culture; altogether there were nine predictor variables.

Using pearson product moment correlation coefficients among predictor and criterion variables and multiple regression equations involving the prediction of each criterion variable from optimally weighted composite of predictor variables as well as corresponding coefficients of multiple correlation, they carried out their prediction study and came up with the following findings:-

1. An achievement examination based on high school subjects (i.e. the 0-level examination) was the most valid predictor of subsequent success in the A-level. The Aptitude test was not a good predictor as the 0-level type examination.
2. It was observed that little was gained as far as prediction was concerned by taking the composite of the 0-level grades with the aptitude test. The 0-level was optimally adequate for prediction purposes. This was the kind of observation made when multiple regression analysis was carried out.

In this study a negative 'Hawthorne effect' is certainly very prevalent. Subjects knowing that aptitude tests and other internally administered achievement tests, the ones administered with A-level,

will not count for their selection or certification should have tended not to take them very seriously. This may contribute to the aptitude failing to predict very well the future performance.

In a more recent study Baldauf(Jr.) and Dawson (1980) with a sample of students from a wide variety of linguistic and tribal cultural backgrounds found that English language proficiency test was a very valid predictor of general academic attainment in a teachers' college in Papua, New Guinea. The sample they took was of students' teachers who had 10 years or 12 years of school and had English as a second language. In many studies English language has tended to be a very good predictor of future performance in tertiary institution. The studies do emphasize the importance of English language competence in tertiary institutions where the language is used as a media of instruction usually regardless on the area of speciality.

Their conclusion was based on correlation analysis between the scores of Michigan test of English Language proficiency and the overall Grade average point (GPA) for students who completed the teacher training.

Other studies which have found English language being the best single predictor are by Mukherje (1965) and by Hwang Kwa-Yann & Dizney (1970) among others. Mukherje using 87 freshmen,

I believe many of them had English as the first language, found that the test of vocabulary (English) turned out to be the best single predictor. He summarized his findings by the remarks, "As is well known vocabulary plays an important role in most college courses. It is no surprise therefore to find that students having a better vocabulary scored higher in the psychology examination". Hwang Kwa - Yann and Dizney (Hwang & Dizney, 1970) using a small sample of Chinese graduate students at an American University and examining correlations between the test of English as a foreign language (TOEFL) grades, grades of English as a second language (ESL) and Grade point average (GPA) made the following observation:-

Though their empirical evidence was limited, TOEFL was a relatively good predictor of grades in ESL for Chinese graduate students at University of Oregon. However, its use to predict the academic success of Chinese graduate students was doubtful.

Though English Language has shown to be a good predictor in most counts in courses where it is used as a medium of instructions, of course there is a big danger of overdepending on it for selection and placement. As it is obvious, it should be used together with other relevant predictor measures.

It is very important that if certain predictor measures are used for selection then they should predict the criterion measure. Otherwise if there is evidence to the contrary then there would be need to look at both the predictor and criterion measures so as

to establish whether they meet the required objectives. Thomas Goolsby (1970) looking at the Validity of the college level examination program (CLEP) for a number of College Sophomore Students had the following remarks to make as a result:-

The evidence in the study led no support to use of CLEP for selection, placement and advisement at the Sophomore level when it is considered alone and especially when grade point average (GPA) is a criterion.

He advised that there was need for a very substantial emphasis on adequate measurement and grading practices with the College. Thus both a vigorous determination and definition of curricular objectives and the construction of criterion measure in a continuing research program are necessary for higher education to meet its responsibilities for selection, placement and advisement.

Though Goolsby is right in suggesting that there is need to do research program on the examinations done in our institutions so as to be sure they met their required objective, it should be pointed out that this is not just necessary when predictor measures fail to predict criterion measures. It should be done even when predictor measures predict criterion measures, for high correlations are just necessary but not sufficient to make us certain that the required objectives are being met. The following study by Hewitt and Goldman (1976) illustrates this very well.

The two scholars on their study entitled "predicting the success of black, Chicano, Oriental and White College Students" concluded that there is danger that investigations like that one of theirs, may draw fallacious conclusions because both predictor measure and criterion measures may be biased in the same way. In such a circumstance, the predictor measures might appear valid and their use justified while reinforcing a vicious cycle of educational discrimination.

Thus we may conclude that though prediction studies are important and it is necessary for predictor measures to correlate with criterion measures and this need be so in particular when predictor measures are used for selection, placement or advisement, but this high correlation between predictor measures and criterion measures is not sufficient. There is need to do research program on the examinations done in our institutions so as to be certain they met their desired objectives.

It should be mentioned that estimates of previous school performance and maximum performance tests (intellective predictors) are not the only categories of possible predictors. There are what we refer to as non intellective predictors but as WatKin & Astilla (1980) and others have found, none has yet achieved adequate success. There are two categories of such non intellective predictors Drenth (1975):-

1. Personality variables, measured by means of personality tests, self-rating or observation scales. There is a vast literature and a great deal of empirical research to demonstrate that this category is not by any means negligible. Motivation, anxiety, interests, attitude and values, stability, adjustment and adaptation are examples of variables that have important influence on school performance. The measurement of such variables is highly complex not to mention the psychometric problems (validity and reliability) involved as well as the complicated way in which the variables cause and influence behaviour and performance.

How
Controlled
here?

2. The miscellaneous category of biographical information, antecedents, physical qualities, socio-economic factors, linguistic, racial, tribal background variables etc. Some of these would prove to have a very high validity in the psychometric senses, but a great many would at the same time be discriminators' in full sense of the word.

Since non-intellective predictors have failed to demonstrate adequate success as far as prediction of future academic performance is concerned then we are to be contented with previous school performance and maximum performance tests (intellective predictors) for selection, admission, placement or advisement. But as indicated there is always need to look critically at the

Who
says?

determination and definition of curricular objectives and the construction of criterion measures, at all levels i.e. at the implementation, evaluation and other levels. This should be done regardless of how well the criterion measure is predicted by the predictor measure. This critical research should not be restricted to the criterion measures. It need be done on predictor measures too.

Admission to next level in the educational system on the basis of previous school performance, in our case O-level can be justified by two statements of reasoning which at first glance might appear to be identical, though in reality are different:-

1. Education or training in a higher level of institution assumes a number of skills or a certain level of knowledge, or may assume both, acquired in the previous school. If the student has not acquired the assumed knowledge or skill, he cannot benefit from the subsequent learning experience. Here we are referring to the "prerequisites".
2. Previous school performance reveals qualities within the students that are also needed in next learning institution. The history of ones attainment is used as a predictor of future school performance since both are supposed to be dependent on the same "learning ability". Thus we expect them to be correlated highly. The probability of success depends on student performance, low performance means low

chance of success and high performance means high probability of success assuming everything else is constant. Here we are referring to the 'potentialities'.

The study attempted to find out how far O-level is meeting the above two obligations in recruiting professional teachers. In simple terms, the study is trying to find out how far we are justified in using O-level grades for selecting students to Pl courses in Primary Teachers Training Colleges. To accomplish this assignment correlation analyses of the predictor measures (O-level) with the criterion measures (performance in Primary Teachers' Colleges) were studied.

Some subjects done in O-level are weighted more than others and also researches elsewhere have found that, for instance, English Language Grades have tended to predict very well general future academic attainment (Baldauf & Sawson (1980) Mukherje (1985) etc.) where English is used as a media of instruction. With the great importance attached to such subjects as English Language and Mathematics, as well as the results of previous researches, it was found necessary to carry out a cause and effect (causality study) on the two aforementioned O-level subjects.

The following were the specific questions the study was attempting to answer:-

- (i) How well does O-level Grades predict the performance in Primary Teachers' Training Colleges ?

- (ii) Can we attribute any predictor measure (0-level grade) to a causal relationship with any criterion measure (performance in Primary Teachers' College final year examination)? i.e. can we say a good grade in 0-level for this measure ensures, causes or determines a good grade in this measure in Primary Teachers' College Examinations ?

The following were the hypotheses tested in an attempt to answer these questions:-

- (i) The 0-level Grades do not ^{significantly} predict performance in Primary Teachers' Training Colleges.
- (ii) There is no ^{significant} causal relationship between any 0-level grades with any Primary Teachers' Training Colleges final year examination performance.
-

C H A P T E R T H R E E

DESIGN OF THE STUDY

This Chapter tries to describe the samples used in the analysis, the variables used, as well as, in great details, the Statistical Methods used in the analysis namely:-

(i) Multiple correlation analysis

and(ii) Path analysis.

A further discussion on multiple correlation analysis appear in the Appendix B.

CHAPTER THREE

METHODOLOGY:-

For this study, the method adopted for selection of unit of analysis was a type of cluster sampling. Out of all the Primary Teachers' Colleges in Kenya the Colleges selected were Shanzu, Mosoriot, Kigari, Kilimambogo and Machakos. These Colleges were the ones treated as clusters and then as usual with one stage cluster sampling all the subjects in these clusters (colleges) were to be used for study. The cluster sampling was adopted as opposed to simple random sampling for convenience as well as for keeping the cost of obtaining data within reasonable limits. The sample obtained was quite large and I have reasons to believe was very representative of the population.

Observe because of some reasons, very common with this type of research it was not possible to get complete data in all the variables considered. For instance a small number of subjects in the college selected had been admitted with other types of examinations other than O-level and a few others even if they had offered O-level there were other abnormalities with their grades. The number of such cases and others not included in the sample were small and were considered to constitute a random sample and hence no bias would ensue in the final sample obtained. The sample studied consisted of a large sample of 1622 subjects from these five colleges. The subjects had been admitted in the colleges for

Pl course in 1984 and did their final college examination in 1986, that is after going through two years of training as Primary school teachers.

The O-level examinations of which are used for selection, are achievement tests done after four years of secondary education. The candidates offer 6 to 9 subjects each tested by one or more examination tests. The subjects range from social sciences (English, Kishwahili, Religious Education, Geography etc.) to natural sciences (Biology, Mathematics, Chemistry etc.) The candidates are classified into Division I (best) II, III or IV or Fail which is judged by their aggregate performance of their best six subjects which need come from certain laid groups of subjects. The examination are centrally and annually administered by the Examination Council to all secondary schools in Kenya. Some natural science subjects do have a practical component in them but more weight is on the achievement test. For the sample considered there were a few of them who had offered technical subjects.

It seems for the sample considered a good number of them had division three, this was approximately two thirds of the sample (several ways may be used to establish this, one is to consider or assume the distribution of the total aggregate of the best six subjects is approximately normal and then using knowledge of standard normal distribution one can get the approximate number of those who had division III or what have you). This was the method adopted. Other methods more manual and more mechanical

and more accurate may be used if need be .

The number of cases used from every one of the five colleges considered were as follows:-

<u>College</u>	<u>No of cases (or n)</u>
Shanzu	315
Mosoriot	213
Kigari	293
Machakos	372
Kilimambogo	<u>429</u>
TOTAL	1622

As indicated earlier all these were to graduate as Pl after passing the final college examination used as the dependent variable here. Thus one to be selected for a Pl course, a two year course, one has to have a division III in O-level or better and then one has to pass the final year college examination. For this sample considered, the subjects taken in the final year college examination were

- 1) Professional studies
- 2) Mathematics
- 3) Kiswahili
- 4) Mathematics
- 5) Science
- 6) Physical Education
- 7) Christian Education or Islamic Education
- 8) History or Geography
- 9) Art and Crafts or Music
- and 10) Agriculture or Home Science

Thus one had to take 10 subjects, the first six subjects listed above were compulsory while the candidate had a choice between studying Christian Religious Education or Islamic Religious Education; between studying History or Geography etc. The examination (or tests) are very much achievement test centrally administered and graded.

While we may be justified in generalizing our findings, this should be done with caution. This is because:-

Given the difficulties interaction creates for social scientists generalization is not absolutely justifiable. A social scientist modelling his work on physical science, thus aspiring to mass empirical generalization, to restructure them into more general laws, and to weld scattered laws into coherent theory finds it unpracticable.

Each research or experiment in social science is unique in its own right. Replication of experiments, or research even, is almost impossible if not absolutely impossible in social science. This is so because the unit of analysis in social science are usually human beings whose behavioural or performance is a result of hosts of other variables not to mention the problems we are faced with when we attempt to measure these psychological variables.

Explain
why
path
analyses

Thus when we give proper weight to local conditions, attention to whatever variables were controlled as well as to uncontrolled as well as to uncontrolled conditions, attention to personal characteristics, and to events that occurred during experimental treatment and measurement, any generalization is a working hypothesis, not a conclusion (Cronbach, 1975). Thus generalisation and conclusions made here should be seen in that context.

For the statistical analyses the following weighting system was adopted for the independent variable

A grade 9 (a fail) was weighted as 1,
 grade 8 (a pass) as 2
 grade 7 (a pass) as 3
 grade 6 (credit) as 4 etc.
 thus grade 1 (distinction) as 9.

The independent variables considered using the weighting system for each subject were:-

1. O-level English language grade (OENG)
2. O-level Mathematics grade (OMATH)
3. The aggregate of the best six grades the O-level (OAGG).

An assumption made here which is justifiable is the grading system for O-level did not vary appreciably from year to year.

Thus the O-level grades had consistent implication from year to year in the sense of the abilities of individuals they were suppose to represent. This was necessary because these subjects had done their O-levels not necessary in the same year. The years of doing O-level varied from 1982 and earlier; and for one to carry out the analyses the assumption was necessary.

The dependent variables were raw scores or standardized scores of the final year college examinations. The standardized scores were used where a combination of two or more subjects had to be done in the analyses otherwise raw scores were used where such combination was not done. The standardization was done on a scale of, mean 50 and standard deviation 10. Observe this kind of standardization is a linear transformation of raw scores and consequently unless combination of variables is done such kind of transformation would not alter the correlation coefficient (both multiple and simple) of the variables. This is why computer time was not wasted in standardizing raw scores for the variables which were not combined as such. Thus each subject (candidate) had 10 such dependent variables indicated earlier namely:-

- | | | |
|----|----------------------|--------|
| 1) | Professional studies | (PROF) |
| 2) | English | (ENG) |
| 3) | Kiswahili | (KISW) |
| 4) | Mathematics | (MATH) |
| 5) | Science | (SCIE) |
| 6) | Physical Education | (PHE) |

Explain
Use of
Z score
Raw score

- | | | |
|---------|--|-----------------|
| 7) | Christian Religious Education | (CRE) <u>or</u> |
| | Islamic Religious Education | (IRE) |
| 8) | History (HIST) <u>or</u> Geography | (GEOG) |
| 9) | Art and Crafts (ART) <u>or</u> Music | (MUSIC) |
| and 10) | Agriculture (AGRIC) <u>or</u> Home Science | (HOME) |

Using these dependent variables (or their combinations) and the three independent variables, a simple correlation analysis, multiple regression analysis and path analysis were done with the sample of 1622 subjects. one of the subsamples (sample for Kigari) was arbitrary chosen to carry out the fit of the structural models postulated in the path analyses study.

STATISTICAL MODELS AND DESIGNS

Two types of statistical data analyses methods used in the study are discussed. These are, multiple correlation (Regression) analysis, and path analysis:-

Multiple Correlation Analysis

The method of multiple correlation is employed to predict a single criterion variable from two or more predictor variables with the minimum amount of (squared) error. The amount of variance in the criterion variable can be determined empirically, and the multiple correlation technique is used to maximize the amount of the criterion's variance predicted, or accounted for, by the predictor variables. This is another way of saying that the technique seeks to maximize that amount of variance in the criterion variable which is unpredicted and hence ascribed to error in the immediate problem.

We denote the criterion variable as Y and the predictor variables by X_1, \dots, X_k . The score of person i on variable X_j will here be denoted as X_{ij} and Y_i for his/her score on criterion variable, Y . The multiple regression (correlation) model has the form:-

$$Y_i = \sum_{j=1}^k \beta_j X_{ij} + e_j \quad j = 1, 2, \dots, k$$

$$Y = X\beta + \epsilon$$

Where

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Note: Observe this model $Y_i = \sum_{j=1}^k \beta_j x_{ij} + e_j, j = 1, \dots, n$

is similar to the familiar linear regression model $y = \beta_0 + \beta_1 x_1 + \dots$

$\beta_k x_k + e$ only that the former model is on the standard form.

For this model $Y = X\beta + \varepsilon$ we have

Y as a random observed vector of $n \times 1$

ε as a random error vector of $n \times 1$

X is an $n \times k$ matrix of known fixed quantities

and β is a $k \times 1$ vector of unknown parameters.

The assumption for this model are:-

(i) the expected value of Y is $X\beta$

i.e. $E(Y) = X\beta$ this implies the expected value

of ε is zero vector

$$E(\varepsilon) = 0$$

and the variance of Y equal to σ^2

where σ^2 does not depend on the β or X

- (ii) ε_i is normally distributed with mean zero and variance σ^2 and that the ε_i are jointly independent.

The estimation of β (the vector β is a parameter) can be done using least square method or maximum likelihood method (other methods are also available for estimating such parameters). σ^2 which is also a parameter may be estimated by considering sum of squares for errors and degrees of freedom (i.e.

$$\sigma^2 = \frac{\varepsilon\varepsilon'}{n-(k+1)}$$

where $n - (k + 1)$ are degrees of freedom)

Thus

$$\hat{\sigma}^2 = \frac{1}{n - (k+1)} (Y - X \hat{\beta}) (Y - X \hat{\beta})'$$

The estimates of β is given by

$$\beta = (X'X)^{-1}X'Y$$

as it may be readily be shown

Usually X is of form

$$X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix}$$

as seen earlier.

Using the estimates of β we obtain the predicted criterion variable. This is

$$Y = X \beta$$

The correlation between observed Y and the predicted $Y_1(\hat{Y})$ is what is termed as multiple correlation coefficient, often denoted by R .

COMPUTATION OF R.

Once $\hat{\beta}$ has been obtained, $R^2 = \hat{\beta} K$. Where K is a vector and its elements are correlations between the predictor variables and the criterion variable.

The square root of the inner product of vector K and vector $\hat{\beta}$ gives the multiple correlation coefficient. This is one of the many methods that might be used to obtain R .

Testing Significance in Multiple Regression.

The significance of a multiple correlation may be tested by calculating a variance ratio, F , or by calculating chisquare using L - criterion which is given by

$$L = 1 - R^2$$

and chisquare itself is given by

$$\chi^2(t) = -[n - 1 - \frac{1}{2}(t + 2)] \ln L$$

Where t is the number of predictor variables; this equal to the number of degrees of freedom of χ^2 .

The variance ratio, F , which is preferred to chisquare is ratio of predicted to non predicted variance. The predicted variance has degrees of freedom t and non predicted variance has degrees of freedom $n-t-1$. Note predicted variance has chisquare distribution with t degrees of freedom and non predicted variance has chisquare distribution with $n-t-1$ degrees of freedom. Hence the ratios of these chisquare divided by their degrees of freedom has an F distribution with t and $n-t-1$ degrees of freedom.

Thus the statistics to use to test the null hypotheses that the population correlation coefficient equals zero is

$$\text{i.e. } H_0: R^2 = 0$$

which is the same test as

$$H_0: \beta_1 = \dots = \beta_t = 0$$

$$\text{or } H_0: \beta = 0$$

is

$$F(t, n-t-1) = \frac{R^2}{t} \bigg/ \frac{1 - R^2}{n-t-1}$$

PATH ANALYSIS

Path analysis, due to Wright (1918) is a technique sometimes used to assess the direct causal contribution of one variable to another in a nonexperimental situation. The problem, in general, is that of estimating the parameters of a set of linear structural equations, representing the cause and effect relationships hypothesized by the investigator. Such a system of equations involves variates of two kinds:

exogenous or cause variables

(predictor)

ξ_1, \dots, ξ_q

and endogenous or effect variables

(criterion)

η_1, \dots, η_p

The technique consists of solving the structural equations for the endogenous variables in terms of the exogenous variables to obtain the reduced form of the equations and then estimating the regression of the endogenous (criterion) variables on the exogenous (predictor) variables from this reduced form.

Each equation in the structural equation model represents a causal link rather than a mere empirical association. The structural parameters do not, in general, coincide with the regression coefficients of the observed variables. The use of structural equation models requires statistical tools which are based upon conventional regression and analysis of variance but the tools do go beyond these analyses.

Here ξ 's and η 's are latent variables. These latent variables are not directly observed but there is a set of observed variables that are related to the latent variables. Thus latent variables are variables which are of direct interest in a theoretical specification of the model but which are rarely equivalent to the manifest variables that are actually measured in an experiment.

The task in path analysis, as indicated above, is that of estimating the parameters of the model hypothesized by the modeler (investigator). We consider the most general form of model. This general form of model is considered to consist of two parts:-

- (i) the structural equation model and
- (ii) the measurement model.

The structural equation model:

The structural equation model specifies the causal relationships among the latent variables and are used to describe the causal effects and the amount of unexplained variance. Let the random vector ξ represent the n latent exogenous variables, the random vector η represents the m latent endogenous variables, and the random vector θ represent the m structural disturbance (unexplained variance). The model is specified as

$$\eta = A\eta + B\xi + \theta$$

$$m \times 1 \quad m \times m \quad m \times 1 \quad m \times n \quad n \times 1 \quad m \times 1$$

We assume that the disturbances are independent of the exogenous variables. The reduced form of the model is

$$\eta = (I - A)B \xi + (I - A)^{-1} \theta$$

where $A = \{ \alpha \text{'s} \}$ causal (path) coefficients
 $B = \{ \beta \text{'s} \}$ causal coefficients relating criterion
to predictor variables

we take

$$C(\xi, \theta) = 0 \quad D(\theta) = \delta_{\theta}$$

$$D(\xi) = \delta_{\xi}$$

Measurement model:

The measurement model specifies how the latent variables or hypothetical constructs are measured in terms of observed variables.

We have

$$X = C \xi + \epsilon$$

Measurement model for the
(exogenous) predictor variables)

$$q \times 1 \quad q \times n \quad n \times 1 \quad q \times 1$$

$$q > n$$

Where we take

$$C(\epsilon, \xi) = 0$$

$$D(\epsilon) = \Psi_{\epsilon}$$

$$C(\epsilon, \theta) = 0$$

$$Y = D \eta + \Delta$$

measurement model for the (endogenous)
criterion variables

$$p \times 1 \quad p \times 1, \quad m \times 1 \quad p \times 1$$

where we take

$$C(\Delta, \eta) = 0$$

$$C(\Delta, \boldsymbol{\varepsilon}) = 0$$

$$C(\Delta, \theta) = 0$$

$$D(\Delta) = \psi_{\Delta}$$

$$C(\Delta, \xi) = 0$$

Thus we are assuming the errors of measurement and uncorrelated with η, ξ and θ but may be correlated among themselves.

The parameters matrices are $A, B, \Phi_{\theta}, \Phi_{\xi}, C, \Psi_{\boldsymbol{\varepsilon}}, D, \psi_{\Delta}$

eight in number

$C(X, Y)$ i.e. the covariance matrix Φ [of dimension $(p + q) \times (p + q)$] of X and Y is

$$\Phi = \begin{bmatrix} D(I-A)B\Phi_{\xi}B'(I-A)^{-1} + \Phi_{\theta}D + \Psi_{\boldsymbol{\varepsilon}}D(I-A)^{-1}B\Sigma_{\boldsymbol{\varepsilon}}C' \\ C\Phi_{\xi}(I-A)^{-1}D' & C\Phi C' + \psi_{\Delta} \end{bmatrix}$$

The elements of Φ are functions of the elements of

$A, B, \Phi_{\theta}, \Phi_{\xi}, C, \Psi_{\boldsymbol{\varepsilon}}, D$ and ψ_{Δ}

as can be seen above. In applications some of these elements are fixed and equal to assigned values. In particular, this is so for elements of D, C, A and B , but fixed values will be allowed

in other matrices too. For the remaining nonfixed elements of the of the 8 parameter one or more subsets may have identical but unknown values. Thus, the elements in $D, C, \Phi_{\theta}, \Phi_{\xi}, A, B, \psi_{\epsilon}, \psi_{\Delta}$ (i.e. all 8 parameter matrices) are of three kinds.

- (i) fixed parameters that have been assigned given values.
- (ii) constrained parameters that are unknown but equal to one or more other parameters and
- (iii) free parameters that are unknown and not constrained to be equal to any other parameter.

The model once hypothesized, it has to be tested to find whether it fits the data.

CHAPTER IV

THE ANALYSIS AND DISCUSSION OF RESULTS

Table 1

The variables and the names of the variables used in the analyses:-

<u>Criterion variable</u>	<u>Name used</u>
Professional Studies	PROF
English	ENG
Kiswahili	KISW
Mathematics	MATH
Science	SCIE
Physical Education	PHE
Christian Religious Education	CRE
Islamic Religious Education	IRE
History	HIST
Geography	GEOG
Art and Crafts	ART
Music	MUSIC
Agriculture	AGRIC
Home Science	HOME

<u>Predictor Variable</u>	<u>Name used</u>
0-level English	OENG
0-level Mathematics	OMATH
Aggregate of the best 6 grades of 0-level	OAGG

TABLE II

The intercorrelation matrix for the variables:-

	OENG	OMATH	OAGG	PROF
OENG				
OMATH	.0817(1580)			
OAGG	.1185(1611)	.3948(1589)		
PROF	.2089(1604)	.1309(1582)	.3278(1614)	
ENG	.3833(1609)	.1616(1587)	.3728(1619)	.4470(1613)
KISW	.2153(1610)	.0118**(1588)	.2601(1620)	.2256(1615)
MATH	.1148(1610)	.5230(1588)	.2831(1620)	.3666(1614)
SCIE	.1631(1610)	.3439(1588)	.3879(1620)	.4231(1614)
PHE	.0887(1611)	.1290(1589)	.2437(1621)	.3933(1614)
CRE	.1741(1558)	.0404**(1536)	.3075(1568)	.4487(1562)
IRE	.1001**(47)	.1028**(47)	.2707**(47)	.2765**(47)
HIST	.1051*(631)	.0241**(621)	.2018(635)	.4559(631)
GEOG	.1517(980)	.2383(981)	.3809(986)	.4613(983)
ART	.1518(821)	.1165(806)	.1859(825)	.2821(821)
MUSIC	.0985*(783)	.2104(776)	.2606(789)	.3351(786)
AGRIC	.2315(865)	.1548(859)	.3361(874)	.3923(868)
HOME	.1640(742)	.1187(726)	.3308(743)	.4546(743)

(correlation coefficient of other variables not included here are given on appendix B, C and D).

note on table II

The number in the brackets were the number of cases used to obtain the coefficient i.e. the number of subjects with the two sets of scores in the complete sample (i.e. n).

**indicates the correlation coefficient is not significantly different from zero even at 0.05 level of significance.

*indicates the correlation coefficient is not significantly different from zero at 0.001 level of significance though significantly different from zero at 0.05 level.

Table II indicates correlation coefficients between all the independent variables and all the dependent variables with all the three independent variables as well as with dependent variable PROF. The correlation with other dependent variables were not given here. Note except for IRE all the other dependent variables have correlations significantly different from zero at 0.05 level of significance with both OENG and OAGG. OAGG turning out clearly to correlate the best with all dependent variables (except for IRE).

With this kind of result one cannot fail to be concerned with the way IRE turned out not to have any significant correlation with all the predictor (independent) variables. Some of the questions one may ask are:-

Why were the IRE correlations with all the independent variables not significantly different from zero ? Could it be the grading system which brought about this ?

or could it be the suitability of the examination or the objective and the determination of the curriculum for IRE which needs considerations ? etc.

Here we obtained results which were not expected for IRE; correlation coefficients not significantly different from zero. This implied that there were no relationship between the performance of IRE with the performance in O-level. It was indeed expected there to be a relationship; because the skills or abilities IRE was measuring are expected to be revealed by previous school performances and consequently one expected ~~to~~ a correlation different from zero for IRE with previous school performance (O-level). IRE consistently showed correlation coefficient not significantly different from zero with even all other dependent variables, except with HIST, a correlation coefficient 0.4959 with 19 subjects, significance level being 0.015.

Because of this and also because of the small number of the subjects who had offered this subject, it was not given any further consideration in the analyses which followed. More discussion of table II is given after presenting table III.

From the high positive, significantly different from zero, correlations between the variables, it shows these variables are essentially measuring just a few common factors. This means parsimonious explanation of college performance can be given by just combining the variables which logically and statistically are the same. Such combination was done here. Note, it is usually reasonable to consider variables which are highly inter-correlated together. Another way may be, just choosing any one of them arbitrarily, then considering it as the representative of all others which is not very reasonable. The most reasonable way is to let every one of them make its contribution and we know no other better way than standardizing them as done here and then combining them.

The combination of standardized dependent variables was as follows:-

ENG and KISW were combined to form a new dependent variable called 'LANG'.

another dependent variable called 'BEHA' was made of CRE, HIST, GEOG, ART, PHE·MUSIC AND HOME.

the third dependent variable was 'PHYS' and was made of MATH SCIE and AGRIC.

As indicated earlier a standardization of score was done on a scale of, mean 50 and standard deviation 10 whenever two or more variables were to be combined to form one variable. Thus, LANG, BEHA and PHYS were formed by standardized scores of the respective subjects; for instance LANG was made of sum of standardized scores, standardized to have a mean of 50 and standard deviation of 10, of ENG and KISW and analyses were carried out with such scores. The following were the results:-

TABLE III

Table showing correlation coefficient of the dependent variables LANG, PROF and PHYS with the independent variables OENG, OMATH AND OAGG:-

	LANG	BEHA	PHYS	PROF	OENG
LANG					
BEHA	.2676(1622)				
PHYS	.1980(1622)	-.5123(1622)			
PROF	.3750(1622)	.2941(1622)	.1950(1622)		
OENG	.3649(1611)	.1498(1611)	.0615*(1611)	.2089(1604)	
OMATH	.1117(1589)	.0042**(1589)	.3141(1589)	.1309(1582)	.0817(1589)
OAGG	.3905(1621)	.2233(1621)	.1854(1621)	.3278(1614)	.4185(1611)

	OMATH	OAGG
LANG		
BEHA		
PHYS		
PROF		
OENG		
OMATH		
OAGG	.3948(1589)	

notes on table III

The number in the parentheses were the number of cases used to obtain the coefficient i.e. the number of subjects with the two sets of scores in the complete sample (n).

**indicates the correlation coefficient is not significantly different from zero even at 0.05 level of significance.

*indicates correlation coefficient is not significantly different from zero at 0.001 level of significance.

The following are the observations we can make from the above two tables, table II and table III:-

It seems all the three independent variables are good predictors of PROF (from table II) the best single predictor of it, is OAGG, followed by OENG and next by OMATH. Observe this observation is made by looking at the correlation coefficient between PROF with OAGG, OENG and OMATH (the independent variables), on table II, which are all significantly different from zero at 0.001 level of significance. Again from table II we observe the three independent variable (OENG, OMATH and OAGG) predict very well, nearly all the dependent variables, PROF, ENG, KISW etc. except IRE and to some extent HIST an observation not expected. The independent variable OMATH failed to correlate significantly, infact no correlation, with dependent variables KISW, CRE and HIST, an observation which does not come as surprise.

OAGG is the best predictor of nearly all the dependent variable except the dependent variable MATH which is best predicted by OMATH. The dependent variable ENG is almost equally predicted by OENG and OAGG but as is observed from the two coefficients, the coefficient for OENG with it (ENG) is just slightly higher of the two, thus, OENG is better of the two as far as prediction is concerned. The results agree very well with the common belief or the common expectation based on the well known fact that "previous school performance reveals qualities within the learner or student that are also needed in the next learning institution". Thus the previous record of academic attainment may be used as a predictor of college performance since assumed to require similar abilities. Thus those who are good in school tend also to be good in college and so on. This is infact why we expect there to be high correlations between independent variables and dependent variables where the abilities or skills are similar.

Looking at table III we observe that the new dependent variable LANG is best predicted, almost equally, by OAGG and OENG and to a lesser extent by OMATH. While the new dependent variable BEHA is only predicted by OAGG and OENG but most suitably by the former, OAGG and has no correlation whatsoever with performance in OMATH. On contrary, but as expected, the new dependent variable PHYS is best predicted by OMATH and to some extent, but not as well as OMATH, by OAGG, OENG is a poor predictor of the dependent variable PHYS.

The negative correlation coefficient between the dependent variable BEHA with the dependent variable PHYS, a coefficient quite greatly different from zero, is of very great interest. It shows the two are in opposite poles i.e. the tendency is, those who are doing well in PHYS tended to do not well in BEHA and vice versa. Thus the two dependent variables are measuring two totally different abilities or skills, which indicates that the abilities or skills required in PHYS are totally different from abilities or skills required in BEHA; and almost totally negatively related. It is very interesting results for it reveals we can split the group into strictly two groups, a group good in PHYS and not at all good in BEHA and another group good in BEHA and not at all good in PHYS

and hardly any overlap between the two groups. Thus subjects good in both PHYS and BEHA are almost none existence. This fact may be explained by their orientation and training probably. It is well for us, for it justifies our consideration of the two separately, for they are totally different dependent variables.

LANG has something common with both BEHA and PHYS. This is so since both are correlated significantly with LANG. While BEHA and PHYS had hardly nothing in common, nonetheless BEHA and PHYS both have LANG component in common. Similarly PROF has something in common with all the other new dependent variables namely LANG, BEHA and PHYS. These results will be put in consideration when postulating the causality model which will be formulated and tested its fit.

A puzzling observation is how come BEHA and PHYS are very highly negatively related but OAGG is predicting both of them very well ? Is OAGG a kind of all weather predictor ? We kind of expect OAGG to predict well one of them, either PHYS or BEHA but not both but as we observe the case is not as expected. OAGG best predicts LANG followed by PROF then BEHA and lastly PHYS and as we find the correlation are all significantly different from zero

even at 0.001 level. The implication of the results is OAGG is a very reliable predictor of any of the dependent variable (or college success). It is a predictor we can rely on for any of the dependent variables and consequently better and more reliable than the other two OMATH and OENG for general performance in the college.

Thus OAGG does have all the basic components, may it be, the component or components found in BEHA or PHYS or PROF, OAGG has all those and this why it is best overall predictor of what we may refer to as college success or general performance in the colleges.

We next look at the multiple regression analyses to gain more meaningful insights at the above observations and then kind of give a summary of the findings encompassing all different analyses done here so far.

The multiple regression analyses for the sample of 1622 subjects considering the new dependent variables PHYS, BEHA, LANG and PROF with the independent variables OENG OMATH and OAGG:-

The following are the summary tables showing every one of these four new dependent variables, PHYS, BEHA, LANG and PROF, with their regular regression coefficients, B, and normalized regression coefficients, BETA. These regression coefficients are the ones which each variable would have if it alone were brought into the regression and given with them are their F-values. Also given is the regression constant which is redundant unless one is interested in regression equation. The other statistics given on the tables are the multiple correlations between the dependent variable with particular independent variable or variables in a stepwise fashion. The first multiple correlation which is of the first independent to be listed will be multiple correlation between that particular independent and the dependent and the next multiple will be of the two independent variables with the dependent variable while the last multiple correlation will be of the three independent variables with the dependent variable. The F-values given on the table gives the significance of the multiple or simple regression coefficient whether they are significantly different from zero. Here a conservative test was adopted using the common F value ($F_{0.95}^{1,1618} \approx F_{0.95}^{1,\infty} \approx 3.84$). Thus since F-test is usually

made as one sided test with value of F larger than $F_{\alpha-1, a(n-1)}^{1-\alpha}$ as critical region or rejection region.

Consequently for all F values greater than 3.84 we shall reject the null hypothesis, that the population parameter i.e. either regression coefficient (or weight) or the multiple correlation is significantly not different from zero and accept the alternative which is, it is different from zero at 0.05 level of significance. Note simple regression coefficient is the ordinary regression coefficient B or Beta as explained and their corresponding F -values for testing their significance are given next to them.

Also note the square of multiple correlation which can be obtained easily gives the proportion of the variance in the dependent variable accounted for by the regression equation.

The summary tables showing the regression coefficients with their F-values for significance and multiple correlation also with their F-values:-

Table IV.1 dependent variable: PHYS

<u>Independent variable</u>	B	Beta	F-values	Multiple correlation	F-values
OENG	0.03	0.01	0.1	0.06(.00)	2.41
OMATH	0.80	0.29	125.2	0.32(.10)	91.20
OAGG	0.06	0.07	6.0	0.32(.10)	63.0
Constant	80.43				

Table IV.2 dependent variable: BEHA

<u>Independent variable</u>	B	Beta	F-values	Multiple correlation	F-values
OENG	0.23	.06	4.7	0.14(0.02)	36.9
OMATH	-0.29	-.09	11.6	0.15(0.02)	18.5
OAGG	0.24	.24	67.2	0.25(0.06)	35.2
Constant	121.92				

Square of multiple correlation are the ones given in the parantheses.

Table IV.3 dependent variable: LANG

<u>Independent variable</u>	B	Beta	F-value	Multiple correlation	F-value
OENG	.27	.24	93.9	.36(0.13)	235.2
OMATH	-.03	-.03	1.5	.37(0.14)	128.6
OAGG	.09	.30	126.5	.45(0.20)	134.5

Table IV.4 dependent variable: PROF

<u>Independent variable</u>	B	Beta	F-value	Multiple correlation	F-value
OENG	0.06	0.08	9.9	0.19(.04)	57.72
OMATH	0.01	0.01	0.25	0.22(.05)	41.5
OAGG	0.05	0.26	86.5	0.31(.10)	58.0
Constant	35.79				

Square of multiple correlation are given on the parentheses.

The following are the conclusions we can make by looking at the four tables in turn starting with the first table IV.1:-

The dependent variable PHYS in table IV:1 which is the combination of MATH, SCIE and AGRIC is well

predicted by OMATH. OMATH accounted for about 10% of variance of the dependent variable PHYS while by including any of other independent variable separately or both simultaneously does not increase this variance any further than 10%. The implication here is PHYS is only well predicted by OMATH and hardly any improvement in prediction is obtained by including the other independent variables in the prediction (regression) equation. Note the above observations were made by inspection of squares of the multiple correlations which are given on the tables. We would arrive at similar results by looking at the regression weights (coefficient). As may be observed the regression of OAGG is just barely different from zero while that of OENG is no different from zero, as we see from F-values. The regression weight for OMATH has a F-value quite large and this indicates its big usefulness in the regression model or its usefulness in prediction of PHYS. The regression weight should be interpreted with caution for these are heavily influenced by the nature of the other variables in the regression equation (model). But all the same we observe OENG has nil usefulness as far as prediction of PHYS is concerned. These are the same observations made by observing regression weight as made above in terms of variance accounted for.

The prediction for dependent variable BEHA (refer table IV.2) by the three independent variables is not as good as for PHYS. We observe we have a negative regression coefficient between this dependent variable, BEHA, with OMATH and this regression coefficient is clearly different from zero at 0.05 level of significance. This is so since the F-value for it, is 11.6 as compared to the critical F-value 3.84. We observe by including it (OMATH) in the regression equation (model) we gain nothing as far as the proportion of variance accounted for, is concerned. That is, there is hardly any rise in multiple correlation from 0.15, the initial multiple correlation of the dependent variable with independent variable OENG. By including OAGG in the prediction model (or regression model (equation)) the variance accounted for by independent variables, the 2%, nearly all of it accounted for by OENG, rises to 6%. As may be judged by inspecting the regression coefficients too, we see OAGG is the best single predictor of BEHA then followed by OENG, though the amount of variance for BEHA they account for is well less than 10%. Only 6% and OAGG accounts for about 4% of it.

The obvious implication here is OMATH has a negative usefulness on the prediction (or regression) model of the dependent variable BEHA. Observe even by

including an independent variable with negative usefulness to the prediction model, it does not lower the multiple correlation coefficient. This should be expected for the square of the multiple correlation indicates the amount of variance for the dependent variable accounted for by the independent variable(s). Variance is always a positive quantity and at least the lowest it can get is zero and consequently, the multiple correlation coefficients never drops by addition of other independent variable no matter whether they have negative usefulness or no usefulness at all, as may be observed on the tables.

What we are saying here is there was no gain in considering OMATH in our prediction model for BEHA. Neither are we saying there was loss in including OMATH in the model. This is not a case of either loss or gain, we gain nothing nor lose anything as far as prediction is concerned.

As observed earlier we would arrive at essentially same results by inspecting the regression weights with their corresponding F-values.

The dependent variable LANG in table IV.3 is

best predicted by OENG as may be expected. OENG accounts for about 13% of variance of the dependent variable LANG which is quite high. By including OMATH which has a negative regression coefficient (negative usefulness) though not significantly different from zero at 0.05 level of significance, the variance of LANG accounted for just rises barely by a unit, to 14%. Thus we do not need to include OMATH in the prediction model of LANG. When we include OAGG in the prediction model the amount of variance accounted for by the three independent variables is 20%, implying that while OENG makes about 13%, OAGG makes about 7% of variance accounted for. Thus OENG is the best predictor of LANG while the second is OAGG which makes an appreciable amount of contribution as far as the variance accounted for is concerned.

May be of interest is how the pattern of the contribution would have been if OAGG was included first in the model. Well, this is a deliberate move to have OENG first for this is a simple variable to obtain while OAGG is a combination of six best subjects of which many a time O-level English or O-level mathematics grades or both may be included in it (the variable OAGG). So no effort was ever wasted in trying to have OAGG first in the regression model (or prediction model) because of this reason.

The dependent PROF in table IV.4 which may be

considered to reveal the professional aspect of the training that is assumed to reveal the quality of ones teaching or that kind of thing, seems to be clearly best predicted by OAGG followed by OENG. As said above we cannot absolutely rely on OAGG as a predictor for it is a composite of several variables which apparently are not uniform (same) as OENG and OMATH from subject to subject. This is so because the best six subjects normally would vary from subject to subject but there is uniformity of single variables like OMATH and OENG. So with this in mind we cannot underscore the importance of OENG in predicting PROF; which accounts for about 4% of the variance of PROF. By including OMATH we barely rise this variance accounted for to 5% and when we include OAGG this goes upto 10%.

In summary, we find OENG is quite good predictor of LANG, PROF and BEHA, dependent variables made of scores of final year examination in colleges, examination administered by Kenya Examination Council. OMATH is best predictor of PHYS. Quite an amount of precision would be gained by including OAGG in the prediction model for LANG and PROF or by using it (OAGG) alone especially so in PROF.

Thus looking at the hypothesis explicitly we find that O-level grades do predict performance in Primary Teachers' training colleges. As found OENG predicts very well performance in LANG, PROF and BEHA while OMATH predicts very well performance in PHYS; and we found quite an amount of precision, as far as prediction is concerned, would be realized by including OAGG in the prediction of PROF and LANG.

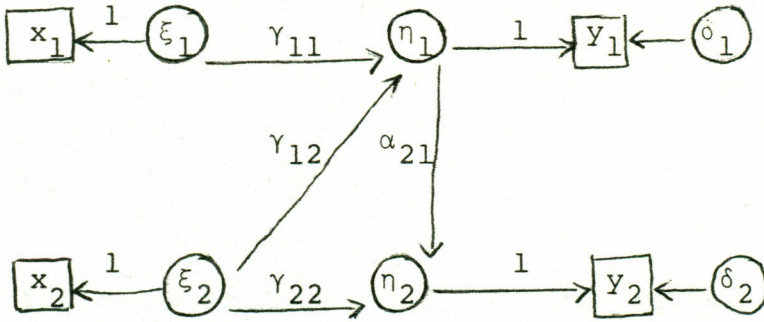
Thus we have a good justification of using O-level grades for selection for P1 teachers.

The second hypothesis is more difficult to test and may arise a lot of controversy in the procedure and feasibility. However it should be pointed out that because of the nature of impossibility of realizing experimental conditions we have no other choice but to resort to this kind of procedure for establishing causality.

We begin by formulating a simple model which we study. If need be, it (model) will be relaxed and modified. The parsimonious structural equation which will first be postulated for this big sample of 1622 subjects is:-

Not
really
necessary
since
it is not
possible
in
S. Garcia

Structural equation model I



x_1 = OENG

y_1 = Social Science

x_2 = OMATH

(a combination of Eng,
KISW, CRE Geog etc.) i.e
combination of BEHA and LANG
call it SOCL.

y_2 = Physical science

(Math, AGRIC and SCIE)

i.e. PHYS

Circles denote unobservable variables (latent variables) while squares denote respective observable variables.

The correlation matrix between the variables of interest was:-

	Y_1	Y_2	x_1	x_2
Y_1	1			
Y_2	.4049	1		
x_1	.2304	.0623	1	
x_2	.0367	.3161	.0817	1

$$Y_1 = \text{Social science} \quad (\text{SOCI})$$

$$Y_2 = \text{Physical science} \quad (\text{PHYS})$$

$$X_1 = \text{OENG}$$

$$X_2 = \text{OMATH}$$

From the above structural equation model we get the following structural (or causal) equations:-

$$\eta_1 = \gamma_{11} \xi_1 + \gamma_{12} \xi_2$$

$$\eta_2 = \gamma_{22} \xi_2 + \alpha_{21} \eta_1$$

hence the reduced structural equations are:-

$$\eta_1 = \gamma_{11} \xi_1 + \gamma_{12} \xi_2$$

$$\eta_2 = \gamma_{22} \xi_2 + \alpha_{21} (\gamma_{11} \xi_1 + \gamma_{12} \xi_2)$$

$$= \alpha_{21} \gamma_{11} \xi_1 + (\gamma_{22} + \alpha_{21} \gamma_{12}) \xi_2$$

The measurement model is

$$X_1 = \xi_1 \quad \gamma_1 = \eta_1 + \delta_1$$

$$X_2 = \xi_2 \quad \gamma_2 = \eta_2 + \delta_1$$

Note the measurement model specifies how the latent variables or hypothetical constructs (ξ and η) are measured in terms of observed variables. We assume both independent variables (X'S) and dependent variables (Y'S) have 1 - 1 correspondence with the respective latent variables and we have no measurement error in independent variables (or cause variables). The measurement error or disturbance term is only found in dependent variables (effect variables).

The structural equation model specifies the causal relationships among the latent variables and are used to describe the causal effects. The unexplained variance will be incorporated in the error term found in the measurement model equation.

It is not difficult to see how measurement model equation and structural equation model may be combined to form a more general form of model and in our case is

$$Y_1 = \gamma_{11} x_1 + \gamma_{12} x_2 + \delta_1$$

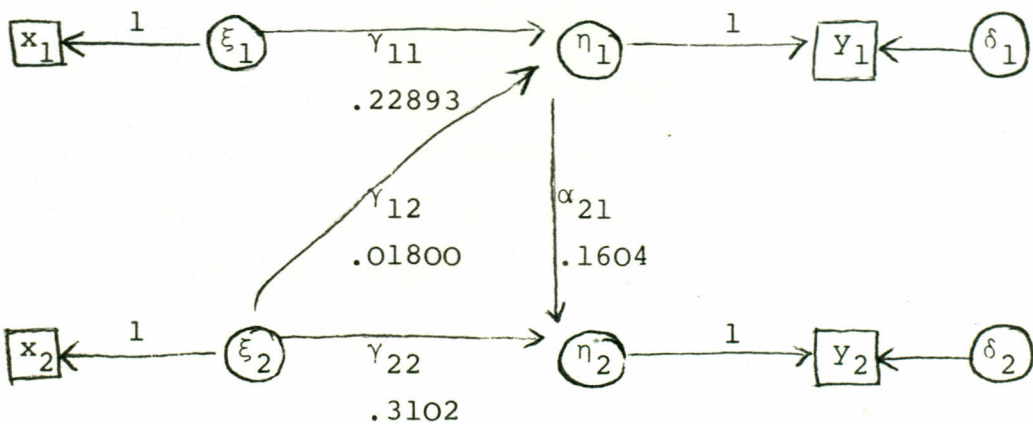
$$Y_2 = \alpha_{21} \gamma_{11} x_1 + (\gamma_{22} + \alpha_{21} \gamma_{12}) x_2 + \delta_2$$

This is what is at times referred to as 'reduced

regression equations'; and using the usual procedures (e.g least square method, maximum likelihood method etc.) the regression coefficients may be found. Maximum likelihood method has more advantages over all the other methods hence it is the commonly used method. (for complete discussion on this refer to Marsden p.v. p.209 an article written by J. Scott Long). These regression coefficients are special coefficients, and are referred to "path coefficients".

We have 'just identified' case. This is so because there are the same number of path coefficients as there are simple or partial regression coefficients.

The path coefficients were obtained together with multiple regression coefficients by the computer, using the SPSS package - SPSSH version 5.01; and using these the following path coefficients given in the diagrams were obtained.



Model I

A path diagram showing the estimates of path coefficients.

From the path diagram we observe that there is hardly any causal relationship between latent variable OMATH and latent variable Social Science (SOCI). This is a finding which would be considered when postulating another model, for surely as it will be found, the above model does not fit the data.

The formal tests for fit of data to the model, I feel has a lot of limitation as I will bringing it out here:-

Given certain independent variables (exogeneous) and dependent variables (endogeneous) there are several ways in which several structural equation models may be formulated. For instance there is quite a number of other ways a model I above may be formulated just by changing the ways the two independent variables may be causally related to the two dependent variables, and for all these many models there is just one test for the fit of the model to the data. This is an area which definitely needs some research on. With this

big limitation of the tests of the fit of the model to the data, the author felt that the estimates of the (population) path coefficients should be given more weight. After all these path coefficients will always be obtained whether the model is accepted or rejected. This above all, always gives an idea of how the causality relationship is, from variable to variable.

The significance of path coefficients is really very meaningful though not usually given. Since this requires one to know the type of sampling distribution these path coefficient have, so as to say when path coefficient is significantly different from zero or not, an area requiring some work on. At the moment I will just compare the magnitude of these path coefficients after the fit of the model to the data has been carried out.

With the limitation mentioned, we nevertheless need to look at the test for the fit of the models. We start with model I. To test the model we take and treat Kigari sample (N=293) as our sample while all samples (N=1622) are treated as the population. Chisquare (χ^2) for test is as follows:-

Minimum value of G (assuming least square method was used to estimate the parameters)

$$G = \text{tr}(\mathbf{I} - \mathbf{S}^{-1}\mathbf{\ddagger})^2$$

where

\mathbf{S} is the sample correlation matrix

$\mathbf{\ddagger}$ is the population correlation matrix

\mathbf{I} is the identity.

For model I $G = 0.06351$ hence $\chi^2 = 0.06351 \times 292 = 18.54$
(note $G \cdot (n-1)$ where n is size of sample has a χ^2 distribution).

The conservative test at 0.05 level of significance has critical values for testing H_0 against H_1 (H_1 is $\mathbf{\ddagger}$ is any positive definite matrix; meaning that the model does not fit the data) as 0.484 and 18.3 as critical χ^2 values. (these χ^2 values for degrees of freedom for lower χ^2 and the maximum value for degrees of freedom for upper χ^2 , significance level being 0.05). We observe that we just barely reject the null hypothesis, accepting the alternative H_1 .

Looking at the model more closely at the individual

path coefficients we notice that some path coefficients are not really different from zero (i.e. the path coefficient for OMATH and SOCI) and hence this consideration would be put into effect when relaxing the model for better fit. The model II (to be discussed later) was tested its fit using similar procedures as above, as shown below:-

Note model II is a relaxed and modified version of model I and has

$$G = 0.093404$$

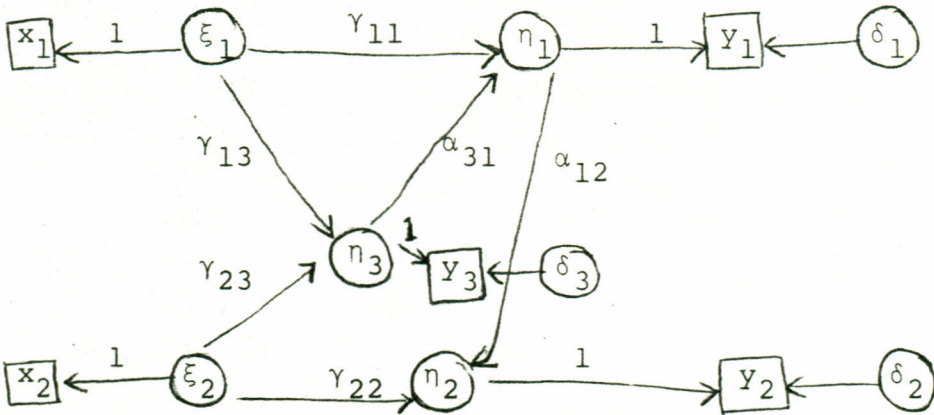
hence has an observed chisquare $(\chi^2) = 0.093404 \times 292$
 $= 27.27$

The conservative test for this model has critical value for χ^2 as 0.831 and 27.5 (obtained as above). We note that we barely accept the null hypothesis. So our relaxation and modification paid off, at least. The above agrees with what we observe when we look at path coefficients (worked out later and can be found on path diagram on page 47). We observe from this, that all the path coefficients of the model are appreciably different from zero, the smallest being between OMATH and PROF, γ_{23} , which is 0.1027^1 .

After relaxation and modification of model I which

1 see appendix A

apparently failed to fit the data narrowly as observed earlier, the model II which is depicted below was considered. The fit of the data into the model is already tested. Note the limitation of this test as explained above:-



Structural equation model
model II

x_1 = OENG
 x_2 = OMATH

y_1 = SOCI
(Social Sciences
combination of ENG, KISW,
CRE etc.)
 y_2 = PHYS
Physical Sciences
 y_3 = Professional Studies
(PROF)

Circles denote unobservable variables (or latent) while squares denote respective observable variables on the above path diagram.

The correlation matrix between the variables of interest was:-

	Y ₁	Y ₂	Y ₃	X ₁	X ₂
Y ₁	1				
Y ₂	-.4049	1			
Y ₃	.3640	.1966	1		
X ₁	.2304	.0623	.1935	1	
X ₂	.0367	.3161	.1218	.0817	1

$$Y_1 = \text{SOCI}$$

$$X_1 = \text{OENG}$$

$$Y_2 = \text{PHYS}$$

$$X_2 = \text{OMATH}$$

$$Y_3 = \text{PROF}$$

The following are the structural (or causal) equations which can be obtained from model II:-

$$\eta_3 = \gamma_{13} \xi_1 + \gamma_{23} \xi_2$$

$$\eta_1 = \gamma_{11} \xi_1 + \alpha_{31} \eta_3$$

$$= \gamma_{11} \xi_1 + \alpha_{31} (\gamma_{13} \xi_1 + \gamma_{23} \xi_2)$$

$$= (\gamma_{11} + \alpha_{31} \gamma_{13}) \xi_1 + \alpha_{31} \gamma_{23} \xi_2$$

$$\eta_2 = \gamma_{22} \xi_2 + \alpha_{12} \eta_1$$

$$= \alpha_{12} (\gamma_{11} + \alpha_{31} \gamma_{13}) \xi_1 + (\alpha_{12} \alpha_{31} \gamma_{23} + \gamma_{22}) \xi_2$$

These are reduced structural equations and the following are the measurement model equations:-

$$X_1 = \xi_1$$

$$Y_1 = \eta_1 + \delta_1$$

$$X_2 = \xi_2$$

$$Y_2 = \eta_2 + \delta_2$$

$$Y_3 = \eta_3 + \delta_3$$

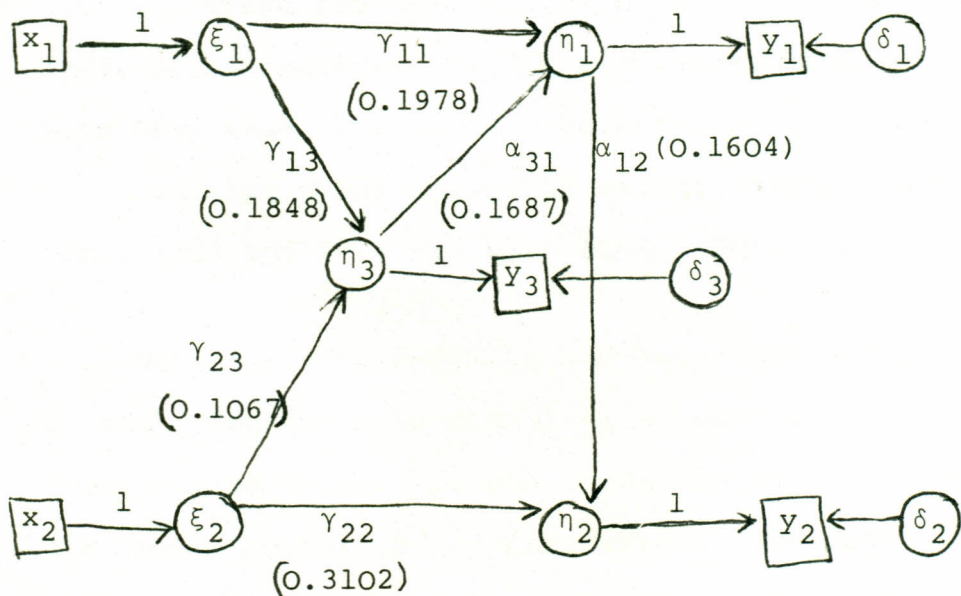
There are many similarities between model I and model II. The combined structural model equation and measurement model equation form a familiar regression model equation in terms of observable variables which can be solved to give the required path coefficients.

$$Y_1 = (\gamma_{11} + \alpha_{31} \gamma_{13})x_1 + \alpha_{31} \gamma_{23} x_2 + \delta_1$$

$$Y_2 = \alpha_{12} (\gamma_{11} + \alpha_{31} \gamma_{13}) x_1 + (\alpha_{12} \alpha_{31} \gamma_{23} + \gamma_{22}) x_2 + \delta_2$$

$$Y_3 = \gamma_{13} x_1 + \gamma_{23} x_2 + \delta_3$$

We have another case of justidentification, and consequently the estimates of the path coefficients can be obtained with ease. The estimates of these path coefficients were obtained using the computer package SPSSH-version 5.01. The path diagram below depicts them clearly.



Model II

Note that not all these path coefficients were obtained directly. Simple computation was done to obtain many of them; of course from the coefficients obtained by means of computer.

The model II has definitely a better fit of the data than model I. It should be noted that the two models have a lot of similarities, even some path coefficients are absolutely identical and consequently many interpretation of model II hold true for model I.

We observe the best causality relationship is between OMATH and PHYS followed by between OENG and Social Science (SOCI). These findings are true of model I too. We find Professional studies (PROF) has a better causal relationship with OENG than with OMATH. Note the kind of chain causal relationships which are quite strong existing between OENG, SOCI and PHYS and then OENG, PROF, SOCI and PHYS.

Here the very definite and interesting observation is that OENG has more causal relationship in quantitative terms. It has causal relationship with all the dependent variables considered in the model. In qualitative terms definitely OMATH turns out to have the best causal relationship with dependent variable PHYS.

It (OMATH) has a weak causal relationship with PROF and consequently this makes its chain causal relationship emanating from it(OMATH) need be taken with reservation.

The implication here is that much of good performance in PHYS is due to previous good performance in OMATH; thus a good grade in OMATH does ensure a good grade in PHYS. We have not gone in working to what extend this is true, a problem of probability. A good performance in SOCI and PROF are due to previous good performance in OENG implying a good grade in OENG does ensure a good grade in both SOCI and PROF. Also here we cannot give any probability of this.

In conclusion it should be pointed out the test for the fit of the model should be consumed with caution. We need always to look at individual path coefficients so as to draw more information from the model we have postulated as we have observed.

Thus looking at the second hypothesis more explicitly we come to a conclusion that a good grade in O-level mathematics ensures or determines a good performance in PHYS

and note PHYS was a combination of MATH, SCIE and AGRIC after standardization of the three variables. We can also say with a great certainty that a good grade in O-level English language ensures or determines a good performance in SOCI and PROF and note what these dependent variables are; SOCI is a combination of several variables such as ENG, KISW, etc. which have been standardized before combination while PROF is professional studies. Thus, there is, infact, causal relationship between OMATH with PHYS and between independent variable OENG with both SOCI and PROF.

Thus in summary the study offered enough evidence to reject the hypothesis that O-level grades do not predict performance in final year college examinations better known as Preservice College examinations. Thus O-level grades do predict very well performance in final year college examination. This can be seen as a necessary though not sufficient justification for using O-level for selection.

The study also has the evidence to show there is a causal relationship between O-level grades and performance in final year college examinations.

SUMMARY AND IMPLICATIONS

The study offered enough evidence to reject the hypothesis that O - level grades do not predict performance in Primary Teachers' Training Colleges. This was found in the simple correlation analysis as well as in the multiple regression analysis carried out in the study.

The independent variable OAGG, the aggregate of the best six O - level grades, was found to be the best predictor of any of the dependent variable (or college success). OAGG is a predictor which can be relied on for any dependent variable and hence better and more reliable than the other two OMATH and OENG considered, for general performance in the colleges. OMATH is O - level Mathematics grade while OENG is O - level English language grade.

OENG was found to be a very good predictor of dependent variables with social science orientation while OMATH was found to predict very well for the dependent variables with Physical Science orientation. It was further found that, quite an amount of precision would be gained by including OAGG in the prediction model or by using it alone for nearly all dependent variables.

Thus the findings do justifies the use of O - level grades for selection for Pl student teachers into Primary Teachers' Training Colleges; Note this is only necessary but not sufficient justification.

In an attempt to test the second hypothesis which required more sophisticated method of path analysis, two models were formulated. The best causal relationship was found between OMATH and dependent variables with Physical Science orientation while OENG was found to have causal relationship with dependent variables with Social Science orientation. OENG had more causal relationship with more dependent variables.

The implication was, a good performance in Social Sciences was to a great extent due to previous good performance in OENG while a good performance in Physical Science was to a great extent due to previous good performance in OMATH. The study did not provide the probability of these.

Thus the study did infact establish beyond any reasonable doubt that there is a causal relationship between O - level grades and performance in final year college examinations.

NOT
Conclu
sive
is it?

REFERENCES

- Akeju S.S. A & Michael W.B.
 "Predicting success in the Federal School of Science, Lagos Nigeria.
Educational & Psychological Measurement, 1970, 30 483 - 486
- Allen M. J. Yen W.M. Introduction to Measurement Theory
 Wadsworth Inc. Belmont, California (1976)
- Anderson, T.W. Introductiton to Multivariate Statistical
 Analysis. New York: Wiley 1958
- Arnold, S.F. The Theory of Linear Models and Multivariate Analysis.
 John Wiley & Sons Inc., 1981.
- Baldauf, Jr., R.B. & Dawson R.L.T. The predictive validity of the
 Michigan test of English language proficiency for Teacher Trainee in
 Papau, New Guinea". Educational & Psychological measurement 1980,
 40, 1201 - 1205.
- Bartlett, M.S. 'Internal and external factor analysis'
British Journal of Psychology,
 1948, 1, 73 -81
- Bickel, P.J. & Doksum, K.A. Mathematical Statistics: Basic Ideas and
 Selected Topics. Holden Day Inc. 1977 San Francisco.
- Borg, W.R. & Gall, M.D. Educational Research and Introducton.
 Third Edition Longman Inc. New York 1979.
- Campbell, D.T. & Stanley, J.C. Experimental and Quasi Experimental Designs
 for Research. American Educational Research Association 1963.
- Carter, D. 'Comparison of different shrinkage formulas in estimating
 population multiple correlation coefficients'. Educaitional
 and Psychological Measurement Vol. 39 No. 2 1979.
- Catell, R.B. (Ed). Handbook of Mutivariable Experimental Psychology. Rank 1966
- Conger, A.J. & Jackson, D.N. 'Suppressor variables, prediction and the
 interpretation of psychological relationships'. Educational and
 Psychological Measurement. Vol. 32 1972, p.579 - 599.
- Conger, A.J. 'A Revised Definition for Suppressor variables:
 A guide to their indentification and interpretation.
Educational and Psychological Measurement Vol. 34,

- Conrad, E. & Maul T. Introduction to Experimental Psychology
John Wiley 1981.
- Cooley W.W. & Lohmes P.R. Multivariate Procedures for
Behavioural Sciences. New York:
Wiley 1972.
- Cooley W.W. & Lohmes P.R. Multivariate Data Analysis:
New York: Wiley 1971.
- Cramer E.M. & Block R.D. 'Multivariate Analysis' Review of
Educational Research, 1966, 36
604 - 617
- Cronbach L.J. Esentials of Psychological testing: New
York: Harper and Row 1970.
- Cronbach L.J. & Meehl P.F. 'Construct validity in Psychological
tests'. Psychological Bulletin,
1955, 52, 281 - 302.
- Cronbach L.J. "Beyond the two disciplines of Scientific
Psychology". American Psychologist: January
1975.
Daily Nation No. 7174 30th December 1983.
- Darlington R.B. "Multiple regression in Psychological Research
and Parctice." Psychological Bulletin, 1968, 69.
- Darlington, R.B., Weinberg S.L. & Walberg H.J. "Canonical
Variate analysis and related techniques"
Review of Educational Research, 1973, 43, 433 -
454.
- Draper, N.R. & Smith, H. Applied Regression Analysis:
New York: Wiley 1966.
- Dunn, O.J. & Clark V.A. Applied Statistics: Analysis
of Variance and Regression: New York:
John Wiley & Sons Inc., 1974

- Dunteman, G.H. & Bailey J.P. "A canonical correlation analysis of the strong vocational interest blank and the Minnesota Multiphasic Personality Inventory for a female college population. Educational and Psychological Measurement, 1967, 27, 631 - 642.
- East African Certificate of Education. Regulations and syllabuses 1974.
- Edwards, A.L. Multiple Regression and the Analysis of Variance and Covariance. W.H. Freeman and Company 1979.
- Gatumu, H.J.N. 'A study of Predictive Validity of 0-level, and aptitude test in relation to the performance at University of Nairobi'. Master's thesis, University of Nairobi, Kenya, 1976.
- Goldberger, A.S. & O.D. Duncan (Ed.) Structural Equation Models in the Social Sciences: Seminar Press Inc. 1973.
- Goldman R.D. & Hewitt B.N. "Predicting the success of Black, Chicano, ~~Oriental~~ and White College Students." Journal of Educational Measurement, 13, 1976, 107 - 117
- Goolsby T.M. 'The Validity of College level examinations programs tests for use at College sophomore level. Educational & Psychological measurement 1970, 30, 375 - 380.
- Gordon R.A. 'Issues on Multiple Regression'. The American Journal Sociology Col. 73. No. 5, 1968.
- Gulliksen, H. 'Methods for determining the equivalence of measures'. Psychological Bulletin, 1968, 70, 534 - 544.
- Gupta, R.O. 'Certain techniques of multivariate analysis applied to different measures of interitem relationships for developing unifactor tests.' Journal of Educational Measurement: 1968, 5.
- Graybill, F.A. An Introduction to Linear Statistical Models - McGraw Hill book Company Inc. 1961.

- Halinski, R. & Feldt, L. 'The selection of variables in multiple Regression.' Journal of Educational Measurement: 1970, 7.
- Hollard J.L. & Richards J.M. (Jr.) "Academic and non academic accomplishment correlated or uncorrelated?" Journal of Educational Psychology, 1965, 56, 165 - 1974
- Hope, K. Methods of Multivariate Analysis: London: University of London Press Ltd. 1968.
- Horst, P. 'Relations among sets of measures' Psychometrika, 1961, 26, 129 - 149.
- Hotelling, H. 'The most predictable criterion'. Journal of Educational Psychology, 1935, 26 139 - 142.
- Hoyt D.P. "College Grades and adult accomplishment. A review of research". Educational Record: 1966, 47, 70-75.
- Hwangkw-Yann & Dizney H.F. "Predictive Validity of test of English as a foreign language for chinese graduate students at an American University. Educational & Psychological Measurement 1970, 30, 475 - 477.
- Jackson D.N. & Chan D.W. 'Maximum Likelihood Estimation in Common Factor Analysis - A Cautionary Note'. Psychological Bulletin, 1980, Vol. 88 No. 2. p. 502 - 508.
- Jöreskog, K.G. "Analysing Psychological data by structural analysis of covariance matrices". Contemporary Development in Mathematical Psychology Volume II: W.H. Freeman & Co., San Francisco, 1974.
- Jöreskog, K.G. 'A general method for analysis of covariance structures' Biometrika, 1970, 57, 2, p.39.
- Jöreskog, K.G. & Sorbom D. LISRELL - analysis of Linear Structural Relationship by the Method of Maximum Likelihood User's Guide Version IV. National Education Resources, Inc. Chicago, 1978.

- Kenny, D.A. Correlation and Causality: A Wiley Interscience Publication, Wiley, 1979.
- Kerlinger, F.N. & Pedhazur E.J. Multiple Regression in Behavioural Research. Holt, Rinehart and Winston, Inc., 1973.
- Lehn, T., Valadovic, R. & Michael E.B. 'The short term predictive validity of a standardized reading test and of scales reflecting six dimensions of academic self concept relative to selected high school achievement criteria for four ethnic groups'. Educational and Psychological Measurement: 1980, 40, 1017 - 1031.
- Mardia, K.V., Kent J.T. & Bibby J.M. Multivariate Analysis: Academic Press Inc. (London) Ltd. 1979.
- McMenar, Q. Psychological Statics: John Wiley and Sons Inc. Inc. 1969.
- Mendenhall, W. Introduction to Linear Models and the Design and Analysis of Experiments. Wadsworth Publishing Company Inc. 1968.
- Meredith, W. 'Canonical correlations with fallible data'. Psychometrika, 1964, 29, 55 - 65.
- Morrison, D.F. Multivariate Statistical Methods. McGraw Hill Inc. 1967.
- Mukherjee, B.N. 'Application fo canonical correlational analysis for learning data'. Psychological Bulletin, 1966, 66, 9 - 21.
- Mukherjee, B.N. "he prediction of grades in introductory Psychology from tests of Primary mental abilities. Educational & Psychological Measurement, 1965, 25, 557 - 565
- Ohnmacht, F.W. & Olson A.V. 'Canonical Analysis of reading

readiness measures and the Frostig developmental test of Visual perception' Educational and Psychological Measurement, 1948, 28, 479 - 484.

Porebski, O.R. 'Discriminatory and Canonical Analysis of technical college data.' British Journal of Mathematics and Statistical Psychology, 1966, 19, 215 - 236.

Rao, C.R. Linear Statistical Inference and its application. John Wiley & Sons Inc. 1965.

Scheaffer, R.L. Mendenhall, W. & Ott, L Elementary Survey Sampling. 2nd edition Wadsworth publishing Inc. 1979.

Sörbom, D. 'Detection of correlated errors in longitudinal data.' British Journal of Mathematical and Statistical Psychology: 1975, 28, 138 - 157.

Stalnaker J.M. "Psychological tests and Public responsibility". American Psychologist; 1965, 20, 131 - 135.

Stewart D. & Love W. 'A genralized canonical correlation index.' Psychological Bulletin, 1968, 70, 160 - 163.

Tatsuoka, M.M. Multivariate Analysis technique for Educational and Psychological Research: John Wiley & Sons Inc. New York, 1966.

Thorndike, R.M. & Weiss, D.J. 'Stability of canonical components.' Proceedings of the American Psychological Association, 1970, 78, 107 - 108.

Turner, ME. & Stevens C.D. 'Regression analysis of causal paths.' Biometrics, 1959, 15, p.236.

Velicer, W.F. 'Suppressor variables and the semipartial correlation coefficient'. Educational and Psychological Measurement, Vol. 38 No.4, 1978.

Walberg, H.J. 'Generalized Regression models in Educational Research.' American Educational Research Journal, 1971, 8, 71 - 91.

Watkins D. & Astilla E "Intellective and non intellective predictors of academic achievement at Filipino University. Educational & Psychological measurement, 1980, 40, 245 - 249.

Wiley, D.E., Schmidt W.H. & Bramble W.J. 'Studies of a class of covariance structure - structure models.' Journal of American Statistical Association, June, 1973.

Other References

Kuhn T.S., The Structure of Scientific Revolutions:

University of Chicago, 1970.

Appendix ATest of significance using maximum likelihood method:-

I tried to see how the test could have been if maximum likelihood method of estimation was used.

for model I and using Kigari as my sample and the whole population as my population I came with.

$$m = .03199 \text{ for model 1}$$

hence

$$\chi^2 = .03199 \times 292 = 9.34$$

critical values are as before .484 and 18.3
 Implying we accept the model fits the data which contradicts our previous observation. I pointed out how these tests need be used with caution. Certainly we need look at individual path coefficient and looking at them we are fully convinced the results using G instead of M are more reliable (consistent) with the path coefficients.

for model II

$$\chi^2 = 14.54$$

Thus accepting the null hypothesis this does not go against our earlier findings using G instead of M.

Appendix A

Table II

Cont'd

	ENG	KISW	MATH	SCIE
OENG				
OMATH				
OAGG				
PROF				
ENG				
KISW	.3188(1619)			
MATH	.3463(1619)	.1492(1620)		
SCIE	.4920(1619)	.2415(1620)	.5782(1620)	
PHE	.3447(1619)	.1323(1619)	.3124(1620)	.4402(1620)
CRE	.3799(1566)	.2446(1567)	.2146(1567)	.3715(1567)
IRE	.0436**(47)	.0329**(47)	.771**(47)	.1390**(47)
HIST	.4670(635)	.2065(634)	.3466(634)	.4418(635)
GEOG	.3852(984)	.1309(986)	.4337(986)	.5604(985)
ART	.1990(824)	.2324(825)	.3309(824)	.4324(824)
MUSIC	.3553(788)	.2002(788)	.4335(789)	.5092(789)
AGRIC	.4203(843)	.1654(874)	.3706(874)	.5103(873)
HOME	.4261(743)	.2047(743)	.3205(742)	.4396(743)

(see notes at end of these tables P. 139)

Appendix B

Table II

Cont'd

	PHE	CRE	IRE	HIST	GEO
OENG					
OMATH					
OAGG					
PROF					
ENG					
KISW					
MATH					
SCIE					
PHE					
CRE	.3995 (1568)				
IRE	.085 ^{**} (47)	-			
HIST	.3039 (635)	.3343 (614)	.495 [*] (19)		
GEO	.3904 (986)	.3662 (954)	.212 ^{**} (28)	-	
ART	.2720 (825)	.2607 (800)	.6371 (22)	.3145 (326)	.3035 (499)
MUSIC	.3668 (789)	.3444 (761)	.2989 ^{**} (25)	.5114 (307)	.4129 (482)
AGRIC	.3157 (874)	.3649 (849)	.570 [*] 8 (22)	.4602 (323)	.4499 (550)
HOME	.4146 (743)	.5152 (715)	.187 ^{**} 1 (25)	.3099 (310)	.3781 (434)

- indicates no correlation coefficient computed,
 these were for subjects which were alternatives e.g.
 CRE and IRE, no candidate took both.

(see notes at end of these tables p. 139)

Appendix C

Table II

Cont'd

	<u>ART</u>	<u>AGRIC</u>
OENG		
OMATH		
OAGG		
PROF		
ENG		
KISW		
MATH		
SCIE		
PHE		
CRE		
IRE		
HIST		
GEO		
ART		
MUSIC	-	
AGRIC	.3189(470)	
HOME	.1548 [*] (352)	

- indicates no correlation coefficient computed, these were for subjects which were alternatives e.g. AGRIC was an alternative of HOME no candidate took both.

(see notes at end of these tables P. 139)

The number in the brackets were the number of cases used to obtain the coefficient i.e. the number of subjects with the two sets of scores in the complete sample.

**indicates the correlation coefficient is not significantly different from zero even at 0.05 level of significance.

*indicates the correlation coefficient is not significantly different from zero at 0.001 level of significance though significantly different from zero at 0.05 level of significance.

APPENDIX E.

Other assumptions of the path analysis model considered:-

$$1. \quad \mathcal{C}(\vec{\eta}, \vec{\delta}) = 0$$

$$2. \quad \mathcal{D}(\vec{\delta}) = \Psi = \begin{bmatrix} \sigma_{\delta_1}^2 & 0 & 0 \\ 0 & \sigma_{\delta_2}^2 & 0 \\ 0 & 0 & \sigma_{\delta_3}^2 \end{bmatrix}$$

$$3. \quad \sigma_{x_1}^2 = \sigma_{x_2}^2 \quad \text{reasonable since both predictor measure have similar reliability.}$$

APPENDIX FFURTHER DISCUSSION OF MULTIPLE CORRELATION ANALYSISPartial and Semipartial Correlation Coefficients and Multiple Correlation, R, Defined in Their Terms:

Suppose X_1 , X_2 and X_3 are predictor variables, then partial correlation $r_{12.3}$ indicates correlation coefficient between X_1 and X_2 with X_3 held constant. It is defined by formula

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

In other words the partial correlation coefficient $r_{12.3}$ removes the variance of X_3 from both X_1 and X_2 . Thus it is correlation coefficient between variable X_1 and X_2 when the influence of X_3 is held constant. We may wish to remove the

influence of X_3 from only one of the other two variables and not both. Then this will give rise to a semipartial correlation coefficient between X_1 and X_2 with the variance of X_3 removed from X_2 say. This is denoted as $r_{1(2.3)}$ and is

$$r_{1(2.3)} = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}$$

i.e. $r_{1(2.3)}$ is the semipartial correlation coefficient which is the correlation coefficient between X_1 and X_2 after variance that X_3 has in common with X_2 has been removed from X_2 .

If the correlation of the predictor variables among themselves are all equal to zero, then any semipartial correlation say $r_{1(2.3)}$ is equal to r_{12} here X_0 is designated as the critterion variable. Thus $r_{1(2.3)}$ is the correlation between critterion variable X_0 , (symbolised as Y elsewhere) and predictor variable, X_3 . This is a very simplified case and finding multiple correlation coefficient for a case like this is very straight forward as seen later.

The multiple correlation coefficient, R, is given by

$$R^2 = R_{0.123 \dots k}^2 = r_{01}^2 + r_{02}^2 + r_{03}^2 + \dots + r_{0k}^2$$

where r_{0i} ($i=1, 2, \dots, k$) is the correlation coefficient between predictor variable X_i and the critterion variable, X_0 .

It should be noted that this formula is only true when the correlation coefficient between predictor variable X_i ($i=1, 2, \dots, k$) and predictor variable X_j ($j=1, 2, \dots, k$) is zero i.e. the predictor variables are independent (uncorrelated or orthogonal), when $i \neq j$.

Also if predictor variables are independent note the regression coefficients β_1, \dots, β_k would remain unchanged (elements of β would remain unchanged) when we drop some of the predictor variables from the regression equation (model), otherwise the regression coefficients would change.

If the correlation of the predictor variables are not equal to zero, which is the most frequent experience, then the case is not as simple as for the above case.

We first consider a simple case in which we have two predictor variables. We found $r_{0(2.1)}$ the semipartial correlation between predictor variable X_2 and criterion variable, X_0 , after variance that X_1 has in common with X_2 has been removed from X_2 .

The multiple correlation coefficient R in this case is

$$R^2 = R_{0.12}^2 = r_{01}^2 + r_{0(2.1)}^2 \quad (2.1)$$

r_{01}^2 , gives the proportion of total variance, $(X_0 - \hat{X}_0)$ or $(Y - \hat{Y})^2$, accounted for by X_1 when regressed on criterion variable X_0 or Y .

Given that X_1 has been included in the regression model the additional proportion of total variance, $(X_0 - \hat{X}_0)^2$, that can be accounted for by X_2 will be given by $r_{0(2.1)}^2$. Now that X_1 and X_2 are in the regression equation (model) the increment due to X_3 would be given by $r_{0(3.12)}^2$. We observe the multiple correlation, R , in the case of 3 predictor variables is

$$R^2 = R_{0.123}^2 = r_{01}^2 + r_{0(2.1)}^2 + r_{0(3.12)}^2$$

Hence the multiple correlation, R , in most general case is:-

$$R^2 = R_{0.12\dots}^2 = r_{01}^2 + r_{0(2.1)}^2 + r_{0(3.12)}^2 + \dots + r_{0(k.123\dots k-1)}^2$$

Thus if k predictor variables were statistically uncorrelated among themselves (i.e they were orthogonal) then the squared individual predictor - criterion correlation r_{0i}^2 ($i = 1, 2, \dots, k$) would represent separate and distinct proportions of the criterion's variables. Consequently, we would simply sum up these values to determine the proportion of the criterion's variance that is accounted for by all the predictor variables. This sum would give multiple correlation coefficient, R , as seen earlier.

Usually, as earlier indicated, the predictor variables are found to be significantly correlated so that their separate predictor - criterion correlations (predictable variances)

will overlap. The problem, then, is resolved to one of 'partialling' out the overlapped predicted variances in the criterion's variance.

Measures of the importance of a predictor variable in multiple Regression Equation

The usefulness of the predictor variable, X_j is defined in relation to the amount of R^2 that would drop if X_j were removed from the regression equation. Of course new regression weights without X_j would need to be calculated (for a non orthogonal case) and new R^2 would be obtained. The square of multiple correlation, R^2 , is the sum of the proportions of variance in the criterion variable "accounted for by" each of the predictor variables included in the regression model.

Beta weights (or elements of vector β) are determined solely by the characteristics of the orthogonal component of the variable under consideration. They have little relation to the validity and are heavily influenced by the nature of the other variables in the regression equation (model). Beta weights can even change in sign as variables are added to or removed from the equation (model).

The size of B_j is not a measure of the usefulness of X_j when the predictor variables are intercorrelated. When the focus is on the prediction of criterion variable, X_0 , the usefulness of each X_j is the measure of greatest interest.

Usefulness of X_j divided $1 - R^2$ gives the squared partial correlation between X_0 and X_j , holding all other variables constant.

The hypothesis that a predictor variable has zero usefulness in the population is equivalent to the hypothesis that the variable has a population beta weight of zero, so the significance tests of these two hypotheses are the same.

When predictor variables are intercorrelated there is really no satisfactory method of determining the relative contribution of the predictor variables to the regression variance or the proportion $(X_0 - \hat{X}_0)^2$ accounted for by each of the predictor variables, X_0 is the criterion variable and \hat{X}_0 is the predicted criterion variable.

The contribution will depend on the order in which the predictor variables are entered in the regression equation (model). When predictor variables are correlated, each variable usually accounts for a larger proportion of total variance of the criterion variable, $(X_0 - \hat{X}_0)^2$ when it is entered first in the regression equation (model) than when it follows other variables. In other words when several predictor variables are highly correlated with criterion variable, X_0 , it is reasonable to believe that each of the predictor variable X_i ($i = 1, \dots, k$) is accounting for the same

common variance with criterion variable, X_0 . In this case, the semipartial correlations for each of these predictor variables may be substantially correlated with criterion variable, X_0 . For instance if X_i , X_j and X_k (all the three predictor variables) are highly correlated with criterion variable, X_0 , then when any one of the three is entered after the other two, it will probably not represent any substantial increase in the regression variance.

Thus if several predictor variables are highly inter-correlated, it makes sense to create a single new predictor variable that is a composite of the several predictor variables whose correlations are high; or else it would suffice to consider just one of the several predictor variables which are highly correlated.

The selection of variables in multiple regression analysis:

There are two principal uses for the multiple regression techniques and these are:

- prediction which is the techniques' main objective
 - and - identification of relevant predictor variables
- (distinction between the two is elusive)

The process of identification helps the investigator in understanding which skills and traits contribute to criterion

variable variance. Thus in other words, specifying those predictor variables "belong" to the equation (model). In this respect the researcher is often interested in obtaining a reduced set of predictor variables from a large initial set. In case of prediction the elimination of unnecessary predictor variables (predictors) will simplify future data collection procedures and can generally be expected to enhance predictive accuracy with subsequent sampling units.

Various method for accomplishing this reduction have been found. Four empirical method and another involving factor analysis are discussed here. These empirical procedures are the ones commonly employed to the sample data to reduce a set of the predictor variables. Each procedure incorporates an F test of statistical significance as the criterion for terminating the selection process. The procedures are:-

1. The forward selection procedure:-

The predictor variable maximally and significantly correlated with the criterion variable is initially selected for entry into the regression equation. At each successive stage the variable providing the maximum significant increase in the multiple correlation coefficient squared, R^2 , is added to the set of variables previously selected.

2. The stepwise regression procedure:-

This approach differs from the first method by allowing for the deletion of variables at a stage

subsequent to their entry. A variable is deleted when, upon the addition of other variables, its contribution to R^2 becomes nonsignificant.

3. The backward elimination procedure on F:

Initially the regression equation for the full set of predictors is determined. then, for each variable in turn both the reduction in R^2 which would occur if the variable were deleted and the F ratio associated with that decrease are computed. The variable which yields the smallest such nonsignificant F ratio is deleted. The process is repeated with other predictors or until the deletion of another variable would produce an R^2 significantly lower than the R^2 for the full set.

4. The backward elimination procedure on the weights:

This technique is identical to the previous procedure in all but one respect. At each step the variable eliminated is the one with the nonsignificant regression weight closest to zero in absolute value.

5. Another technique for reducing the number of predictors is to factor analyze the set of all available predictors and then use some of the resulting factors in a regression equation in place of the original predictor variables. If the number of factors extracted equals the original number of predictor variables, then it can be shown that the multiple regression equation constructed to predict the criterion variable from the factors is equivalent to the comparable equation constructed from the original variables. The two equations will make identical predictions for any individual since the weight given to each original

variable in the equation based on factors exactly equals the weight given that same variable in the regression equation based on the original variable. Therefore, any improvement resulting from the use of factors as predictors can occur only when the number of factors used is smaller than the number of factors used is smaller than the number of original predictor variables.

Each one of these techniques work best in some situations.

Utilization of multiple regression model for selection and Classification

In most selection situation the decision to accept or reject is based upon the results on several tests. The various sub-scores can be used as a basis for decision making in several different ways.

We look at two models in selection:-

(1) Compensatory and 2) conjunctive models

Multiple regression analysis is based on a linear combination of the data from sub-distributions. An individual's total score is obtained as a sum of scores and consequently, it is possible for someone who is weak in one area to compensate by being strong in another and so be selected. Hence multiple regression model is an example of a compensatory model.

In some situation a linear combination of scores is not suitable for obtaining cut off limits. This is where it is not possible to compensate for weakness in certain respects by

strength in others. If the individual is to be accepted, he must exceed the cutoff limit on each of the subtests. A multiple cutoff method is used. This is what is termed as conjunctive model. A pilot must have good eyesight and good theoretical training and good coordinating ability. He cannot compensate for poor eyesight no matter how good his theoretical training or coordinating ability might be. This being a practical illustration of the model.

A research problem that arises again and again is to classify individuals into groups on the basis of their scores on tests. The simplest case is to place people (or other units of which we have measures of) into two groups on the basis of scores on one test. Thus for instance by using test scores, we assign individual into two (or more) groups.

Thus the criterion variable takes two values say, 1 or 0 distinguishing for instance pass and fail. Using the scores of predictor variables as usual and a criterion vector of 1's and 0's, we solve the regression in usual manner. The resulting equation, the discriminant function (discriminant analysis), maximally discriminates the members of the sample into 1 or 0 i.e. pass or fail.

Shrinkage and correcting for it:

R^2 as found earlier is an index of the amount of variance in a criterion variable that is accounted for by variance in the predictor variables. This statistic tends to become a positive biased estimator of the squared population correlation coefficient (ρ^2) whenever the number of predictor variables increases or the number of cases (N) decreases. This is termed as 'Shrinkage'.

Several formulae have been given for correcting for this shrinkage such that R^2 becomes a less biased estimator of ρ^2 .

Error of R:

The classical formula for the standard error of multiple correlation involving n variables (1 criterion and n-1 predictor variables) is

$$S_{R_{0.12\dots n-1}} = \frac{1 - R_{0.12\dots n-1}^2}{N}$$

If N is large and if value of $R_{0.12\dots n-1}$ is not too high, this formula will provide satisfactory approximation.