# The PLS-PM Approach to Causal Modeling: An analysis of Learners' achievement in Kenya Certificate of Secondary Education examination of 2014

BY
JOSEPH NDIRANGU KURIA
I56/74801/2014

A RESEARCH PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS, UNIVERSITY OF NAIROBI IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN SOCIAL STATISTICS

**DECLARATION**

I hereby declare that this project paper is my original work, which has not been undertaken anywhere else in any university for the award of a degree.

**Joseph Ndirangu Kuria** ...... ...... ...... Date ...... ...... ......

**APPROVAL**

This project has been submitted for examination with my approval as the university supervisor

**Supervisor**

**Dr Nelson Owuor** ........................ Date........................

## DEDICATION

This work is dedicated to my wife Esther Njeri Ndirangu, and my loving children Annabel, Derrick and Samantha. Thank you for the financial and spiritual support you accorded me during the period of the course. You were also patient with me during the long periods of absence from home due to work and the demands of the course.

**ACKNOWLEDGEMENT**

I am very thankful to my supervisor Dr Nelson Owuor for the great supervision and guidance and for being there at the point of need. My sincere appreciation also goes to all lecturers in the School of Mathematics for being great teachers, adequately preparing me to be a statistician, and challenging me to put the knowledge to practice.

I would also like to thank the Director, CEMASTEA, for allowing me to reorganize my work to avail time for classes. Members of staff of CEMASTEA were also kind in supporting me during the course. Special thanks to members of the Mathematics department, Mr Matiri for providing reference materials and Mr Carillus Ogwel who has been an academic mentor and kept urging me on.

To my classmates, you were great people. Special thanks to Bichanga Simon, Vincent and Victor Otieno, Liz and Miriam. You made the course even more interesting through the group discussions.

## ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| KCSE | Kenya Certificate of Secondary Education |
| PLS-PM | Partial Least Squares Path Modeling |
| KNEC | Kenya National Examination Council |
| SMASSE | Strengthening Mathematics and Science in Secondary Education |
| SMASE | Strengthening Mathematics and Science in Education |
| INSET | In–Service Education and Training |
| CEMASTEA | Centre for Mathematics, Science and technology Education in Africa |
| GOK | Government of Kenya |
| NILES | Nonlinear Iterative Least Squares |
| NIPALS | Non-linear Partial Least Squares |
| PLS | Partial Least Squares |
| SEM | Structural Equation Modeling |
| PLS–SEM | Partial Least Squares Structural Equation Modeling |
| ECSI | European Customer satisfaction Index |
| EPSI | European Performance satisfaction Index |
| STI | Science Technology and Innovation |
| MoEST | Ministry of Education science and Technology |
| JICA | Japan International Cooperation Agency |
| ICT | Information Communication Technology |
| MTP | Medium Term Plan |

**ABSTRACT**

Students' performance in mathematics and sciences is closely associated with the scientific and technological innovations worldwide.The Government of Kenya recognizes the important role mathematics and science must play in achieving 'Vision 2030' and invests resources in raising the quality of teaching mathematics, science and technology.

English, a second or even the third language in Kenyan communities, is the language of instruction and assessment in schools.

As the social sciences develop, hypothesized relationships become increasingly more complex, and therefore the need to use more versatile models. Partial Least Square Path Modeling is one of such models. Partial least squares path modeling allows research to study the measurement and structural models of the variables. We can illustrate the structural regressions in complex causal structures by means of Partial Least Squares Path Modeling. Group comparisons,i.e. comparisons of model estimates for different groups of observations,can also be carried out.

The purpose of this study was to investigate the causal effects of the proficiency and achievement in languages on the achievement in mathematics and the sciences; i.e. chemistry, biology and physics. From the study, it can concluded that languages contribute highest to the development of science process skills which in turn contribute to the development of numeracy skills.

In the study, multi-group comparison was also done between the school types to show the differences in contribution of the construct variables. It was established that the path coefficients between literacy and numeracy are significantly different at 5% confidence level between boys and girls schools and between girls and mixed schools. This implies that there are differences in the contribution of literacy skills to the development of numeracy skills from one school type to another,and therefore development of these skills call for different efforts for different school types. The study recommends that, other than the government focusing on mathematics and sciences only, there should be focus on the development of literacy skills in English and Kiswahili. Capacity development activities for mathematics and science teachers should also include enhancement of literacy skills in English and Kiswahili. Policy–makers should also consider planning for capacity development training for English and Kiswahili teachers.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

A good mathematics and science education is crucial for scientific and technological development. Mathematics and science are important school subjects in the Kenyan curriculum that aim to address the need for workforce and general population that is scientifically literate. In recognition of the value of mathematics, the Ministry of Education, Science and Technology (MoEST) in Kenya has made mathematics a compulsory subject in both primary and secondary schools (Republic of Kenya, 1964). This is because mathematics has direct relationship with other subjects, particularly technical and sciences.

Students' performance in mathematics and the sciences at Kenya Certificate of Secondary Education (K.C.S.E) level has been poor as documented in various reports such as the SMASSE Baseline Report of 1998, and the Kenya National Examination Council (K.N.E.C.) yearly reports. All stakeholders in education should should come up with measures to change this poor performance since mathematics and the sciences are strategic for technological, scientific and industrial development. According to the Education and Training Medium Term Plan(MTP)of 2013-2018, Science, Technology and innovation (STI) in education is crucial in ensuring efficiency and productivity. Education and training is important as a vehicle to apply science and technology to deliver development in the country. There is need therefore, for the education system to produce scientists who will participate in growing the economy. Strengthening Mathematics and Science in Education (SMASE) is a programme in the Ministry of Education, Science and Technology (MoEST) which aims at providing In-service Education and Training (INSET) to teachers of mathematics and sciences. The SMASE programme grew from the SMASSE project, a joint venture between the Kenyan government through MoEST, and the Government of Japan through Japan International Cooperation Agency(JICA) that aimed at upgrading the capabilities of students in mathematics and sciences. This was necessitated by the persistent poor performance in Mathematics and Science (Biology, Chemistry and Physics) by learners in national examinations. Some of the issues that SMASSE was to address were the attitude of teachers and students towards science and mathematics; teaching methodology; content mastery; developing and using teaching / learning resources; administration and management.These issues were to be addressed through INSET for teachers of mathematics and science.

SMASSE Impact Assessment Surveys (2006-8) reported a change of attitude towards work planning by teachers. It was observed that teachers were involving learners in practical activities, experiments and group discussions more than at the inception of the project. Teachers were now more involved in developing teaching /learning resources through improvisation. However, going by the KNEC reports, these changes in teaching /learning did not translate to better grades in mathematics and sciences. One of the key features of the public discourse regarding education at the end of the four-year secondary education is the students mean score and the courses to pursue at the university and tertiary institutions. This performance is determined by the performance in the core subjects namely English, Kiswahili and mathematics and the elective subjects.

## 1.2 Background of the Study

KCSE examinations are administered at the end of four years secondary education. The examination are offered to all the candidates in the three core subjects, namely English, Kiswahili and mathematics and in the subjects a student selected in Form Three. Among the subjects to be selected are at least two science subjects from Biology, Physics and Chemistry. A candidate can also select to do the three science subjects.

Kenya is a linguistically diverse country with about 42 tribes each using its dialect. Kiswahili, an indigenous language, is the national language and is therefore used in most of the formal and informal communications. English is the official language of Kenya and, in terms of policy, the medium of instruction from Standard four onward. Teachers are expected to use English as the medium of instruction, with occasional interjections in Kiswahili to enhance learners' understanding in secondary schools.

English is a second or even a third language to many learners in Kenyan schools; coming after mother tongue and Kiswahili. English language is taught at school and at the same time is used as the medium of instruction. Kiswahili is the language used for conversations in school,learner– learner , and sometimes teacher–learner interactions.

Kenya National Examination Council (KNEC) evaluates the performance of students through examinations it offers. The performance of students in the national examinations has been a key measure of the success of educational institutions in terms of management, discipline and provision of essential infrastructure and teaching-learning resources. Examinations can also provide the basis for evaluating the curriculum implementation at the school level.

Examinations are used to measure the level of candidates' achievements, clarify the level of education, and readiness for further education, training and employment after the learning in a school. The concern in the assessment of students in mathematics and sciences is whether assessments measure language proficiency as well as actual content knowledge.

## 1.3 Statement of the Problem

The Kenya National examination council (KNEC) continues to report dismal performance in mathematics and the science subjects despite government intervention measures to enhance students' performance. Some of the intervention measures have focused on enhancing teachers' pedagogical skills and attitude change of students and teachers towards mathematics and the science subjects. There is advocacy for the use of practical activities and experiments in the teaching /learning process. Teachers are encouraged to use locally available materials in the teaching / learning process so as to relate the concepts to learners' real environment. Improvisation of teaching /learning resources is also encouraged so that learners are involved in experiments even where conventional materials are not available. In all these, there is high emphasis on proper planning for the lessons by the teachers. This is aimed at ensuring that teachers plan for the most appropriate and most effective activities and try them before the lessons. Forums are also created for teachers to share experiences on the best practices in teaching mathematics and the sciences. The forums also involve school administrators who are sensitized on prioritizing resources for teaching/learning mathematics and sciences.

ICT integration in teaching/learning of mathematics and science was also introduced with the aim of making the subjects interesting and the otherwise abstract concepts concrete. Despite all these intervention measures, the performance in mathematics and the science subjects is still poor and the enrollment in some science subjects low.

## 1.4 Objectives of the Study

The study sought to establish the relationship between achievement in English and Kiswahili and the achievement in mathematics and the science subjects.

## 1.5 Research questions

The study was guided by the following research questions:

1. Are learners competent enough in the language of instruction and assessment to effectively use it in learning mathematics and sciences?

2. Does proficiency in English and Kiswahili contribute to the achievement in Mathematics and Sciences?

3. Are there differences in the contribution of languages to the performance in mathematics and sciences in the various school types?

## 1.6 Assumptions of the Study

The study made the following assumptions:

1. That the KNEC examination results reflect the true performance /achievement of candidates.
2. That schools' mean performance in English and mathematics remains fairly constant from year to year.
3. That the teaching and learning of English and Kiswahili in secondary schools help in the development of literacy skills essential to the learning of mathematics and science.
4. That the achievement in English and Kiswahili is a measure of proficiency in the languages.

## 1.7 Significance of the study

The study is important for it will guide the government in the value addition to the SMASE INSET for teachers. For many years, the SMASE INSET has confined itself to Science and mathematics disregarding the role played by the language of instruction and assessment. The study will also provide information on how the factors under consideration determine the performance in the two subjects and therefore enabling the designing of intervention measures. The findings in this study will generally inform policy on access and the provision of quality education.

## 1.8 Conceptual Framework for the study

Literacy can be defined as the knowledge and skills that form the foundation for learning, communication, language use and social interaction. Literacy skills range from the basic ability to read, write, listen and comprehend, to higher level processing skills. A learner with literacy skills is able to interpret,elaborate, monitor and deduce on communication presented to him. Literacy skills form the foundation for future learning and participation in society and employment. Literacy skills include the capacity to read, understand and appreciate various forms of communication through spoken language, printed text,digital media and broadcast media.

Numeracy skills refer to the the basic mathematical skills that involve mathematical knowledge, problem solving and communication skills. It encompasses the ability to use mathematical understanding and skills to solve problems encountered in real life situations enabling one to fit in society. To have this ability,

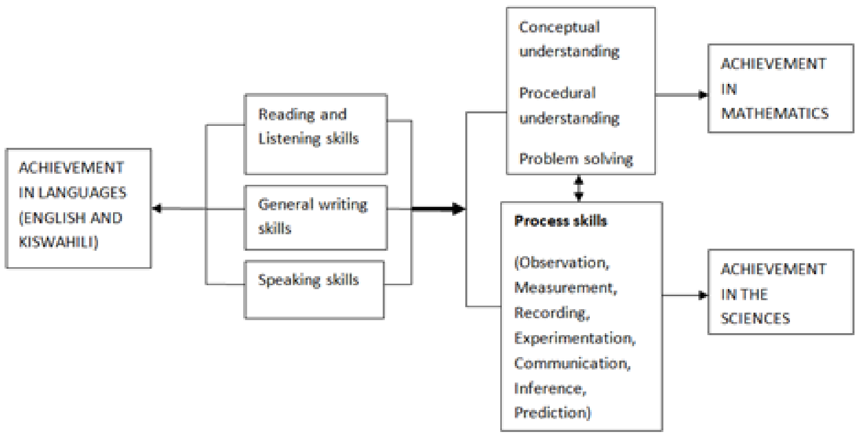


Figure 1.1: Conceptual framework

a young person needs to be able to think and communicate quantitatively, to make sense of data, to have a spatial awareness, to understand patterns and sequences, and to recognize situations where mathematical reasoning can be applied to solve problems.

The conceptual framework of this study is illustrated by Fig 1.1. It defines the interplay of the multiple factors that determine achievement in English, mathematics and the sciences. Achievement in English is determined by the mastery of literacy skills like reading and listening skills, general writing and speaking skills. The achievement in mathematics is determined by the conceptual understanding, procedural understanding and problem solving skills enhanced by the mastery of skills in English. Achievement in the sciences on the other hand is enhanced by the mastery of the science process skills. Science process skills reinforce conceptual and procedural understanding and problem solving skills in mathematics and vice versa. Learning in Science and mathematics reinforce each other.According to the National Council of Teachers of Mathematics(NCTM) (1980), the development of skills in logical reasoning and problem-solving is a goal of both science and mathematics instruction. The development of the science process skills rely heavily on learners' reading and listening skills, writing skills and speaking skills. Students read textbooks, read procedures for conducting experiments, and write their observations and reports.

# Chapter 2

# Literature Review

## 2.1 Introduction

In this chapter, literature on PLS-PM in social sciences is reviewed. This chapter also discusses the literature related to achievement in mathematics and sciences and the proficiency in English in schools.

## 2.2 PLS-PM in social sciences

For close to two decades from 1960 , Herman Wold spent time developing and improving a series of algorithm based on ordinary least squares regressions to handle modeling problems. A number of methods were proposed over the years among them Nonlinear Iterative Least Squares (NILES), Nonlinear Iterative Partial Least squares (NIPALS), Partial Least Squares basic design and Partial Least Squares Soft modeling. The modeling methods were applied to data from social sciences, e.g. economics, psychology and sociology. Svante Wold (1980) applied PLS principles to Chemistry and food industry.

Structural Equation Modeling (SEM) can test theoretically supported linear and additive causal models hence making it popular as a second-generation multivariate data analysis method in marketing research.(Kaplan, 2004; Statsoft, 2013). SEM enables researchers to visually examine the relationships that exist among variables of interest. SEM has continually gained popularity due to the need to test complete theories and concepts (Rigdon, 1998). Much of SEM's success can be attributed to its ability to evaluate the measurement of latent variables, while also testing relationships between latent variables (Babin et al., 2008).
PLS-SEM has recently received considerable attention in a variety of fields which include marketing (Hair et al., 2012b), strategic management (Hair et al., 2012a), management information systems (Ringle et al., 2012), operations management (Peng and Lai, 2012), and accounting (Lee et al., 2011). This is

because of PLS-SEM's ability to handle data with distribution issues such as multicollinearity,non-normality and small sample size that routinely occur in the social sciences.
In international marketing research, PLS-SEM provides a powerful framework for estimating causal models with latent variables and systems of simultaneous equations with measurement errors (Jörg Henseler et al, 2009). PLS path modeling is a statistical technique for estimating parameters of conceptual models. A critical review of the PLS application in international marketing reveals that this methodology has increased in popularity, especially for multi-group analyses of PLS results for different nations.
In the study titled 'Transition from university to employment 'and utilization of graduates', Bruno Chiandotto et al (2002) used Structural Equation Modeling (SEM) in analyzing customers' perceived quality in the ECSI methodology (European Customer Satisfaction Index). The model was to represent the satisfaction of the student/end user with latent variables to be gauged through a set of observable indicators. The results of the analysis confirmed the validity of the application of the ECSI-SEM models in customer satisfaction studies and also stimulated interest in the implementation of more detailed analysis.

The concept of EPSI rating (European Performance Satisfaction Index) that developed from ECSI: European Customer Satisfaction Index was introduced in Martensen et. al. (2000) and had been adapted to measure student perception of exogenous latent variables as institution image, expectations, quality of human and non-human elements of teaching and learning, and endogenous variables such as perception of value, student satisfaction and loyalty. The EPSI model is a Structural Equation Model with latent variables linking customer satisfaction to its drivers. The EPSI model has been adapted to studies on student experiences in institutional image, expectations, quality, value, satisfaction and loyalty. In the studies, it was observed that it is an essential tool in future quality enhancement in Higher Education, at both programme and institutional level, and for bench-marking. According to Jög Henseler et al (2009), PLS path modeling is suitable for prediction-oriented research. The methodology assists researchers who focus on the explanation of endogenous constructs. Many researchers argue that PLS path modeling is most important in exploration and prediction, and recommend it in early stages of theoretical development in order to test and validate exploratory models. The advantages of PLS path modeling mostly considered by researchers in their work include: PLS-PM delivers latent variable scores which are measured by one or several indicators; PLS-PM avoids small sample size problems and can therefore be applied in some situations when other methods cannot; PLS-PM can estimate very complex models with many latent and manifest variables; PLS-PM has less stringent assumptions about the distribution of variables and error terms; and PLS-PM can handle both reflective and formative measurement models.

## 2.3 Performance in mathematics and science

Many studies have been carried on the causes of poor performance in mathematics and sciences. Mwenda et al (2013), in a study of factors contributing to students' poor performance in mathematics in public secondary schools in Tharaka South found out that attitude factors towards the subject, teaching methodology, teaching resources and background factors (of the family and school) determine the performance in mathematics. Despite, enumerating background factors as a cause of poor performance, very little was said about this in the research.

Mbugua, Kibet, Muthaa and Nkonke, (2012) were of the view that performance in mathematics by students can be improved by provision of proper staffing, provision of teaching and learning materials, curriculum reviews, motivation of students and teachers, change of attitude towards the subjects by teachers and students, and reducing the burden of fees and levies. Karue and Amukowa, (2013) were of the opinion that provision of instructional materials, library, laboratory and other physical facilities, head teachers developing good rapport with parents, reducing students and teachers ratio to manageable size are some of the ways of improving performance in mathematics.

#### Language and Learning in Mathematics

English language is used as the medium of instruction in the teaching of mathematics in Kenyan secondary schools. It is also the language used for assessing achievement in mathematics. In schools, "language is both the instrument and the vehicle of teacher-student interaction" (Smith and Ennis, 1961:112) and thus

*"the conduct of classroom instruction is inescapably involved in the use and interpretation of language–written, printed, and above all, spoken. Few indeed are the acts of teaching that entail no verbal dimension, that proceed without some verbal interplay between teacher and students. For just as the act of teaching, however else defined, is an effort to induce learning, so is the language of teaching a taproot to learning" (Smith and Ennis, 1961:113).*

It can therefore be surmised that for a student to achieve in mathematics, he/she must have a good command of the language of teaching/learning and assessment. To fair well in national examinations, a Kenyan student must have a good command of English as the language of instruction and assessment. Effective mathematics instruction is characterized by a well integrated development of students' conceptual understanding, procedural fluency, and problem solving. As they develop these abilities, students must become familiar with mathematics vocabulary and with representation of mathematical ideas in multiple ways. Language plays a significant role in mathematics. Therefore, direct instruction of key vocabulary is a critical element in raising student achievement in mathematics.

Fillmore (1982) asserts that the language of textbooks and instruction "frequently calls for a high degree of familiarity with words, grammatical patterns, and styles of presentation and arguments that are wholly alien to ordinary in-

formal talk". Some of the language vocabularies and patterns used in resources and discussions in the mathematics class may be especially difficult for second-language learners to understand.
The interest in the relationship between language and learning in general is not new. Some theorists (e.g., Whorf, 1956) have suggested that language determines and defines how we think. The Australia National Numeracy Review Report (2008), synthesized evidence on effective numeracy teaching to support the goal of improving numeracy outcomes for Australian students. The report recognized the importance of language in mathematics learning.

The ability to read mathematics in a second language is obviously influenced by a variety of language skills. Clark (1975) proposed a model in which concepts are viewed as the result of the learner's experience, with language facilitating the learner's conceptual development through discussion and instruction. Language is also applied to the content of mathematics in the representation of experience through diagrams and mathematical notation. The relationship between proficiency in language of instruction and mathematics achievement is not clearly understood, although it is reasonable to assume that mastery of mathematical concepts assume that there is some proficiency with the language used to express, characterize, and apply those concepts. As Thorndike (1912) noted, "Our measurement of ability in arithmetic actually is a measurement of two different things: sheer mathematical insight and knowledge, on the one hand; and acquaintance with language, on the other" Research on the relationship between English language and the learning of mathematics, as well as on the role language plays in assessing mathematical concepts and skills is scarce. This research, therefore, is an attempt to study the relationship between achievement in language of instruction and assessment and the achievement in mathematics and sciences at KCSE in Kenya schools.

### 2.3.1 Language and learning in the Sciences

English language is the medium of instruction and assessment in the science subjects, Biology, Chemistry and Physics in secondary schools in Kenya. According to Halliday (1993), language of instruction plays a significant role in learning disciplinary content which includes a unique vocabulary, discourse patterns, and forms of communication.

Language allows students to participate in classroom activities thereby accessing the subject content as defined in the syllabus. In order to understand concepts in science, students must learn how scientific knowledge is constructed, represented and communicated. Barber (2001) asserted that the ways students make sense of science test items are influenced by the values, beliefs, experiences, communication patterns, teaching and learning styles. The Language of instruction, assessment and cultural factors in students' test performance are tightly intertwined. However, most studies have focused on students' background and cultural factors.

The foundation for scientific instruction and learning is formed by the science process skills. The basic science process skills consist of the following: Observation; Measurement; Experimentation;recording; Communication; Inference; and Prediction skills. The science process skills are the tools used to construct science concepts and are what students use to investigate the world around them. Many researchers have posited that how students learn to think scientifically and the language is intertwined (Gee, (2005); Michaels and Sohmer, (2000)).

Thinking scientifically involves the appropriation of the ways that scientists use language. According to Halliday and Martin (1993), scientific communication consists of a variety of genres, each consisting of distinct patterns of linguistic features, arranged such that certain aspects of scientific knowledge and reasoning can be communicated efficiently. Scientific researchers recognize that inquiry processes require language skills, thus inquiry is taken as an opportunity to incorporate explicit development of science-related literacy skills.

This study, therefore, attempts to establish the relationship between achievement (proficiency) in the language of instruction and the achievement in the sciences; Biology, Chemistry and Physics.

# Chapter 3

# Methodology

## 3.1 Introduction

This chapter explains the model that was used in this study. The study involved trying to establish the relationship between achievements in the languages; English and Kiswahili and the achievement in mathematics and the sciences using partial least squares path modeling (PLS-PM).

## 3.2 Research Design

The study used KNEC examination results for KCSE 2014. KNEC conduct public academic technical and other national examinations within Kenya at basic and tertiary levels. The Kenya Certificate of Secondary Education (KCSE) examination is done after one completes four years of study in secondary school. The KCSE 2014 results are for 486430 candidates from all the secondary schools in the 47 counties in Kenya.
A total of 65535 candidates sat for KCSE in the three science subjects, namely Biology, Physics and Chemistry in 2014.
The study involved analyzing examination results in English, Kiswahili, Mathematics, Chemistry, Physics and Biology for the 65535 candidates to establish the relationship between the languages,mathematics and the sciences. The method used for this study is the Partial Least Square path Modeling (PLS-PM).

## 3.3 Partial Least Square-Path modeling

Structural Equation Models (SEM) are statistical models that allow researchers to study real world complexity by taking into account a number of causal relationships among latent concepts each measured by several observed indicators. The system can be studied defining the causality network among latent variables (LV), each measured by observed indicators defined as Manifest Variables

(MV).
The Partial Least Squares (PLS) approach to Structural Equation models is also known as the Partial Least Square Path Modeling (PLS-PM). PLS Path Modeling is a component-based estimation method (Tenenhaus 2008). It is an iterative algorithm that separately solves out the blocks of the measurement model and then estimates the path coefficients in the structural model.
PLS-PM is regarded as a soft modeling approach because there are no strong assumptions with respect to the distributions of the data, the sample size and the measurement scale. PLS Path Modeling is a component-based alternative for estimating Structural Equation Models that we can use in understanding the structure of the data. PLS path modeling is also a method for analyzing a system of linear relationships between multiple blocks of variables.

### 3.3.1 Manifest and latent variables

Given a domain and population of interest, the researcher selects from the domain variables which are to be measured. A manifest attribute is any of the many attributes of a population that can be observed and measured. Manifest variables are also known as indicators, items, surface attributes or measurable variables.
A Latent variable is an internal attribute. It can also be taken to be an unobservable characteristic of the population. Latent variables cannot be measured directly, but their effects are reflected on surface attributes.Latent variables are commonly referred to as internal attributes, hypothetical constructs, composites, hypothetical variables, theoretical concepts, intangibles, and factors.
In this project, the manifest variables are the grades in English, Kiswahili, Mathematics, Biology, Physics, Chemistry and the mean grade. The latent variables are literacy, process skills, numeracy and graduation.

### 3.3.2 PLS-Path Model

A PLS path model is composed of two models:

1. An outer model relating the manifest variables (MV) to their latent variable (LV). Also called measurement model.
2. A inner model relating latent variables (LV) to other latent variables(LV). Also called a structural model.

#### The Outer Model

A latent variable (LV) is an unobservable variable (or construct) indirectly described by a block of observable variables $X_{ij}$ which are called manifest variables (MV) or indicators. The outer model formulation depends on the direction of the relationship between the latent variables and the corresponding manifest variables. The following ways can relate the MV to their LV: the reflective

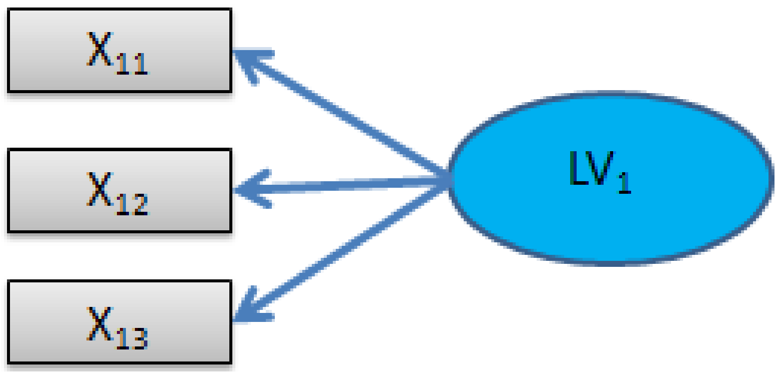


Figure 3.1: Diagram of a reflective block

way, the formative way and the multiple effects indicator for multiple causes (MIMIC) way.

**The Reflective model**

For a reflective model, the manifest variables related to a latent variable is assumed to measure a unique underlying concept. The latent variable is considered as the cause of the Manifest variables for a case of reflective model. We can represent $LV_1$ measured with three indicators as shown in the figure 3.1:

The outer model relationships in reflective cases are considered to be linear. Hence we have:

$$X_{ij} = \beta_{0jk} + \beta_{jk} LV_j + error_{jk}$$

In the reflective model, the internal consistency has to be checked. Each block is to be assumed to be homogenous and uni-dimensional and the manifest variables linked to the same latent variable should covary: changes in one indicator should imply changes in the others.

In the reflective scheme, the outer model reproduces exactly the factor analysis model in which each manifest variable is a linear function of the underlining factor.

**The formative model**

In the formative scheme, the Manifest variables are considered to be the cause of the Latent variables. In it, each block of manifest variables represents different dimensions of the underlying concept.

The latent variable is obtained as a linear combination of the manifest variables

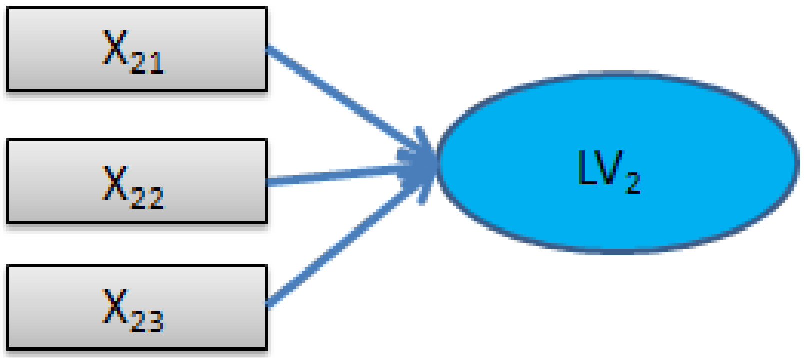


Figure 3.2: Diagram of a formative block

in the outer model. In this case, indicators need not to covary: changes in one indicator do not imply changes in the others. Measures of internal consistency are not necessary.

A latent variable $LV_2$ with three indicators would be represented as shown in fig.3.2: This can be represented as

$$LV_j = \beta_{0j} + \beta_{jk}X_{jk} + error_j$$

$\beta_{jk}$ are the loadings.
$\beta_{0j}$ is the intercept term,
The error terms account for the residuals.

#### Regression specification

The linear relationships for the two ways of representing the outer model are as shown below: for reflective model

$$E(X_{jk}/LV_j) = \beta_{0jk} + \beta_{jk}LV_j$$

for formative model

$$E(LV_j/X_{jk}) = \beta_{0j} + \beta_{jk}X_{jk}$$

A model with all arrows pointing outwards from the latent variables is called a Mode A. In such a case, all latent variables have reflective measurements. A model with all arrows pointing inwards towards a latent variable is called a Mode B. In such a case, all latent variables have formative measurements.

A model containing both, formative and reflective latent variables is referred to as MIMIC or a mode model.

### The MIMIC model

The MIMIC model describes a mixture of the reflective and formative models. The measurement model for a MIMIC block is the following:

$$X_j = \beta_{0j} + \beta_j LV + error_j$$

for $j = 1 to P_1$ where

$$LV = \sum_{h=P_i+1}^{P} w_h X_h + \delta$$

The$P_1$ first manifest variables follow a reflective way and the $P - P_1$ last ones a formative way. In the reflective and formative scheme, the parameters to be estimated are the external or outer weights $(w_{pq})$ and the loadings $(\lambda_{pq})$.

### The Project PLS Path Model

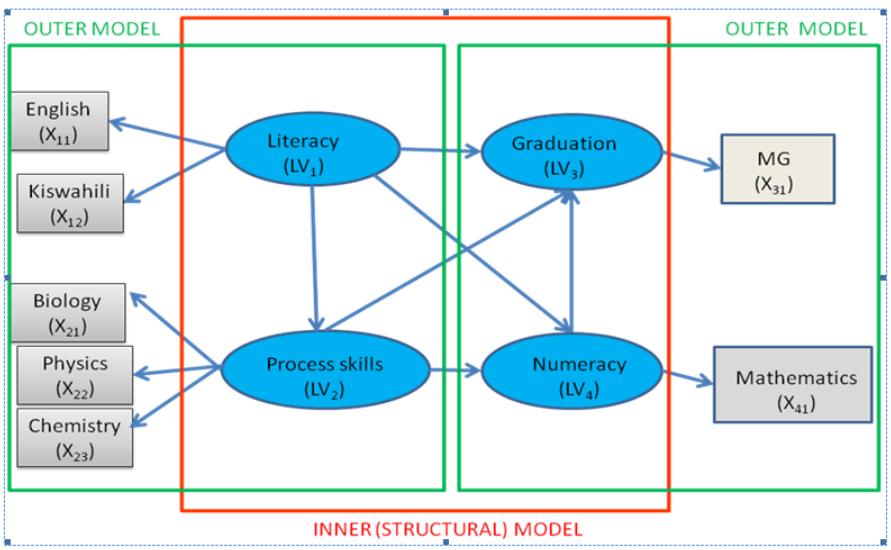


Figure 3.3: The PLS path model for the project

Latent variables in the structural model can be categorized as either exogenous or endogenous. Endogenous latent variables characterizes latent variables that are explained by others in the inner model. An endogenous variable has at least one path leading to it and represents the effects of other variable(s). An exogenous latent variable is used to characterize latent variables with no preceding ones in the structural model.An exogenous variable has path arrows pointing outwards and none leading to it.In figure 3.3, literacy($LV_1$)is an exogenous latent variable while process skills($LV_2$), Graduation($LV_3$) and numeracy($LV_4$)are endogenous latent variables.

1. The measurement model

   $X_{1k} = \lambda_1 k LV_1 + error$ for $k = 1, 2$

   $X_{2k} = \lambda_2 k LV_2 + error$ for $k = 1, 2, 3$

   $X_{31} = \lambda_3 LV_3 + error$

   $X_{41} = \lambda_4 LV_4 + error$

   Where $\lambda_{jk}$ are loadings

2. The structural model

   The causality model in figure 3.3 leads to linear equations relating the LVs between them.

   $$LV_j = \beta_{0j} + \sum_{i \leftrightarrow \mathrm{J}} \beta_{ij} LV_i + error_j$$

   The causality model must be a causal chain. This means that there is no loop in the causality model. This kind of model is called recursive. The structure equations corresponding to the figure above are:

   $$LV_2 = \beta_{20} + \beta_{21} + error$$

   $$LV_3 = \beta_{30} + \beta_{31} + \beta_{32} + \beta_{34} + error$$

   $$LV_4 = \beta_4 0 + \beta_{41} + \beta_{42} + error$$

a structural model can be summarized using inner design matrix. The design matrix is a 0/1 square matrix with dimensions equal to the number of LVs. Rows and columns represent the LVs. A cell $(i, j)$ is filled with a 1 if $LV_j$ explains $LV_i$ , and 0 otherwise. For the project PLS path model,the inner design matrix is written as:

| LV | Literacy | Process skills | Numeracy | Graduation |
|---|---|---|---|---|
| Literacy | 0 | 0 | 0 | 0 |
| Process skills | 1 | 0 | 0 | 0 |
| Numeracy | 1 | 1 | 0 | 0 |
| Graduation | 1 | 1 | 1 | 0 |

**The Weight Relations**

In PLS-Path model, latent variables are estimated as a linear combination of their manifest variables. An estimated $LV_j$ is called a score, denoted as $Y_j$:

$$\widehat{LV_j} = Y_j = \sum_k W_{jk} X_{jk}$$

PLS-Path modeling is referred to as a component-based approach because latent variables are calculated as a weighted sum of their indicators, similar to principal component analysis.

**Measurement Model Assessment: Reflective Indicators**

Reflective indicators are supposed to measure the same underlying latent variable. Reflective block should be homogenous and uni-dimensional. Three aspects of reflective measures are evaluated:

1. Uni-dimensionality of the indicators
2. Check that indicators are well explained by its latent variable
3. Assess the degree to which a given construct is different from other constructs.

The following tools can be used to checking for the block homogeneity and uni-dimensionality:

1. Cronbach's alpha: A block is considered homogenous if this index is larger than 0.7 for confirmatory studies.
2. Dillon-Goldstein's rho: a block is considered homogenous if this index is larger than 0.7
3. Principal component analysis of a block: a block may be considered uni-dimensional if the first eigenvalue of its correlation matrix is higher than 1, while the others are smaller (Kaiser's rule).

In the formative model, each manifest variable or each sub-block of manifest variables represents a different dimension of the underlying concept. Thus, unlike the reflective model, the formative model does not assume homogeneity or uni-dimensionality of the block. The latent variable is defined as a linear combination of the corresponding manifest variables, thus each manifest variable is an exogenous variable in the measurement model. These indicators need not covary: changes in one indicator do not imply changes in the others and internal consistency is not an issue.

### 3.3.3 PLS-PM Algorithm

PLS-PM algorithm involves a series of simple and multiple ordinary least squares regressions. The estimation of the path coefficients involves running many least squares regressions as structural equations in the model. Consequently, obtaining the loadings involves computing simple correlations. In PLS Path Modeling, an iterative procedure permits estimation of the outer weights and the latent variable scores. The estimation procedure is named partial since it solves blocks one at a time by means of alternating single and multiple linear regressions. The path coefficients are estimated afterwards by means of a regular regression between the estimated latent variable scores in accordance with the specified network of structural relations.

PLS Path Modeling is an iterative algorithm that separately solves out the blocks of the measurement model and then, in a second step, estimates the path coefficients in the structural model. Therefore, PLS-PM best explains the residual variance of the latent variables and, potentially, also of the manifest variables in any regression run in the model (Fornell and Bookstein 1982). The procedure followed in PLS Path Modeling algorithm involve three major stages:

1. Get the weights to compute latent variable scores
2. Estimating the path coefficients (inner model)
3. Obtaining the loadings (outer model)

Stage 1: The iterative process

1. Compute the external approximation of latent variables
2. Obtain inner weights
3. Compute the internal approximation of latent variables
4. Calculate new outer weights

Repeat step 1 to step 4 until convergence of outer weights

**Initial arbitrary outer weights**

We start the iterative process by assigning arbitrary values to the outer weights, e.g. we can initialize all weights equal to one: $W_{jk} = 1$.

$$W_1 = (W_{11}, W_{12})$$

$$W_2 = (W_{21}, W_{22}, W_{23})$$

$$W_3 = (W_{31})$$

$$W_4 = (W_{41})$$

**External estimation**

The PLS-PM algorithm defines a system of weights to be applied at each block of MV in order to estimate the corresponding LV, according to the weight relation:

$$Y_k \propto X_k \widehat{W_k}$$

for k=1,2,3
Decomposing the formula for each LV in project PLS Path model we have following:

$$Y_1 \propto 1X_{11} + 1X_{12}$$

$$Y_2 \propto 1X_{21} + 1X_{22} + 1X_{23}$$

$$Y_1 \propto 1X_{31}$$

$$Y_1 \propto 1X_{41}$$

The symbol $\propto$ is used to indicate that each score $Y_j$ depends on its manifest variables $X_{jk}$. The manifest variables should first be centred (or standardized)

$$Y_j = \pm \sum_k W_{jk} X_{jk}$$

The $\pm$ sign shows the sign ambiguity. This ambiguity is usually solved by choosing the sign making the outer estimate positively correlated to a majority of its manifest variables.

$$sign\left[\sum_k sign\{cor(X_{jk}, Y_j)\}\right]$$

The standardized LVs are finally expressed as:

$$Y_j = W_{jk} X_{jk}$$

The weights $W_{jk}$ are the definite outer weights. In the project PLS Path model we have:

$$Y_1 = 1X_{11} + 1X_{12}$$

$$Y_2 = 1X_{21} + 1X_{22} + 1X_{23}$$

$$Y_1 = 1X_{31}$$

$$Y_1 = 1X_{41}$$

### Obtain Inner weights

Inner weight estimates of $LV_j$, denoted as $Z_j$ are computed as follows:

$$Z_j = \sum_{i \leftrightarrow \mathrm{J}} e_{ij} Y_j$$

where $e_{ij}$ are inner weights

$$Z_1 = \sum_{i \leftrightarrow \mathrm{J}} e_{i1} Y_i = e_{21} Y_2 + e_{31} Y_3 + e_{41} Y_4$$

$$Z_2 = \sum_{i \leftrightarrow \mathrm{J}} e_{i2} Y_i = e_{12} Y_1 + e_{32} Y_3 + e_{42} Y_4$$

$$Z_3 = \sum_{i \leftrightarrow \mathrm{J}} e_{i3} Y_i = e_{13} Y_1 + e_{23} Y_2 + e_{43} Y_4$$

$$Z_4 = \sum_{i \leftrightarrow \mathrm{J}} e_{i4} Y_i = e_{14} Y_1 + e_{24} Y_2 + e_{34} Y_3$$

### Updating Outer weights

In mode A, we obtain the outer weights $W_{jk}$ estimates with simple regressions of each indicator $X_{j1}$,$X_{j2}$,...,$X_{jk}$ on their latent score $Y_j$

$$\widehat{W_{jk}} = (Y_j^T Y_j)^{-1} Y_j^T X_{jk}$$

### Path Coefficients

The structural path coefficients are estimated by ordinary least squares in the multiple regression of $Y_j$ on the $Y_i's$ related with it:

$$Y_j = \sum_{i \leftrightarrow \mathrm{j}} \beta_{ij} Y_j$$

The least squares solution is:

$$\beta_i j = (Y_j^T Y_j)^{-1} Y_j Y_j$$

### Loadings

Loadings are calculated as correlations between a latent variable and its indicators.

$$\widehat{\lambda_{jk}} = cor(X_{jk}, Y_j)$$

# Chapter 4

# DATA ANALYSIS AND RESULTS

## 4.1 INTRODUCTION

This chapter discusses the data analysis and highlights the findings of the study.

## 4.2 Data analysis software

Although PLS-PM developed in the mid-1960s (Wold, 1973, 1985), there was a lack of PLS path modeling software until mid 2000s. A number of software have been developed for PLS path modeling like LVPLS 1.8 (Lohmöller 1987), PLS-GUI (Li 2005), PLS-Graph (Chin 2003), VisualPLS (Fu 2006a), WarpPLS, LVPLS, XLSTAT, SmartPLS (Ringle et al. 2005) and R. In this study, R software version 3.3.0 (2016-05-03) will be used for data analysis.

### 4.2.1 Descriptive statistics

Table 4.1: Descriptive statistics

| **Subject** | **Mean** | **Median** | **Standard deviation** |
|---|---|---|---|
| **English** | 7.459 | 8.000 | 2.320 |
| **Kiswahili** | 7.538 | 8.000 | 2.698 |
| **Mathematics** | 7.214 | 7.000 | 3.596 |
| **Biology** | 7.607 | 8.000 | 2.881 |
| **Physics** | 6.466 | 6.000 | 3.007 |
| **Chemistry** | 7.175 | 7.000 | 3.025 |

### 4.2.2 Visualizing distribution of the data

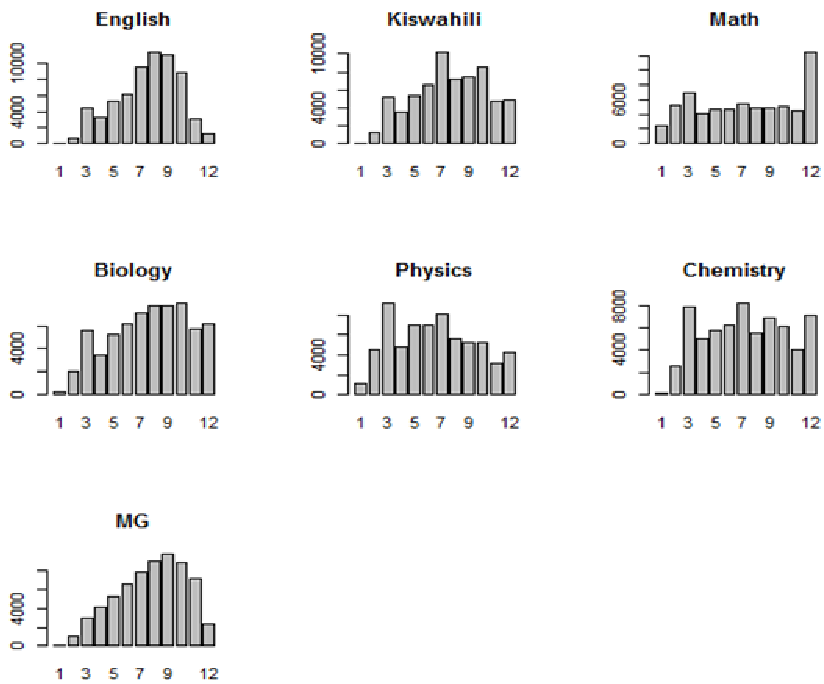


Figure 4.1: Distribution of the data

English, Biology and mean grade (MG) are skewed to the right. All the other manifest variables are fairly normally distributed.

### 4.2.3 Correlation of the Indicators

The manifest variables have high correlation coefficients, ranging between 0.66 and 0.91. This indicates a high multi- collinearity among the variables.

### 4.2.4 Path diagram of the inner model

There is no global goodness-of-fit criterion in PLS path modeling. Chin (1998) put forward a two-step process criteria to assess partial model structures. The steps are:

1. Assessment of the outer model and

2. Assessment of the inner model.

| Subject | English | Kiswahili | Math | Biology | Physics | Chem | MG |
|---|---|---|---|---|---|---|---|
| **English** | 1.00 | 0.77 | 0.67 | 0.72 | 0.68 | 0.71 | 0.84 |
| **Kiswahili** | 0.77 | 1.00 | 0.66 | 0.77 | 0.68 | 0.74 | 0.86 |
| **Mathematics** | 0.67 | 0.66 | 1.00 | 0.71 | 0.83 | 0.80 | 0.86 |
| **Biology** | 0.72 | 0.77 | 0.71 | 1.00 | 0.76 | 0.84 | 0.90 |
| **Physics** | 0.68 | 0.68 | 0.83 | 0.76 | 1.00 | 0.84 | 0.87 |
| **Chemistry** | 0.71 | 0.74 | 0.80 | 0.84 | 0.84 | 1.00 | 0.91 |
| **MG** | 0.84 | 0.86 | 0.86 | 0.90 | 0.84 | 0.87 | 1.00 |

Table 4.2: Correlations

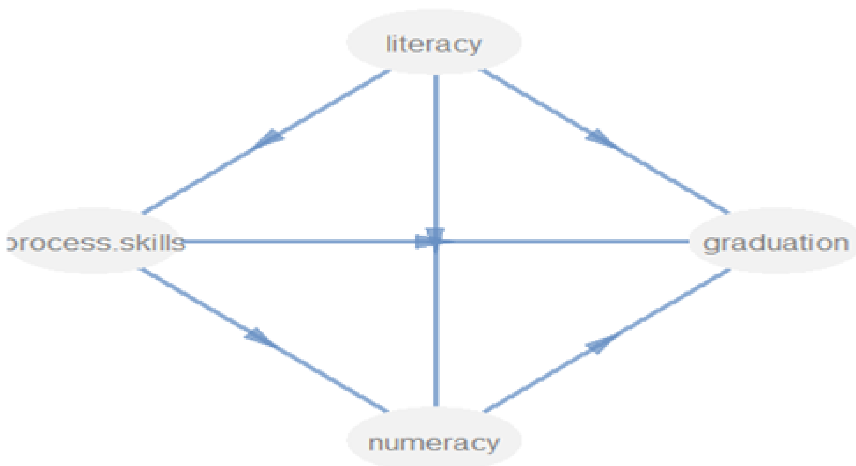


Figure 4.2: Project's inner model

### 4.2.5 Assessment of measurement Model

Reflective measurement models should be assessed for reliability and validity. The criterion which is checked is internal consistency reliability. The traditional criterion for internal consistency is Cronbach's $\alpha$ which provides an estimate for the reliability based on the indicator inter-correlations.

***Uni-dimensionality of indicators***

1. The Cronbach's alpha

   The Literacy block has an alpha of 0.87; Process skills has an alpha of 0.92; Numeracy has an alpha of 1.00 and graduation has an alpha of 1.00. All the alpha values are above 0.7 and therefore acceptable.

2. Dillon-Goldstein's rho

Table 4.3: Uni–dimensionality of indicators

| LV | Mode | MVs | C.alpha | DG.rho | eig.1st | eig.2nd |
|---|---|---|---|---|---|---|
| **Literacy** | A | 2 | 0.87 | 0.94 | 1.77 | 0.23 |
| **Process skills** | A | 3 | 0.92 | 0.96 | 2.63 | 0.24 |
| **Numeracy** | A | 1 | 1.00 | 1.00 | 1.00 | 0.00 |
| **Graduation** | A | 1 | 1.00 | 1.00 | 1.00 | 0.00 |

Table 4.4: DG rho

| LV | DG.rho |
|---|---|
| **Literacy** | 0.94 |
| **Process skills** | 0.96 |
| **Numeracy** | 1.00 |
| **Graduation** | 1.00 |

The Literacy block has a rho value 0.94; Process skills has an rho value of 0.96; Numeracy has rho value of 1.00 and graduation has a rho of 1.00. All the rho values are above 0.9.

3. First eigenvalue

| LV | eig.1st | eig.2nd |
|---|---|---|
| **Literacy** | 1.77 | 0.23 |
| **Process skills** | 2.63 | 0.24 |
| **Numeracy** | 1.00 | 0.00 |
| **Graduation** | 1.00 | 0.00 |

Table 4.5: Eigenvalues

The first eigenvalue is greater than 1 for literacy and process.skills; and is equal to 1 for computational and graduation. All the second eigenvalues are less than 1. This implies that the all blocks are uni-dimensional.

### Loadings and communalities

The loadings are correlations between a latent variable and its indicators, whereas communalities are the squared correlations. Loadings explain how each manifest variable relates to each construct. A measure in question is said to be able to discriminate when it is strongly related to the construct it attempts to reflect, and does not have a stronger connection with another construct. As can be seen in figure 4.3, all the loadings are positive. All the Loadings are above 0.70. This indicates that the constructs explain over 50% of the indicators' variance.

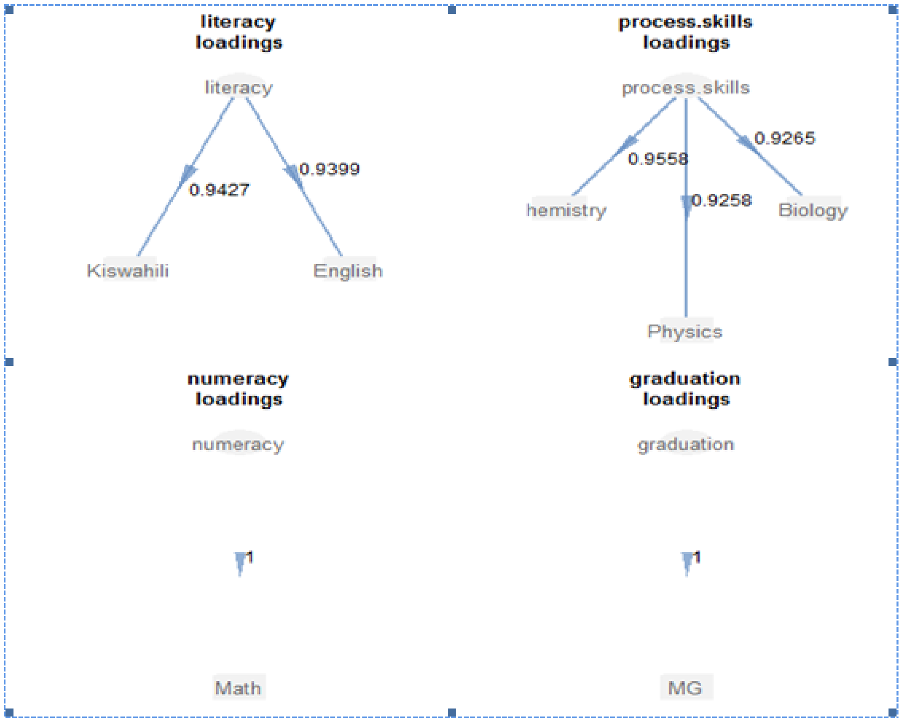


Figure 4.3: Loadings

As can be seen in figure 4.4, all the weights for the literacy block indicators, process skills block indicators, numeracy indicator and graduation indicator are positive. This means that the Cronbach's alpha and the Dillon-Goldstein's rho are adequate for this model.

**Cross-loadings**

Cross-loadings can also be used to check for discriminant validity. An indicator is not expected to have a higher correlation with another latent variable than with its respective latent variable.

As can be seen in table 4.7, all the indicators have the highest correlation with their latent variable indicating the suitability of the model.

### 4.2.6 Structural Model Evaluation

Evaluation of the inner path is done after confirmation of reliability and validity of the outer model estimates. Figure 4.5 shows that the the path coefficients linking literacy to process skills and process skills to numeracy are high at above 0.7. The path coefficient linking literacy to numeracy is lowest at 0.0779.

| Name | block | weight | loading | communality | redundancy |
|---|---|---|---|---|---|
| **English** | Literacy | 0.525 | 0.940 | 0.883 | 0.00 |
| **Kiswahili** | Literacy | 0.537 | 0.943 | 0.889 | 0.00 |
| **Biology** | process skill | 0.352 | 0.926 | 0.858 | 0.569 |
| **Physics** | process skill | 0.355 | 0.926 | 0.857 | 0.568 |
| **Chemistry** | process skill | 0.362 | 0.956 | 0.914 | 0.605 |
| **Mathematics** | Numeracy | 1.000 | 1.000 | 1.000 | 0.701 |
| **MG** | graduation | 1.000 | 1.000 | 1.000 | 0.966 |

Table 4.6: Loadings and communalities

| Name | block | Literacy | process skills | Numeracy | graduation |
|---|---|---|---|---|---|
| **English** | Literacy | 0.940 | 0.750 | 0.669 | 0.837 |
| **Kiswahili** | Literacy | 0.943 | 0.782 | 0.662 | 0.865 |
| **Biology** | process skill | 0.791 | 0.926 | 0.714 | 0.901 |
| **Physics** | process skill | 0.726 | 0.926 | 0.834 | 0.865 |
| **Chemistry** | process skill | 0.769 | 0.956 | 0.799 | 0.907 |
| **Mathematics** | Numeracy | 0.707 | 0.836 | 1.000 | 0.860 |
| **MG** | graduation | 0.904 | 0.952 | 0.860 | 1.000 |

Table 4.7: Crossloadings

### Inner model

To evaluate the quality of the structural model three indices are examined:

1. the $R^2$ determination coefficients
2. the redundancy index
3. the Goodness-of-Fit (GoF)

### Coefficients of determination $R^2$

The $R^2$ values are assessed to determine the predictive power of the structural model. The $R^2$ are the coefficients of determination of the endogenous latent variables.

| Name | Type | $R^2$ | Block communality | Mean Redundancy | AVE |
|---|---|---|---|---|---|
| **Literacy** | Exogenous | 0.000 | 0.886 | 0.000 | 0.886 |
| **process skills** | Endogenous | 0.662 | 0.876 | 0.580 | 0.876 |
| **Numeracy** | Endogenous | 0.701 | 1.000 | 0.701 | 1.000 |
| **graduation** | Endogenous | 0.966 | 1.000 | 0.966 | 1.000 |

Table 4.8: Inner model summary

$R^2$ indicates the amount of variance in the endogenous latent variable explained by its independent latent variables.From table 4.8, it can be seen that the $R^2$ values are above 0.6, which is high.

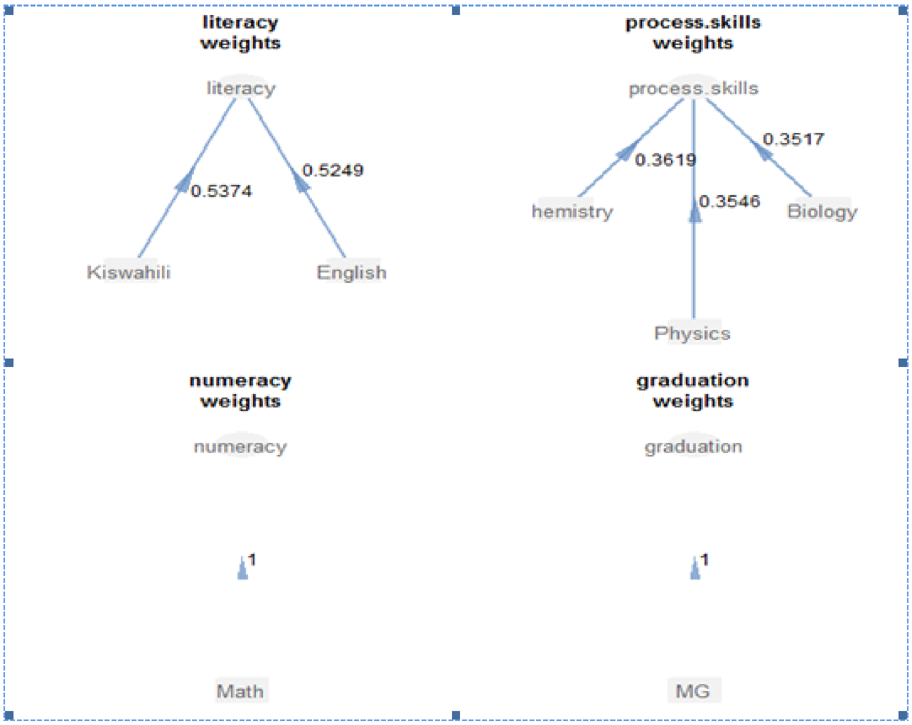


Figure 4.4: Weights

### Redundancy

Redundancy reflects the ability of a set of independent latent variables to explain variation in the dependent latent variable.It measures the percentage of the variance of indicators in an endogenous block that is predicted from the independent latent variables associated to the endogenous LV. Redundancy can also be defined as the amount of variance in an endogenous construct explained by its independent latent variables.

The redundancy index for the j-th manifest variable associated to the k-th block is:

$$Rd(LV_k, mv_{jk}) = loading_{jk}^2 R_k^2$$

From table 4.9, it can be seen that the mean redundancy for the endogenous blocks is above 0.6. This means redundancy is high, which means high ability to predict.

### Convergent validity

Convergent validity measures the extent to which a construct converges in its indicators by explaining the item's variance. Convergent validity signifies that a set of indicators represents one and the same underlying construct, which can be

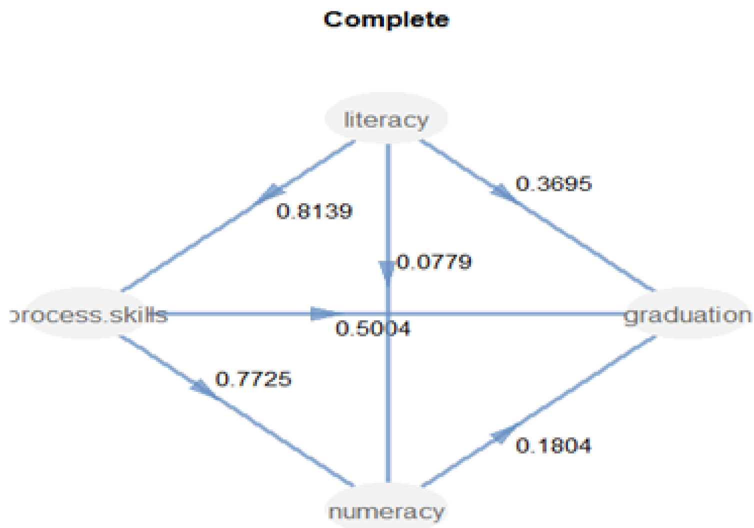


Figure 4.5: Inner model with path coefficients

| Name | Type | $R^2$ | Mean Redundancy |
|---|---|---|---|
| **Literacy** | Exogenous | 0.000 | 0.000 |
| **process skills** | Endogenous | 0.662 | 0.580 |
| **Numeracy** | Endogenous | 0.701 | 0.701 |
| **graduation** | Endogenous | 0.966 | 0.966 |
| **graduation** | Endogenous | 0.966 | 0.966 |

Table 4.9: Mean redundancy

demonstrated through their uni-dimensionality. Convergent validity is assessed by the average variance extracted (AVE) for all items associated with each construct. The AVE value is calculated as the mean of the squared loadings for all indicators associated with a construct. An acceptable AVE is 0.50 or higher, as it indicates that on average, the construct explains over 50% of the variance of its items.From table 4.8, it can be seen that the AVE of the constructs in the model is above 0.8 and therefore acceptable. This implies that the convergent validity is sufficient in this model.

### GoF

The Goodness of fit index is a measure that accounts for the model quality at both the measurement and the structural models. GoF assess the overall prediction performance of the model. The GoF index for the model is 0.827. This implies that the predictive power of the model is 83%.

### Structural Regressions

The path coefficients of the PLS structural model can be interpreted as standardized beta coefficients of ordinary least squares regressions. It is important to review the regression results of each endogenous construct.

1. **Process skills**

Table 4.10: Regression with process skills as response variable

| | **Estimate** | **Std error** | **t value** | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| **Intercept** | $-3.85 \times 10^{-17}$ | 0.0023 | $-1.69 \times 10^{-14}$ | 1 |
| **Literacy** | $8.14 \times 10^{-1}$ | 0.0023 | $3.57 \times 10^{2}$ | 0 |

2. **Numeracy**

Table 4.11: Regression with numeracy as response variable

| | **Estimate** | **Std error** | **t value** | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| **Intercept** | $-1.43 \times 10^{-16}$ | 0.002 | $-6.66 \times 10^{1}$ | 1.00 |
| **Literacy** | $7.79 \times 10^{-2}$ | 0.004 | $2.11 \times 10^{1}$ | 0.00 |
| **Process skills** | $7.72 \times 10^{-1}$ | 0.004 | $2.09 \times 10^{2}$ | 0.00 |

3. **Graduation**

| | **Estimate** | **Std error** | **t value** | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| **Intercept** | $-1.41 \times 10^{-16}$ | 0.000 | $-1.93 \times 10^{-13}$ | 1.00 |
| **Literacy** | $3.69 \times 10^{-1}$ | 0.00 | $2.94 \times 10^{2}$ | 0.00 |
| **Process skills** | $5.00 \times 10^{-1}$ | 0.00 | $3.09 \times 10^{2}$ | 0.00 |
| **Numeracy** | $1.80 \times 10^{-1}$ | 0.00 | $1.36 \times 10^{2}$ | 0.00 |

Table 4.12: Regression with graduation as response variable

It can be observed that all the p-values for the path coefficients are less than 0.05. This implies that the path coefficients are significant and can be used in prediction in the regression equations.

$$Processskills = 0.814(Literacy) + \epsilon$$

$$Numeracy = 0.0779(Literacy) + 0.772(Processskills) + \epsilon$$

$$Graduation = 0.369(Literacy) + 0.500(Processskills) + 0.180(Numeracy) + \epsilon$$

### Bootstrap Validation

Bootstrap re-sampling is used to get confidence intervals for evaluating the precision of the PLS parameter estimates. This is done after checking the nature of the results of the outer and inner models. We examine the bootstrap confidence interval provided by the percentiles 0.025 and 0.975 especially for the path coefficients.

| | Original | Mean.Boot | Std.Error | perc.025 | perc.975 |
|---|---|---|---|---|---|
| **Literacy→ process skills** | 0.81386 | 0.81386 | 0.00137 | 0.81175 | 0.81649 |
| **Literacy→ numeracy** | 0.07795 | 0.07806 | 0.00402 | 0.07065 | 0.08415 |
| **Literacy→ graduation** | 0.36950 | 0.36966 | 0.00116 | 0.36757 | 0.37189 |
| **process skills→ numeracy** | 0.77249 | 0.77233 | 0.00357 | 0.76630 | 0.77967 |
| **process skills→ graduation** | 0.50043 | 0.50026 | 0.00151 | 0.49762 | 0.50344 |
| **numeracy→ graduation** | 0.18039 | 0.18040 | 0.00149 | 0.17767 | 0.18357 |

Table 4.13: Results for bootstrap validation

From table 4.13, it can be observed that all the path coefficients do not contain zero. Hence all the path coefficients are significant at 5% confidence level.
From the obtained results, we can say that achievement in languages contribute to a student's performance in science and mathematics ; achievement in science contribute to a student's performance in mathematics; and that achievement in languages, mathematics and science contribute to the final mean score of a student.

## 4.3 Multi-group comparison of PLSM-path models

PLS Path models for the school types may be different at many levels and there is need to compare the path models. There are four major types of differences:

1. Causal network level: differences in the assumed causal-effect network linking the latent variables.
2. structural level: these are differences in magnitude of the structural coefficients (i.e. the path coefficients)
3. Measurement level: this refers to the way in which the latent variables are defined by their indicators.
4. Latent variables level: this implies that the mean value of latent variables across models may be different.

### 4.3.1 School type models

Given school types, it is important to examine whether there are differences between Girls' and Boys' schools; Girls' and Mixed schools; and Boys' and Mixed schools. This involves calculating PLS path models separately for Girls' schools, Boys' schools and mixed schools.

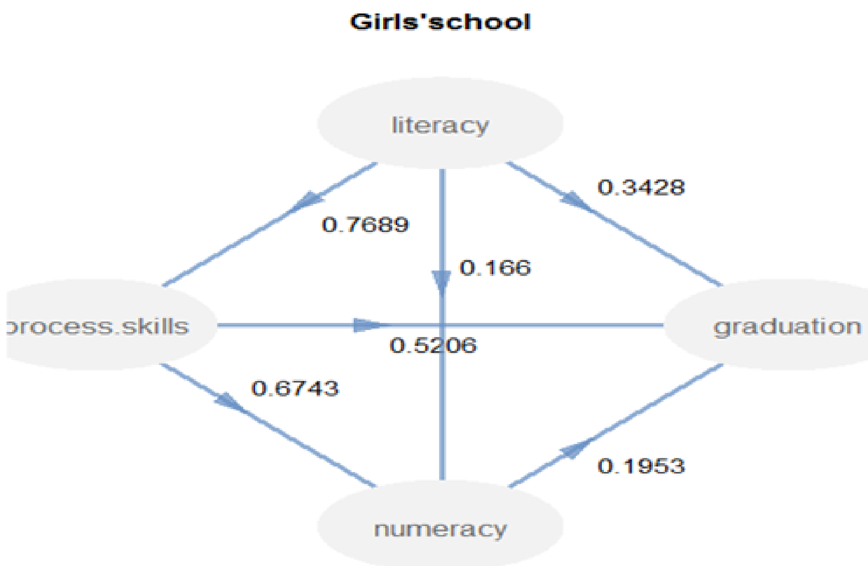


Figure 4.6: Inner model with path coefficients for Girls' schools

We examine the path coefficients of the structural models in order to compare the models. The path coefficients are different in all the school type models. From figures 4.6, 4.7,and 4.8, it can be seen that the path coefficient between literacy and numeracy is highest for girls' schools at 0.166 and lowest for boys' schools at 0.0366. The path coefficient between literacy and process skills is highest for mixed schools at 0.7914 and lowest for boys' schools at 0.7653. The path coefficient between process skills and numeracy is highest for mixed schools at 0.7998 and lowest for girls' schools at 0.6743.
It is important to examine how different the path coefficients are.

### 4.3.2 Comparing Groups: Bootstrap t-test

1. **Boys and Girls schools** Group comparison in PLS-PM for path coefficients

   As we can see from table 4.14, two of the path coefficients between Boys' schools and Girls' schools are significantly different at 5% confidence level,

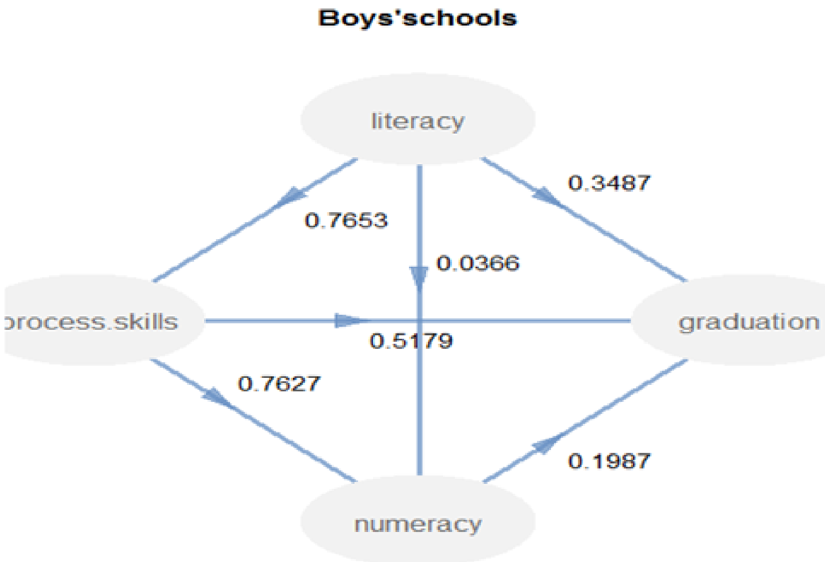


Figure 4.7: Inner model with path coefficients for Boys' schools

| | global | group.B | group.G | diff.abs | t.stat | deg.fr | p.value | sig.05 |
|---|---|---|---|---|---|---|---|---|
| **Literacy→ process skills** | 0.7573 | 0.7653 | 0.7689 | 0.0036 | 0.7891 | 37101 | 0.2150 | No |
| **Literacy→ numeracy** | 0.1080 | 0.0.0366 | 0.1660 | 0.1294 | 12.4126 | 3710 | 10.0000 | Yes |
| **Literacy→ graduation** | 0.3441 | 0.3487 | 0.3428 | 0.0060 | 1.5493 | 37101 | 0.0607 | No |
| **process skills→ numeracy** | 0.7087 | 0.7627 | 0.6743 | 0.0884 | 9.3892 | 37101 | 0.0000 | Yes |
| **process skills→ graduation** | 0.5269 | 0.5179 | 0.5206 | 0.0028 | 0.7383 | 37101 | 0.2302 | No |
| **numeracy→ graduation** | 0.1932 | 0.1987 | 0.1953 | 0.0034 | 1.0907 | 37101 | 0.1377 | No |

Table 4.14: Comparison of boys and girls schools

i.e. the path coefficients between literacy and numeracy and between process skills and numeracy.

2. **Boys and Mixed Schools**

   Group comparison in PLS-PM for path coefficients

   From the results in table 4.15, it can be seen that all the path coefficients are significantly significant except one. The path coefficient between literacy and numeracy is not significantly different at 5% confidence level.

3. **Girls and Mixed Schools**

   Group comparison in PLS-PM for path coefficients

   As we can see in table 4.16, four of the path coefficients between Girls' schools and mixed schools are significantly different. Only the path coefficient between numeracy and graduation is not significantly different at 5% confidence level.

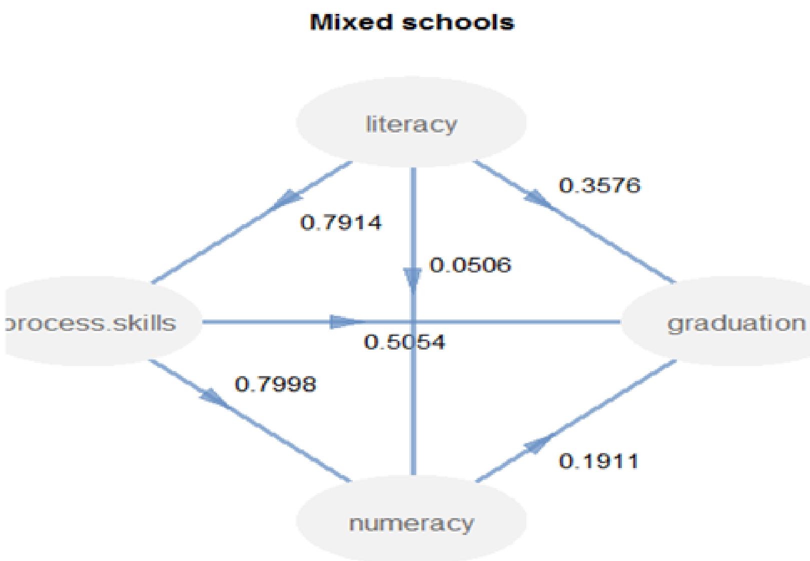


Figure 4.8: Inner model with path coefficients for Mixed schools

| | global | group.B | group.M | diff.abs | t.stat | deg.fr | p.value | sig.05 |
|---|---|---|---|---|---|---|---|---|
| **Literacy→ process skills** | 0.8230 | 0.7653 | 0.7914 | 0.0261 | 7.4026 | 51268 | 0.0000 | Yes |
| **Literacy→ numeracy** | 0.0297 | 0.0.0366 | 0.0506 | 0.0140 | 1.5879 | 51268 | 0.0562 | No |
| **Literacy→ graduation** | 0.3692 | 0.3487 | 0.3576 | 0.0088 | 2.5709 | 51268 | 0.0051 | Yes |
| **process skills→ numeracy** | 0.8154 | 0.7627 | 0.7998 | 0.0371 | 5.0264 | 51268 | 0.0000 | Yes |
| **process skills→ graduation** | 0.5000 | 0.5179 | 0.5054 | 0.0125 | 2.9697 | 51268 | 0.0015 | Yes |
| **numeracy→ graduation** | 0.1796 | 0.1987 | 0.1911 | 0.0076 | 2.1494 | 51268 | 0.0158 | Yes |

Table 4.15: Comparison of boys and mixed schools

| | global | group.G | group.M | diff.abs | t.stat | deg.fr | p.value | sig.05 |
|---|---|---|---|---|---|---|---|---|
| **Literacy→ process skills** | 0.8105 | 0.7689 | 0.7914 | 0.0224 | 5.2439 | 41905 | 0.0000 | Yes |
| **Literacy→ numeracy** | 0.1024 | 0.0.1660 | 0.0506 | 0.1154 | 11.1759 | 41905 | 0.0000 | Yes |
| **Literacy→ graduation** | 0.3697 | 0.3428 | 0.3576 | 0.0148 | 3.6398 | 41905 | 0.0001 | Yes |
| **process skills→ numeracy** | 0.7637 | 0.6743 | 0.7998 | 0.1255 | 13.2177 | 41905 | 0.0000 | Yes |
| **process skills→ graduation** | 0.4939 | 0.5206 | 0.5054 | 0.0153 | 3.2869 | 41905 | 0.0005 | Yes |
| **numeracy→ graduation** | 0.1860 | 0.1953 | 0.1911 | 0.0042 | 1.0387 | 41905 | 0.1495 | No |

Table 4.16: Comparison of girls and mixed schools

# Chapter 5

# CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion and Discussion

In the complete model, Literacy contributed more to the development of science process skills at 0.8139 than to the numeracy skills (at 0.0779). Science process skills contributed 0.7725 to the development of numeracy skills. Literacy and science process skills combined determine the development of numeracy skills, i.e. numeracy=f (science skills, literacy)
Achievement in literacy(English and Kiswahili)contributed 0.8139 to the achievement in the science subjects.Achievement in the science subjects contribute 0.7725 to the achievement in mathematics.Achievement in languages contribute directly and indirectly to the development of numeracy skills.
Language contributed more to the achievement in mathematics for girls' schools than for boys' schools and mixed schools. Achievement in science subjects contributed more to the achievement in mathematics for boys' schools than for girls' schools.
Language contribute highest to the achievement in science subjects in mixed schools as compared to boys' and girls' schools.
In all the cases, the direct contribution of language to the achievement in mathematics is low as compared to the contribution to the achievement in science. Contribution of science to the achievement in mathematics is high in all the cases

Languages contribute to the development of science process skills which in turn contribute to the development of numeracy skills.

## 5.2 Recommendation

The study have shown that the achievement in languages contribute to the achievement in sciences and mathematics. The study recommends that to address the dismal performance in mathematics and sciences at KCSE, the government should not only focus on mathematics and sciences but should also focus on the development of literacy skills in English and Kiswahili. Capacity development activities for mathematics and science teachers should include enhancement of literacy skills in English and Kiswahili.
Policy–makers should also consider planning for capacity development training for English and Kiswahili teachers.

## 5.3 Suggestion for further research

This study investigated the relationship between achievement in languages and achievement in mathematics and sciences for the KCSE 2014 candidates who were taking all the three science subjects. A similar study can be done for candidates who were selecting the science subjects to study two of them.

**References**


1. GoK(2007). Kenya Vision 2030 Medium Term Plan II Education and Training 2013–2018:Towards a globally Competitive and prosperous Kenya. Nairobi: Government Printers.

2. KIE. (2002). *Secondary Education Syllabus: Volume two.* Nairobi: Kenya Institute of Education.

3. Gilberto J. Cuevas *Journal for Research in Mathematics Education, Vol. 15, No. 2, Minorities and Mathematics.* (Mar., 1984), pp. 134-144. Accessed online

4. Lewis R. Aiken, Jr. *Language Factors in Learning Mathematics. Review of Educational Research*, Vol. 42, No. 3. (Summer, 1972), pp. 359-385. Stable URL:http://links.jstor.org/sici?sici=00346543

5. KNEC (2007) KCSE (2006) Examination Report. Nairobi, Government Printers.

6. SMASSE Project: SPIAS reports (1998/2006)(unpublished)

7. Meiers, M. (2010).*Language in the mathematics classroom. The Digest, NSWIT.* Retrieved Oct 15th,2015, from http://www.nswteachers.nsw.edu.au

8. GoK(2013).*Kenya Vision 2030 Medium Term Plan in Education and Training 2013-2018: Towards a Globally Competitive and Prosperous Kenya.* Nairobi, Government Printers.

9. Ostlund, Karen,(1998). "What the Research Says About Science Process Skills." Electronic Journal of Science Education vol. 2, no. 4 ( http://unr.edu/homepage/jcannon/ejse/ostlund.html.

10. Sanchez, G. ( 2013). PLS Path Modeling with R. Trowchez Editions. Berkeley. http://www.gastonsanchez.com/PLS Path Modeling with R.pdf

11. V. Esposito Vinzi, W.W. Chin, J. Henseler, H. F. Wang (eds.) (2010). Handbook of Partial Least Squares: Concepts, Methods and Applications. Springer- Heidelberg Berlin. Accessed online

12. Sarstedt, M., et al. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy.* http://dx.doi.org/10.1016/j.jfbs.2014.01.002

13. Hair J.F., et al. (2014) Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. ResearchGate. https://www.researchgate.net/publication/2580407

14. Armin M., et al (2012). semPLS: Structural Equation Modeling Using Partial Least Squares. *Journal of Statistical Software*, Volume 48, Issue 3. http://www.jstatsoft.org/

15. Tenenhaus, M., et al (2004). PLS path modeling. Computational Statistics and Data Analysis. Sourced online at www.sciencedirect.com

16. Henseler, J.Ringle (2012). Using Partial Least Square Path Modeling in International Advertising: Basic concepts and Recent Issues. Research Gate. https://www.researchgate.net/publication/233408729